UNIVERSITAT DE BARCELONA

Application of computational docking to the characterization and modulation of protein-protein interactions of biomedical interest

Mireia Rosell Oliveras

UNIVERSITAT DE BARCELONA

FACULTAT DE FARMÀCIA I CIÈNCIES DE L'ALIMENTACIÓ
(Supervisor: Juan Fernández Recio)

APPLICATION OF COMPUTATIONAL DOCKING TO THE CHARACTERIZATION AND MODULATION OF PROTEIN-PROTEIN INTERACTIONS OF BIOMEDICAL INTEREST

MIREIA ROSELL OLIVERAS

UNIVERSITAT DE BARCELONA
FACULTAT DE FARMÀCIA I CIÈNCIES DE L'ALIMENTACIÓ

DOCTORAT EN BIOMEDICINA
Codi HDK05

APPLICATION OF COMPUTATIONAL DOCKING TO THE CHARACTERIZATION AND MODULATION OF PROTEIN-PROTEIN INTERACTIONS OF BIOMEDICAL INTEREST

Memòria presentada per Mireia Rosell Oliveras per optar al títol de Doctora per la Universitat de Barcelona

Tesi doctoral realitzada sota la direcció del Doctor Juan Fernández Recio, al Barcelona Supercomputing Center i l'últim darrer any al ICVV-CSIC.

Tesi adscrita al Departament de Bioquímica i Biologia Molecular de la Facultat de Farmàcia i Ciències de l'Alimentació (Universitat de Barcelona).

Josep Lluis Gelpi Buchaca

Firmado digitalmente por Josep Lluis Gelpi Buchaca
Nombre de reconocimiento (DN): c=ES, l=Barcelona, o=Universitat de Barcelona, cn=Josep Lluis Gelpi Buchaca
Fecha: 2020.09.01 11:22:39 +02'00'

Director	Tutor	Ph.D. candidate
Dr. Juan Fernández Recio	Dr. Josep Lluís Gelpí Buchaca	Mireia Rosell Oliveras

Barcelona, 2020

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

CSIC
Consejo Superior de Investigaciones Científicas

ICVV

Universitat de Barcelona

DECLARACIÓ D'ORIGINALITAT

Jo, MIREIA ROSELL OLIVERAS, matriculada al programa de doctorat de BIOMEDICINA de la Universitat de Barcelona declaro que la tesi titulada "Application of computational docking to the characterization and modulation of protein-protein interactions of biomedical interest" és original, que la investigació que vaig realitzar compleix els codis ètics i les bones practiques i que la tesi no inclou plagi. Soc conscient i accepto per escrit que la meva tesi se sotmetrà al procés adequat per provar l'originalitat dels meus resultats.

31 de Juliol de 2020

Als meus pares

Acknowledgments

Primerament estic molt agraïda a en Juan per tots els coneixements transmessos tots aquests anys, tots els seus consells i suggerencies, tota la ajuda i recolçament així com la seva paciència. De nou, moltes gràcies per fer possible aquest procès dur però enriquidor.

Seguidament, moltes gràcies a tots els companys del BSC i del ICVV. Sempre es un orgull treballar amb vosaltres. Especial agraïment a en Brian, la Chiara, la Lucía, en Miguel i en Luis pel seu recolçament i aprenentatge constant. Una especial menció a tots els companys del grup que tinc especials records i apreci, perque cadascún m'ha aportat alguna cosa, Didier, Dàmaris, Jorge, Ha, Bruno, Iain, Sergio, Laura, Silvia, Mark i Augustina. Agraïments als companys transitoris del ICVV en especial a la Carmen F, a la Carmen C i a en Luis el compi de batalles! I finalment, gràcies a en Mark Wass per la bona acollida en el seu grup a la Universitat de Kent, pocs messos però intensos de feina. Gràcies Mark i a tot el grup, Henry, Magda, Miguel, Morena, Marit i Ariadna.

A més a més, volia donar les gràcies a tot l'equip POCTEFA, ha estat un plaer compartir el projecte amb tots volsaltres, ha estat molt enriquidor professionalment i acadèmicament. Gràcies Xavier dC, Javier S, Yolanda V, Natàlia, Josu, Elena, Juanjo i Alfonso.

Volia donar les gràcies també tots els companys de Life. A més a més, a la planta 0 es va crear un grup multidisciplinar de gent que d'una manera o altre he après molt, gràcies pels dinars, cafès, i moments de desconexió Pau, Dani G, Rti, Enrico, Dani L i Pierlauro.

Paral·lelament aquesta tesi no hagués estat possible sense el suport incondicional de tota la gent que m'estima i m'aprecia, a tots vosaltres, moltíssimes gràcies per compartir rialles, moments i forces.

Per acabar, gràcies infinites a la meva família, a cadascún de vosaltres, gràcies.

"You cannot undo a knot without knowing how it is done."
Aristotle

Abstract

The study of the 3D structural details of protein interactions is essential to understand biomolecular functions at the molecular level. In this context, the limited availability of experimental structures of protein-protein complexes at atomic resolution is propelling the development of computational docking methods that aim to complement the current structural coverage of protein interactions. One of these docking approaches is pyDock, which uses van der Waals, electrostatics, and desolvation energy to score docking poses generated by a variety of sampling methods, typically FTDock or ZDOCK. The method has shown a consistently good prediction performance in community-wide assessment experiments like CAPRI or CASP, and has provided biological insights and insightful interpretation of experiments by modeling many biomolecular interactions of biomedical and biotechnological interest. Here, we describe our approach using pyDock for the structural modeling of protein assemblies and the application of its modules to different biomolecular recognition phenomena, such as modeling of binding mode, interface, and hot-spot prediction, use of restraints based on experimental data, the inclusion of low-resolution structural data, binding affinity estimation, or modeling of homo- and hetero-oligomeric assemblies.

The integration of template-based and ab initio docking approaches is emerging as the optimal strategy for modeling protein complexes and multi-molecular assemblies. We will review the new methodological advances on *ab initio* docking and integrative modeling. The seventh CAPRI edition imposed new challenges to the modeling of protein-protein complexes, such as multimeric oligomerization, protein-peptide, and protein-oligosaccharide interactions. Many of the proposed targets needed the efficient integration of rigid-body docking, template-based modeling, flexible optimization, multi-parametric scoring, and experimental restraints. This was especially relevant for the multi-molecular assemblies proposed in the CASP13-CAPRI46 joint rounds. We will present the results for the 7th CAPRI edition and CAPRI Round 46, the third joint CASP-CAPRI protein assembly prediction challenge.

One of the known potential effects of disease-causing amino acid substitutions in proteins is to modulate protein-protein interactions (PPIs). To interpret such variants at the molecular level and to obtain useful information for prediction purposes, it is important to determine whether they are located at protein-protein interfaces, which are composed of two main regions, core and rim, with different evolutionary conservation and physicochemical properties. Here we have performed a structural, energetics and

computational analysis of interactions between proteins hosting mutations related to diseases detected in newborn screening. Interface residues were classified as core or rim, showing that the core residues contribute the most to the binding free energy of the PPI. Disease-causing variants are more likely to occur at the interface core region rather than at the interface rim ($p < 0.0001$). In contrast, neutral variants are more often found at the interface rim or at the non-interacting surface rather than at the interface core region. We also found that arginine, tryptophan, and tyrosine are over-represented among mutated residues leading to disease. These results can enhance our understanding of disease at the molecular level and thus contribute towards personalized medicine by helping clinicians to provide adequate diagnosis and treatments.

The phenotypic effects of non-synonymous genetic variations leading or predisposing to disease can be rationalized on the basis of the functional and structural impact in the mutated protein, including the perturbation of the interaction network and molecular pathways in which such protein is involved. Therefore, understanding these effects at the molecular level is essential to build accurate disease models and to achieve higher precision in diagnosis and therapeutic intervention. In this context, we can computationally characterize the effect of pathological mutations on specific protein-protein interactions ("edgetic"), based on their protein structure, if available, or on docking models. Protein-protein interactions that are clearly stabilized or destabilized by these mutations can be potential targets for therapeutic intervention. We have analyzed the predicted energetical effect of mutations on PPIs by applying a variety of computing methods to model the mutation and compute the change in binding affinity (FoldX, mCSM, pyDock combined to SCWRL3). We validate the predictive energetical impact through experimental mutations contained in SKEMPI 2.0 and apply these approaches in pathological and neutral single amino acid variants (SAVs) afterward (from ClinVar/Humsavar and gnomAD). Based on this, we have identified pathological mutations that clearly affect the analyzed interactions by stabilizing or destabilizing them.

As discussed above, protein-protein interactions are important for biological processes and pathological situations and are attractive targets for drug discovery. However, rational drug design targeting protein-protein interactions is still highly challenging. Hot-spot residues are seen as the best option to target such interactions, but their identification requires detailed structural and energetic characterization, which is only available for a tiny fraction of protein interactions. This thesis covers a variety of computational methods that

have been reported for the energetic analysis of protein-protein interfaces in search of hot-spots, and the structural modeling of protein-protein complexes by docking. This can help to rationalize the discovery of small-molecule inhibitors of protein-protein interfaces of therapeutic interest. Computational analysis and docking can help to locate the interface, molecular dynamics can be used to find suitable cavities, and hot-spot predictions can focus the search for inhibitors of protein-protein interactions. A major difficulty for applying rational drug design methods to protein-protein interactions is that in the majority of cases the complex structure is not available. Fortunately, computational docking can complement experimental data. An interesting aspect to explore in the future is the integration of these strategies for targeting PPIs with large-scale mutational analysis.

Contents

List of tables

1. Introduction

2. Objectives

3. Integrative modeling of macromolecular assemblies: validation and new challenges

4. Computational analysis of protein-protein interactions affected by pathological mutations

5. Docking-based identification of suitable transient cavities for the modulation of protein-protein interactions in drug discovery

6. General discussion

7. General conclusions

8. Appendices

List of figures

1. Introduction

2. Objectives

3. Integrative modeling of macromolecular assemblies: validation and new challenges

4. Computational analysis of protein-protein interactions affected by pathological mutations

5. Docking-based identification of suitable transient cavities for the modulation of protein-protein interactions in drug discovery

6. General discussion

7. General conclusions

8. Appendices

Abbreviations

ABS:	Average Buried Surface
ADPs:	Atomic Desolvation Parameters
AND:	Deoxyribonucleic acid
ANM:	Anisotropic Network Model
RNA:	Ribonucleic acid
A:	Adenine
aa:	Amino acid
Å:	Angstrom
AIDS:	Acquired Immune Deficiency Syndrome
ALS:	Amyotrophic Lateral Sclerosis
ASA:	Accessible Surface Area
BND:	Bound
BM4:	Protein-protein docking benchmark 4
C:	Cytosine
CAPRI:	Critical Assessment of Predicted Interactions
CASP:	Computational Assessment of techniques for protein Structure Prediction
COV:	Coverage
D:	Domain
DNA:	Deoxyribonucleic acid
e.g.:	For example
etc:	Etcetera
EM:	Electron Microscopy
FDA:	Food and Drug Administration
FFT:	Fast Fourier Transform
G:	Guanine
GPCR:	G protein-coupled receptors
HBB	Hemoglobin subunit beta
HBZ	Hemoglobin subunit zeta
HBA	Hemoglobin subunit alpha
HIV	Human immunodeficiency virus
HPV	Human papillomavirus
HP	Haptoglobin
HS:	Hot-spot
i.e.:	*id est* (that is)
K:	Kelvin
Kcal:	Kilocalory
LD:	Langevin Dynamics
MD:	Molecular dynamics
MM:	Molecular Mechanics
mRNA:	messenger RNA
MSA:	Multiple Sequence Alignment
NGS:	Next-generation sequencing
NIP:	Normalized Interface Propensity
mM:	milimolar
NaCl:	Sodium chloride
NMR:	Nuclear Magnetic Resonance
ns:	Nanoseconds
PCC:	Pearson Correlation Coefficient
PDB:	Protein Data Bank
PPI:	Protein-protein interaction
PPV:	Positive Predicted Value
ps:	picoseconds
RMSD:	Root-Mean-Square Deviation
SAVs:	Single Amino acid Variants
SAXS:	Small Angle X-ray Scattering
SI:	Sequence Identity
SNVs:	Single Nucleotide Variants
T:	Thymine
tRNA:	transfer RNA
U:	Uracil
UNB:	Unbound
WT:	Wild Type
3D:	Three-dimensional
ΔΔGbind:	Binding affinity changes

1. INTRODUCTION

Part of this section has been published in *Adv Protein Chem Struct Biol.* [1], *Exp Opin Drug Discov.* [2], *Encyclopedia of Life Sciences* [3] and *Methods Mol Biol* [4].

1.1. Genetic basis of disease

1.1.1. From genes to proteins

In living organisms, genetic information is stored, replicated, and functionally expressed thanks to essential polymeric biomolecules such as nucleic acids (formed by nucleotides) and proteins (formed by amino acids). The most simple unit of genetic information is formed by deoxyribonucleic acid (DNA). The complete set of genetic information in a living organism defines the so-called genome, and this information is present in each cell of the organism. Genome is formed by genes and non-coding genetic information located between, inside and outer of genes. Genes, from the biochemical point of view, is DNA that encodes the information required to synthesize a functional biological product. The protein is the predominant functional biological product, but also it is maintained in specific classes of ribonucleic acid (RNA). All biological processes are managed by these functional biological products, as well as the maintenance, storage, and metabolism of itself. Then, focusing on DNA, it is a polymeric molecule formed by deoxyribonucleotides, composed of one nucleobase, which in living organisms is mostly either a purine-derived base: adenine (A) and guanine (G), or a pyrimidine-derived one: cytosine (C) and thymine (T), plus one deoxyribose (sugar), and a phosphate group. Two complementary DNA strands adopt a highly stable double helix conformation. Such complementarity of two DNA strands is due to specific base-pair chemical interactions: between A and T, with two hydrogen bonds, and between G and C, with three hydrogen bonds. The phosphate groups constitute the backbone of DNA, by joining the nucleotides and also giving directionality to the polymeric chain. The oxygen of phosphate is covalently bound to the C5' position of the sugar. Then, the phosphate group joins the following nucleotide in C3' sugar's position giving a directionality 5' to 3'. The distribution of the four nucleotides found in DNA in the virtual totality of living organisms is conserved along with the species through evolution [5]. DNA provides substantial information for replicating, transcribing, and translating the organism's genetic data. The exonic regions of DNA contain all the information needed to produce proteins [6]. These regions contain a succession of triplets of nucleotides (called codons), and their combination codes up to 20 types of amino acids. The number of amino acids composing a protein is variable, and it depends on the nucleotide sequence of the DNA that codes for the specific protein. The starting triplet for synthesizing protein sequences is ATG; this is the unique codon for the methionine amino acid. On the other hand, there are three stop codons assigned for finalizing the synthesis:

TAA, TAG, and TGA (Figure 1.1). Ribonucleic acid (RNA) is formed by ribonucleotides, which are composed of a nucleobase, which in living organisms is mostly either a purine-derived base: adenine (A) and guanine (G), or a pyrimidine-derived one: cytosine (C) and uracil (U), plus one ribose (sugar), and a phosphate group. RNA is crucial for gene transcription and translation. DNA transcription consists of the biosynthesis of a messenger RNA (mRNA) molecule that is complementary to the DNA sequence, considering that G and C bases are complementary, as well as A and U ones. Thus, mRNA contains the information to translate the gene into a protein in the ribosome. For this, a specific transfer RNA (tRNA) that pairs a specific codon with the corresponding amino acid carries such amino acid to the ribosome, a multi-molecular complex formed by proteins and RNA, where it is incorporated into a polypeptidic chain according to the mRNA sequence.

Amino acids are thus the basic element of proteins, which contain amine (-NH_2) and carboxyl (-*COOH*) groups, along with a side-chain (*R* group), covalently bound to the α-carbon. The α-carbon substituents in the amino acids found in natural proteins adopt L stereoisomery. In natural proteins, 20 different α-L-amino acids are found (Figure 1.1) [7]. The specific side-chain of each amino acid determines its physicochemical properties, such as acidity, polarity, or volume. Amino acids form a polypeptidic chain by covalent binding the carboxyl group from a given amino acid to the amine group from the following one. The chain of amino acids, so-called a polypeptide, has a directionality N'-terminal (N'-term) to C'-terminal (C'-term) according to their concatenation process. A polypeptidic chain formed by a small number of amino acids (up to ~20-25 aa) is generally called a peptide, while a longer chain is called a protein. The length of the peptidic sequence is determinant for their function and structure. Proteins can adopt a large variety of conformations, very often with a specific three-dimensional (3D) structure, for which the distribution of the amino acids along the sequence is critical. The physicochemical properties of the amino acid side-chains enable them to form disulfide bridges, salt bridges, π-stacking, hydrogen bonds, ionic and hydrophobic interactions, as well as van der Waals attracting/repulsive interactions in the folded chain. All these interactions (non-covalent bonds, except the disulfide bridges) contribute energetically to the stabilization of the protein 3D structure.

The primary structure of proteins consists of a flexible thread formed by the polypeptidic chain. According to the physicochemical properties "conferred" by the specific distribution of amino acids, this sequence folds in a repeating arrangement (so-

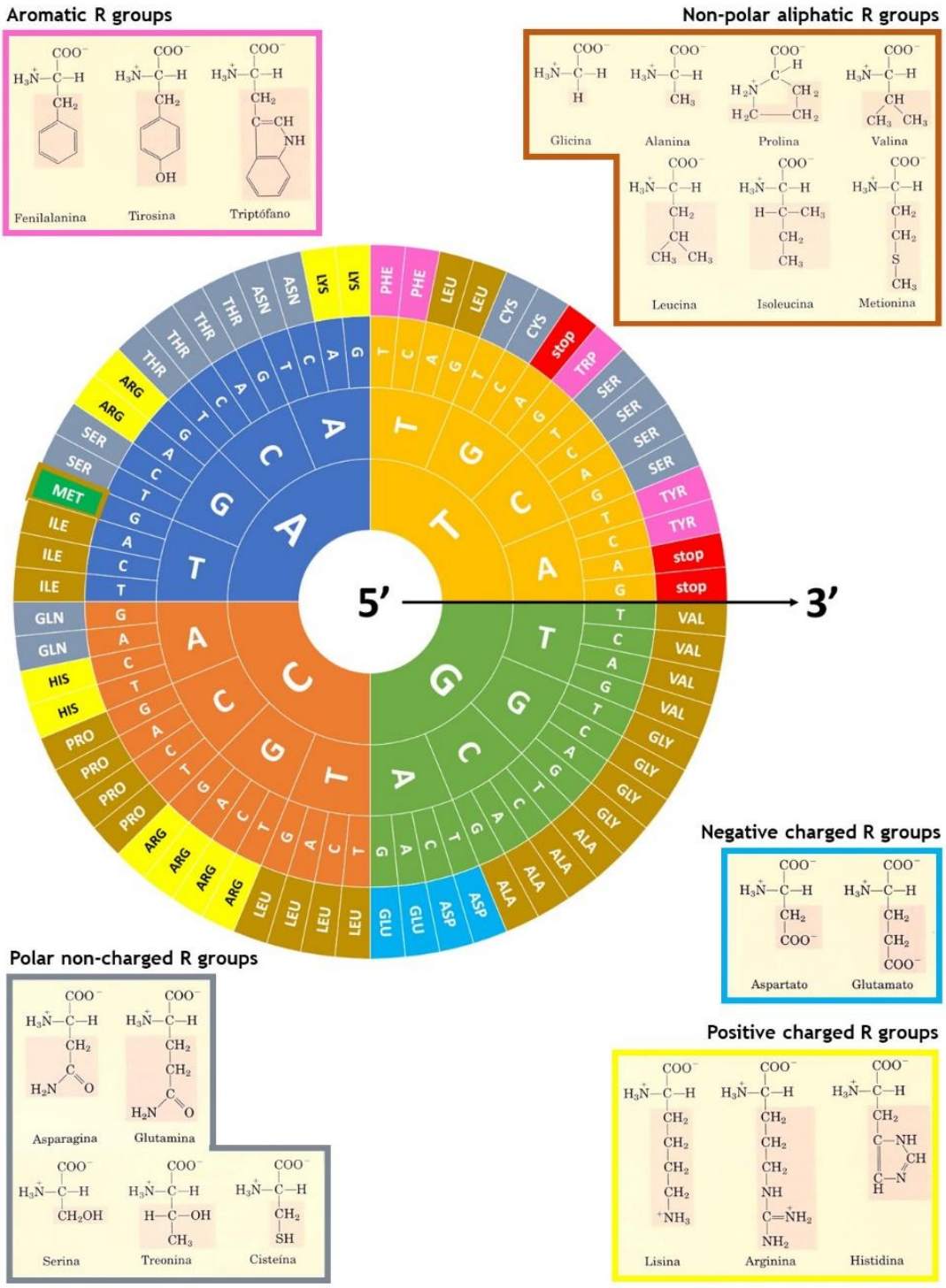


Figure 1.1 The formation of amino acids is based on the combination of nucleotides grouped in codons.
In green, the Methionine, the unique codon for the starting of transcription. In red, the three stop codons for the ending of transcription. The codons code to the formation of specific amino acid. The physicochemical properties of amino acids are based on the assigned R-group (adapted from [7]).

called secondary structure) of minimal 3D structure units. These minimal and specific 3D structures are β-sheet, α-helix, and random-coils. Last one, forming also connecting loops that do not fold into a specific structure. Then, the tertiary structure of proteins is defined as the folding of diverse secondary structures contained in the polypeptidic chain. The assembly of different polypeptide chains in complexes forms the quaternary structure. Therefore, every coded and expressed protein has one or several specific functions and can act as an individual entity or as part of oligomeric assemblies. Proteins rarely act alone, and they usually interact with other proteins and biomolecules to form intricate interaction networks that determine or modulate their function. Thus, protein-protein interactions (PPIs) also play essential roles in biological systems [8]. The complete set of interactions between proteins in a living organism defines the so-called interactome [9], whose description is critical to understand the behavior of the entire biological system, especially when is integrated with other data generated from the "omics" sciences (genomics, proteomics, transcriptomics, etc.).

1.1.2. Genetic variants in health and disease

Genetics plays a role in the majority of diseases, with multiple gene mutations determining the onset and severity of the disease, or modulating the risk of developing it. In recent years, the use of next-generation sequencing (NGS) technologies [10] has boosted human genomics research, and the efforts of large worldwide projects (The 1000 Genomes Project Consortium 2010; The 100,000 Genomes Project, Genomics England; Pan-Cancer Analysis of Whole Genome, International Cancer Genome Consortium) are helping to identify many genes and variants associated to disease. Besides, there are increasing efforts in the scientific community to aggregate and harmonize exome and genome sequencing data from a variety of large-scale sequencing projects, as is the case of gnomAD (Genome Aggregation Database) (http://gnomad.broadinstitute.org/). Thousands of disease-related variants are annotated in publicly available databases, such as UniProt Humsavar (https://www.uniprot.org/docs/humsavar), HGMD [11], OMIM [12], ClinVar [13], which have been useful to understand disease mechanisms, devise biomarkers for disease risk prediction and diagnosis [14], and facilitate the application of personalized therapies [15]. Now the challenge is the functional interpretation of variants and the prediction of their effects, with the ultimate goal of acquiring a better knowledge of disease at the molecular level.

The majority of genetic variants found in the human population are single-nucleotide variants (SNVs), caused by the alteration of a single nucleotide in the DNA sequence. Around 58% of the ~13,000 exomic SNVs carried per person are estimated to be missense SNVs or in-frame indels, which lead to amino acid changes in the translated protein [16]. Such single amino acid variants (SAVs) can alter protein structure and function, which may have an impact at the phenotypic level (e.g., predisposing to or causing a given disease). Overall, there are over 100,000 known genetic variants associated with different disorders, but functional information is missing for the majority of them [17]. The recent advances in genetic edition with CRISPR (clustered regularly interspaced short palindromic repeats) systems can facilitate the characterization of mutations at the phenotypic level [18], but a deep understanding of the impact of such mutations at the molecular level and its relationship to disease will be needed to improve prevention, diagnosis, and treatment of pathological conditions [19].

1.2. Functional interpretation of disease-related mutations at the molecular level

Since the same protein and its protein-protein interaction network has different phenotypic outcomes according to the specific effects of disease mutations [20]. Then, there is a need to understand the structural and energetic effects of these mutations at the molecular level.

1.2.1. Impact of mutations on protein structure and function

Missense mutations can have a direct effect on protein function, by altering active sites or binding interfaces, but can also have other structural effects, such as affecting protein folding and/or stability, inducing conformational rearrangements, or altering transport and location, which in turn may also have indirect effects on function.

As mentioned above, proteins usually interact with other proteins and biomolecules to form intricate interaction networks that determine or modulate their function. They can form more or less stable multimeric complexes involved in a variety of functions, such as the ribosome, nuclear pore, spliceosome, etcetera, or they can form specific and dynamic interaction networks like in cell signalling or metabolic pathways [21].

1.2.2. Protein-protein interaction networks

The complexity of protein-protein interaction (PPI) networks is highly related to the dynamics of the cell. The dynamic activity in a cell is transcribed as functional modules, and these modules are defined as groups of different proteins that interact but that are not necessarily present at the same time and space [22, 23]. These functional modules can be detected by clustering the groups of nodes in the network according to their connectivity. Therefore, based on the information about the genes, disease disorders, and proteins altered by known pathological mutations, human disease networks can be defined as graphs in which a disease node is linked with a gene node containing a mutation associated to the disease, which in turn is connected to other gene-based on the interactions of the encoded proteins [24].

Many efforts have been focused on the annotation PPI networks in order to understand better and accurately the biological processes. The result of such efforts are rendered in PPI networks such as Interactome3D [25] where the human protein-protein interaction data is generated by integrating information available from nine major public PPI databases: Intact [26], MINT [27], DIP [28], MPIDB [29], MatrixDb [30], InnateDb [31], BioGRID [32], BIND [33] and HPRD [34]. Another remarkable PPI network database is STRING [35]. In addition to this, there is an increasing interest in extending this information in disease-related networks. Representative examples are the altered PPIs networks by RASophaties [20] or dSysMap (Interactome3D-based integrating disease missense mutations) [36]. On the other hand, the most ambitious project is Reactome, which consists in the extension of a classic metabolic map, based in a molecular pathway networks database that integrates PPIs, metabolic reactions, disease pathways and drug-response data, as well as literature-Orcid data [37].

1.2.3. Disease-related mutations can affect protein interaction networks

Protein-protein interactions are involved in the majority of diseases, and thus are key to understand, prevent and correct pathological situations. In many cases, the perturbation of a given protein-protein interaction due to the environment, genetics or other reasons is the cause of the disease. Therefore, the role of protein interaction networks in complex diseases is increasingly evident, and new concepts in biomedicine are appearing such as network medicine or endophenotype networks [38]. Functional genomics studies have helped to

understand the involvement of disease-associated mutations in protein-protein interactions (PPIs). In other cases, perturbed PPIs might not be the first cause of a pathological situation, but they could affect the behavior of a given network or metabolic pathway and thus explain the phenotype.

Indeed, recent functional profiling studies showed that the majority of analyzed disease-related mutations were perturbing PPIs: half of them showing complete loss of interactions, perhaps by global conformation effects, and the other half showing "edgetic" effects, that is, that affected only a specific subset of interactions [17] (Figure 1.2). This study shows how the effect of pathological mutations on the entire disease network can differ depending on the amino acid change and its location in the protein. A strong change in protein folding or stability will likely affect all the interactions of the mutated protein. In this context, pleiotropic effects, which are mutations in the same gene that show different disease phenotypes [17, 39], could be explained because such mutations affect binding to different proteins. Understanding the specific impact of a protein mutation on the protein interaction networks involved in different biological processes (e.g., signaling, metabolism,

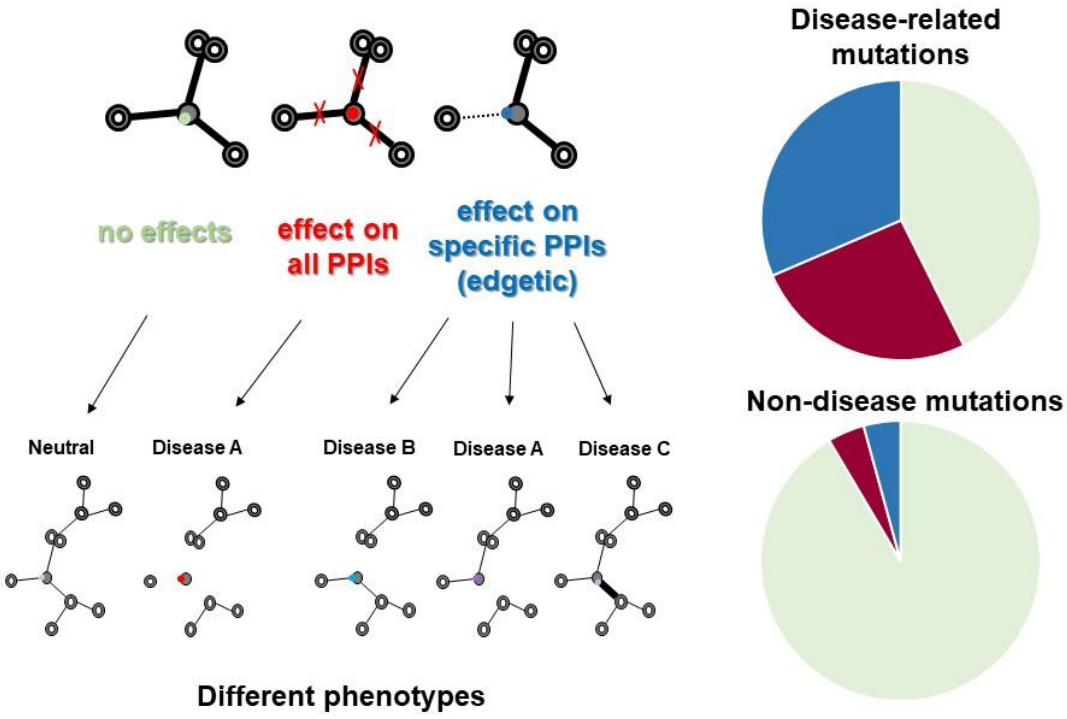


Figure 1.2 Possible effects of disease-related mutations on protein-protein interactions (PPIs).
Pie charts show the results from a recent experimental study on 197 disease-related mutations and 47 non-disease variants [17], showing no effects on PPIs (green), dramatic effect on all PPIs (red) and effect on specific PPIs (blue).

ribosome, transcription factors, etc.) would be key in bridging the gap between genetic information and phenotypic effects. Hence, the effect of a given pathological mutation on conformational variability, post-translational events, or splicing will be important in order to fully understand its role on the PPI networks.

However, it would be impractical and highly expensive to perform this type of experimental functional profiling at the genomic scale, with the added limitation that current tests as the above described, do not cover all possible interactions for a given protein. In this context, knowing the structural details at atomic resolution of a protein and all the interactions in which such protein is involved, it would be important in order to interpret the functional impact of a mutation on the protein interaction networks.

1.3. Structural characterization of mutations in protein-protein interactions

Disease-related mutations of residues directly located at a protein-protein interface can induce structural and physicochemical changes that may alter its stability, modify specific intermolecular atom-atom interactions, or even have an effect on the interface conformational dynamics. Recent statistical analyses on protein-protein complex structures show that disease-causing SAVs are enriched in the protein-protein interface residues as compared to the non-interacting surface residues [40]. Moreover, within the interface regions, disease-causing mutations are more likely to be located at the solvent-inaccessible central part (‘core’) rather than at the partially solvent-accessible one (‘rim’) (Figure 1.3), contrary to what is observed for neutral polymorphisms [41]. Thus, structural characterization of SAVs involved in protein-protein complexes can help to interpret their effect on specific interaction networks and pathways and close the gap between genotype and phenotype.

The 3D structures of thousands of protein-protein complexes are available at the Protein Data Bank (PDB) (www.pdb.org) and information is compiled in useful databases, such as we have introduced in the previous section involving PPI networks. It is the case of Interactome3D (https://interactome3d.irbbarcelona.org/) [25], with structural data for binary protein-protein complexes from several organisms, STRING (https://string-db.org/) [42], which includes predicted interactions based on computational methods from co-evolution, text mining, co-expression or gene-neighborhood data, as well as experimental

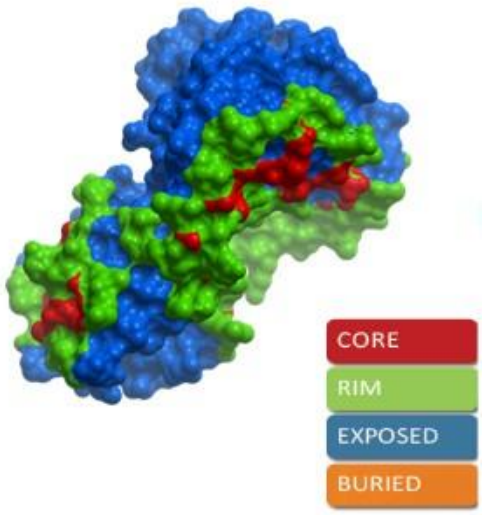


Figure 1.3 Classification of ETFA residues as buried, exposed or interface (core and rim) based on the 3D structure of its complex with ETFB (PDB 1EFV).

biophysical and structural data and links to biological pathways, or dSysMap (https://dsysmap.irbbarcelona.org/) [36], with the specific location of disease-related mutations on known protein-protein interfaces. This type of data facilitates the large-scale characterization of disease-related mutations at the molecular level. Despite this, in the majority of cases the structural coverage of the interaction network for the mutated protein is incomplete since only a tiny portion of the existing complexes have available structure. In human, there are around 20000 human proteins [43] (14000 in Interactome3D, 2019_1 version). The total number of PPIs in human is not exactly known, with some estimates ranging from 130,000 [44] to 650,000 [45] interactions. However, there is an available 3D structure for only less than 7,000 interactions (Interactome3D, 2019_1 version). Currently, 89% of PDB structures have been solved by X-ray crystallography, as compared with 8% that were solved by NMR Spectroscopy and 2.5% by Electron Microscopy (EM). Releases in 2019 indicate a significant increment of use of EM in 12%, while the other main methods decrease by 84% in X-ray and 3.5% in NMR. This suggests an advance of techniques solving multimeric complexes. Therefore, structural data alone cannot give a complete explanation for the majority of mutations regarding their impact on protein interaction networks.

1.4. Computational modeling of protein-protein interactions

Computational methods are increasingly used to complement existing structural data on proteins and other biomolecules, including protein complexes. Computational docking can be applied for modeling protein-protein complexes with no available structure, and different approaches and protocols have been reported, as detailed below (Figure 1.4).

1.4.1. Template-based docking

A protein-protein complex can be modeled based on the available 3D structure of complexes between proteins that are homologous to the ones of interest [46]. With this strategy, called template-based docking, there are quite reliable models for around 5,000 human protein-protein complexes (Interactome3D, 2019_1 version). However, this template-based modeling approach has limited applicability, since for the majority of interactions only templates with remote homology can be found, which dramatically decreases the predictive success [47]. In addition, the oligomeric state of the individual proteins or even that of the complex is not always conserved among homologs, which can introduce significant error in the modeled complexes.

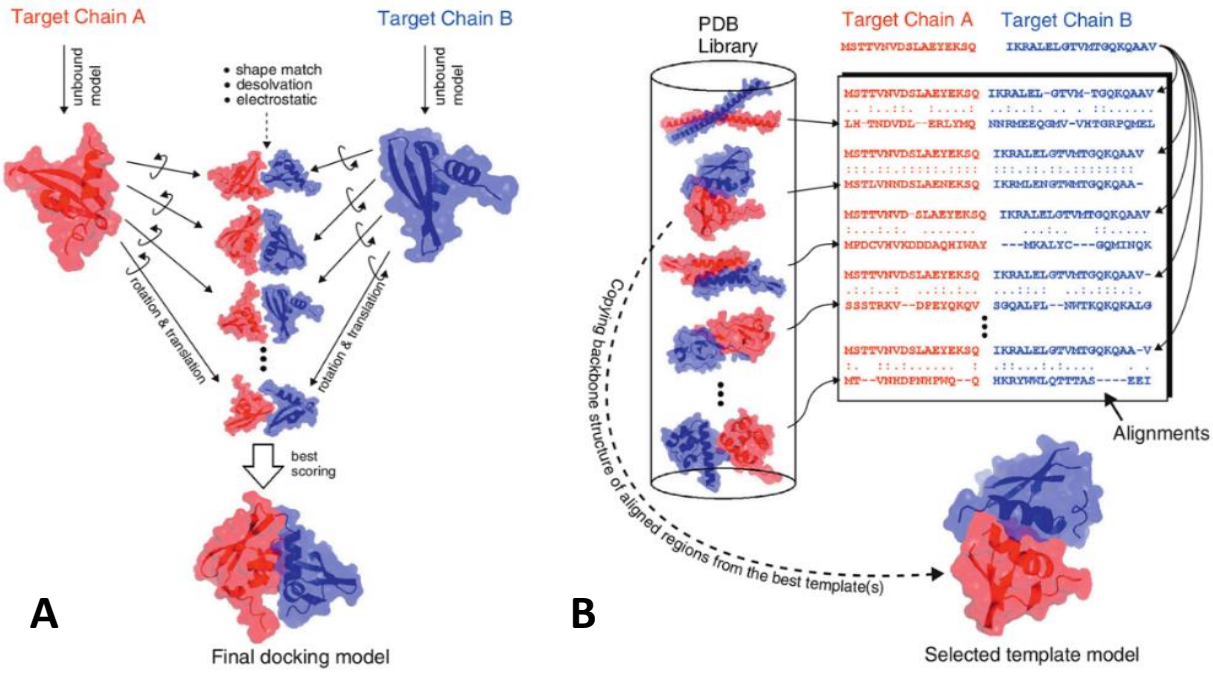


Figure 1.4 The two main methods existing in computational modeling of protein-protein interactions [46].

With the increasing availability of complex structures, in recent years, the attention is focused on template-based structural modeling of complexes, based on the standard principles of homology-based modeling. The term template-based docking is specifically used when a model is built by superimposing the structures (or models) of the unbound subunits onto the corresponding subunits of a template complex structure [46]. One advantage is that template-based modeling can be applied to multi-molecular complexes, not just binary complexes as *ab initio* docking. In addition, it has been suggested that templates are available for the large majority of cases in which interacting subunits have structural information [48]. However, the general availability of good-quality templates that could be reliably used for template-based predictions seems much lower [49]. Actually, for the majority of known interactions, only templates with remote homology are available [44], for which direct application of template-based methods leads to poor predictions [47].

For modeling protein-protein complexes when the unbound structures are available, despite knowing good-quality templates, the superposition method is critical for predicting successful protein-protein interactions. Multiple sequence alignment (MSA) methods such as Clustal Omega [50], T-coffee [51] or MUSCLE [52] sometimes are not enough to assign an accurate alignment, and there is the need to include structural alignment methods to achieve better accuracy in the alignments. Some examples of structural alignment methods are: DALI [53], STAMP [54], TM-Align [55], SuperPose [56] or VAST[57]. On the other hand side, for unknown unbound structures, besides of an accurate alignment, the modeling step also is crucial to obtain successful predictions of PPI complexes. Most of the modeling tools are sequence alignment dependent, and then it is important to get a good accurate alignment before modeling. Several methods are available for predicting protein structures and multi-protein structures. It is the case of I-TASSER [58] and HHpred [59] for a unique protein or MODELLER [60] or RosettaDock [61, 62] for multi-protein modeling structures.

1.4.2. *Ab initio* docking

For interactions with no available template, *ab initio* computational docking methods aim to provide structural models for PPIs based on the structure of their unbound components. *Ab initio* computational docking can provide acceptable models within the top 10 predictions in up to 40% of the cases, according to reported evaluation studies of different methodologies in current protein-protein docking benchmark version 5.0 [49, 63, 64].

Methodologically, two major aspects are considered in virtually all docking protocols: a sampling procedure, to search for different binding orientations, and a scoring procedure, which evaluates these decoys in order to identify the correct binding mode. Major sampling strategies are shape matching, exhaustive global search, and stochastic sampling. Shape matching methods include DOCK [65] or Patchdock [66], where binding regions are represented by a graph generated from geometric features, such as convex, flat and concave regions. Exhaustive global search is based on innovative Fast Fourier Transform (FFT) algorithms [67], which use a grid-based 3D representation of the proteins that facilitates the identification of the best rigid-body docking orientations at a low computational cost. Some of the most popular methods are FTDock [68], ZDOCK [69] or MolFit [70]. The method HEX [71] and later FRODOCK [72] used polar Fourier correlations to accelerate docking calculations. Other different approaches using stochastic search based on global-energy optimization are ICM-DISCO [73, 74], RosettaDock [75], HADDOCK [76], SwarmDock [77], or LightDock [78]. These energy-based methods are still valid for complexes formed by rigid proteins, and have the advantage that they can include conformational flexibility through molecular mechanics (MM). After generating many different docking decoys by the above-described methods, it is important to evaluate all these docking poses and identify the correct binding modes. For example, InterEvScore is an interesting scoring approach based inter-molecular interface contacts based in the inclusion of evolutionary constraints [79]. In this context, scoring plays an essential role, and different strategies have been reported for this, based on statistical potentials, empirical functions, or energy-based description, the latter usually including van der Waals, desolvation or electrostatics terms. Several scoring algorithms that can be independently applied to previously generated docking models are available, such as pyDock [80], ZRANK [81], or SIPPER [82]. In this regard, both FFT- and energy-based sampling approaches can be combined in the pyDock protocol [80, 83], with very efficient energy-based scoring (Figure 1.4-A). This protocol is publicly available at a web server (http://life.bsc.es/servlet/pydock) [83]. Many methods, in addition to scoring, introduce flexible refinement of the docking models, such as FireDock [84], ICM-DISCO [74], HADDOCK [76] or RossetaDock [75].

1.4.3. Assessment of protein-protein interaction predictions

Computational docking methods are being continuously evaluated at the Critical Assessment of PRedicted Interactions (CAPRI) blind assessment [85]. This CAPRI community experiment (http://www.ebi.ac.uk/msd-srv/capri/), has boosted the docking field by providing new challenges, validation tools, and motivating discussions for the last 20 years. In recent years, CAPRI has also been integrated into CASP (Computational Assessment of Techniques for Protein Structure Prediction) rounds focused on multi-molecular assemblies [86], since in many protein structure prediction targets the oligomerization state is highly relevant. Hence, the recent CASP-CAPRI joint rounds imposed an additional difficulty level in the field of prediction of protein interactions, with a number of multimeric targets in which the docking approaches were fundamental to model many of the oligomeric interfaces. From the most recent results in CAPRI, successful methods in docking derived from efficient integration of experimental and evolutionary information rather than from real advances in automatic modeling itself. Indeed, remaining unsolved challenges in automatic docking are: interactions involving flexible proteins, weak complexes and multi-molecular complexes, for which new developments are needed.

1.4.4. Docking-based tools for the prediction of protein-protein interfaces

The use of energy-based functions in docking can help to describe the energetic aspects of protein-protein association [87, 88]. A large variety of statistical potentials, empirical functions and energy-based algorithms can be applied to computationally characterize a protein-protein complex structure. Many of these functions are compiled and made publicly available in the database CCharPPI [89]. They can be used to analyze surface patches, estimate binding affinities, or identifying those residues that contribute the most to the binding affinity (so-called “hot-spot” residues) [90, 91]. In this context, a variety of computational methods have been reported for predicting the protein-binding regions at the surface of the unbound protein structures. One popular tool is ProMate, which uses a combination of biophysical properties [92]. Other interface prediction methods are PINUP [93], which is also based on a combination of empirical energy functions, or cons-PPISP [94], a neural network predictor that uses sequence profiles and solvent accessibilities of each residue and its spatial neighbors. It is important to remark that all these methods are

applied to the structure of one of the unbound proteins and therefore, can only predict general interface residues, which are not necessarily specific for one particular partner protein. In the case of a protein interacting with other different proteins using different interfaces, such predicted interface residues are not useful to distinguish these specific interfaces.

In addition to docking prediction, pyDock provides a variety of additional modules to analyze fundamental problems in biomolecular recognition. The main advantages of this method are that it does not require the structure of a complex and that hot-spot predictions are specific for two given interacting proteins (as opposed to general interface prediction on one unbound protein as above described). The module pyDockNIP, which analyzes the frequency of interface residues in low-energy docking models from pyDock [95], has been reportedly applied to identify interface hot-spot residues [96], which can be relevant for drug discovery targeting protein-protein interactions with small molecules. The module pyDockSAXS is the first systematically tested approach in using protein docking models to complement low-resolution structural data from Small Angle X-ray Scattering (SAXS) [97-99]. On the other side, interface residue data from bioinformatics predictions, mutational experiments, NMR, cross-linking, etc., can be included as distance restraints with pyDockRST module [100]. Although originally aimed to protein-protein docking, pyDock can also be applied to model protein interactions with other biomolecules, such as protein-RNA [101]. On a more practical side, the pyDock methodology has provided biological insights and helped to interpret experiments in different cases of biomedical and biotechnological interest. One remarkable example is the structural study of host-pathogen complexes, in which pyDock docking together with energetic analysis helped to Another case of interest is the application of pyDock within a broad structural analysis of members of the family of Hetero Amino acid Transporters (HAT), such as the integrative modeling of the assembly of transmembrane LAT2 and its ancillary protein 4F2hc, using a combination of modeling, docking, electron microscopy and cross-linking experiments [102].

1.4.5. Current challenges predicting protein-protein interfaces

The application of computational docking to large-scale modeling of the interactome currently has some limitations. The main problem is the low predictive success of current docking tools, which would require to generate many potential models for each interaction,

with high interpret molecular mechanisms for uncertainty in the identification of the correct orientations in weak and flexible interactions [103]. But there are also several additional practical problems. One of the problems is the structural coverage of the interacting proteins. Indeed, there is a complete structure for only 13% of the proteins involved in the known human interactome (data from Interactome3D, 2019_1 version), while an additional 13% of them can be modeled by homology. Another 40% of the proteins have available structure or model just for a part of the sequence, in which case, it would be important to identify which are the interacting domains, to confirm whether they have available structure for docking. However, most of the available protein interaction databases provide sets of binary interactions, and there is no enough information on the domains that are directly involved in the interaction. Another potential problem for docking is the oligomerization state of the interacting proteins, as well as that of the protein complex. Lack of inclusion of this important information in the docking calculations might induce to incorrect binding modes or wrong interpretation of results. In many proteins, there is no experimental evidence of their biological oligomerization state, and this information is taken from the biological unit in the x-ray crystal structure deposited in the PDB, based on PISA scoring [104]. However, there are cases in which biological interfaces are difficult to distinguish from crystal packing interfaces. The web server Eppic (http://www.eppic-web.org/) [105] evaluates the pairwise interfaces in a protein crystal based on multiple sequence alignments (MSA) of closely related homologs and predicts the likely quaternary structure of the protein. Other computational resources to evaluate the oligomerization interfaces in a crystal structure are ProtCiD database (http://dunbrack2.fccc.edu/ProtCiD) [106], QSbio (http://www.QSbio.org) [107], or PPI3D web server (http://bioinformatics.ibt.lt/ppi3d) [108]. For modeled proteins, the oligomerization state can be inferred from the templates used for modeling, which increases uncertainty.

1.5. Energetical characterization of mutations on protein-protein interactions

1.5.1. Identification of hot-spot interface residues

Structural knowledge or reasonable modeling of the interface might not be sufficient to understand the role of a given mutation regarding protein-protein interactions. Usually,

only a small subset of the interface residues are energetically important for the interaction, the so-called hot-spots residues, which are usually defined as those ones in which their change in binding affinity upon mutation to alanine is larger than 2 kcal/mol [109]. Such hot-spot residues tend to be enriched in disease-causing mutations as compared with the rest of the interface [41]. Since hot-spot residues are found more frequently in the interface core [110], a simple geometrical analysis of protein-protein interfaces can suggest potential hot-spot residues, but this is not very precise. More detailed energy-based calculations on the complex structure for hot-spot predictions have been reported, e.g. FoldX energy function [111] or Robetta physical model [90]. Machine learning is also used in different hot-spot prediction approaches, such as PCRPi, which combines sequence conservation, energy score and contact number information [112], PPI-Pred, based on surface shape and electrostatics [113], or Pocket-Query, which provides an assortment of metrics useful for predicting hot-spots [114]. Other methods are HotSpot Wizard [115], based on the integration of structural, functional and evolutionary information provided by several databases; DrugScorePPI [116], derived from experimental alanine scanning results; iPred [117], using pairwise potential atom types and residue properties; or ECMIS [118], using a new algorithm combining energetic, evolutionary and structural features. These methods can be used to analyze interfaces at a large-scale. In this context, it is interesting to mention the PCRPi database (PCRPi-DB), which contains computationally annotated hot-spot residues in all protein-protein complexes for which a high-resolution 3D structure is known [119]. The above hot-spot prediction methods are based on the 3D structure of a protein-protein complex. However, the structural determination of all protein-protein complexes in human (structural interactome) remains one of the biggest challenges in structural biology. Indeed, there is available structural information for only a tiny fraction of all the protein-protein interactions that are estimated to occur in human [25]. In this context, computational methods can help to characterize a protein-protein interaction for which there is no structural information. In this context, when no complex structure is available, docking-based prediction of hot-spot residues with pyDock is an interesting alternative [96].

1.6. Computational prediction of binding affinity changes upon mutation

The effect of mutations at the local level can be better understood by analyzing the structural information of protein-protein interfaces in more detail. It is known that protein-binding sites are involving around 28 residues on average, forming mostly flat interfaces of around 1000 $Å^2$ on average [120], which are much larger than traditional protein-ligand sites [121]. Mutation of hot-spot residues are more likely to be related to disease [122, 123]. Recent studies show that hot-spot residues can also be relevant in homo-oligomeric interfaces [124]. Hot-spot residues at self-assembly play an essential role in binding, and their mutation could induce the formation of non-natural assemblies and cause disease [124, 125]. Then, structural data (either experimental or modeled) can help to identify disease-related mutations that are directly located at a protein-protein interface. When analyzing all the structural aspects in protein-protein interactions, we should not forget that proteins are not static entities. In addition to their intrinsic conformational variability in solution, very often they show different conformational states (as well as isoforms generated by alternative splicing, or post-translational modifications, such as methylation, phosphorylation, glycosylation, etc). Therefore, the definition of whether a given protein is active or inactive, or is engaged in a protein interaction or not, will depend on which of their different possible molecular states are considered. In this line, recent work has redefined the concept of protein-protein interaction networks, by including the conformational variability of the interacting proteins [126]. However, there are often situations in which knowing the structure of the complex is not sufficient to estimate the impact of a mutation. For example, mutations H27A and H180A in human prolactin (hPRL) interacting with its receptor (hPRLr) have been experimentally shown to have a completely different impact on the free energy of binding of these two proteins. While hPRL H180A has a strong impact on binding affinity (+2.45 kcal/mol), the hPRL H27A mutation has no significant impact (-0.07 kcal/mol) [127]. However, from the available structure of the hPRL/hPRLr complex (PDB 3MZG), we can observe that these two residues are both located at the protein-protein interface (Figure 1.5), which does not directly explain their different effects. Therefore, knowing that a given mutation is located at a protein-protein interface is not sufficient to describe their impact on the interaction, so further energetic characterization is needed. A variety of experimental studies have reported the binding affinity changes ($\Delta\Delta G_{bind}$) for many mutations, and most of these values have been

compiled in public databases, such as SKEMPI [87, 88]. The majority of the experimental $\Delta\Delta G_{bind}$ values that are reported in the literature correspond to destabilizing mutations (defined as those with $\Delta\Delta G_{bind}$ > 1.0 kcal/mol), and only a minority of them are stabilizing mutations (i.e. $\Delta\Delta G_{bind}$ < -1.0 kcal/mol). However, the amount of experimental data that is available for disease-related mutations is very small, and the experimental determination of binding affinity changes of a mutation is not suitable for large-scale characterization of disease-related mutations. In this context, computational methods are increasingly needed for the energetic characterization of mutations.

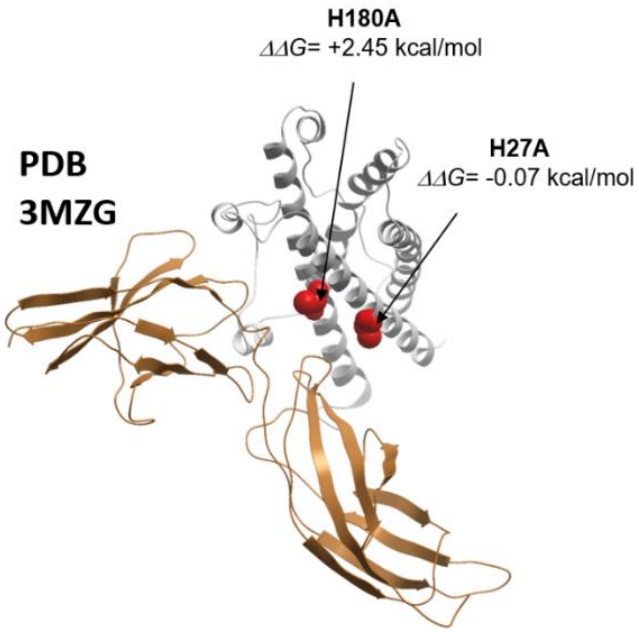


Figure 1.5 Different effects of SAVs located at protein-protein interface.
Complex structure of human prolactin (hPRL), represented in white ribbon, bound to its receptor (hPRLr), in gold ribbon (PDB 3MZG). In CPK are shown two interface mutations with different impact on binding free energy.

The energy-based scoring function in pyDock can also be applied to describe the contribution of residue side-chains to binding affinity in a protein-protein complex structure. This can be used as a fast estimator of the binding affinity changes upon mutation to alanine, as implemented in the pyDockEneRes web server (https://life.bsc.es/pid/pydockeneres) [128]. When the structure of the protein complex is available, the description of the energetic impact of mutations to residues other than alanine needs modeling of the mutated residues, which introduces more uncertainty in the predictions. Recent energy-based methods like FoldX or machine learning methods, such as mCSM [129] or iSEE [130]; or a combination of both like MutaBind [131] can provide

good predictive rates on specific sets of cases, but validation on external sets is strongly advised to have a more realistic estimation of their predictive capabilities.

An interesting aspect to explore is the integration of these strategies for targeting PPIs with large-scale mutational analysis. Characterization of the effect of a pathological mutation on the network of PPIs in human could help to identify potential targets for therapeutic intervention. Candidate targets for that would be protein-protein interactions whose binding energy are directly affected by the mutation. The majority of mutations affecting protein-protein interfaces are expected to be destabilizing, but a significant percentage of them (20% according to SKEMPI database) makes the interaction stronger. For instance, over-expression of interleukin-8 (IL8) has been associated with KRAS mutations in tumors, and thus the interaction of IL8 with their receptors CXCR1 or CXCR2 has been proposed as an attractive therapeutic target in cancer [132]. It could also happen that a protein-protein interaction acquires a new unwanted role due to homeostatic compensation of the network upon a given mutation or other pathological situations. As an example, IL-2 levels are raised in normal immune response, but unwanted high IL-2 levels are also found in pathological situations as in autoimmune disease or graft rejection in organ transplantation. In these situations, the otherwise normal interaction of IL-2 with IL-2R produces unwanted effects, and thus constitutes a known drug target for which several inhibitors have been identified (see Table 1.1) [133]. In all these situations, the modulation of a target protein-protein interaction by a small-molecule could restore normal function, and thus such a compound could have the potential to be further developed into a therapeutic drug.

1.7. Modulation of protein-protein interactions

Rational drug design has been the object of increased attention by the pharma industry, mostly propelled by accumulated expertise on biological systems, large-scale availability of sequence, structural and functional data, and better computer-based models. One of the most successful examples was the series of HIV protease inhibitors that were computationally designed and developed in the early 1990s, which provided the first effective clinical treatments against AIDS [134]. However, despite all the investments and the increased availability of computational resources and big data, the number of new molecular entities approved by Food and Drug Administration (FDA) every year is reaching a plateau [135]. The reasons for this are multiple, but one of the key factors is that

the majority of available small-molecule drugs target only three major protein families: GPCRs, ion channels and nuclear receptors [136]. In order to expand the target space, it is important to take into account above mentioned regarding how proteins act forming intricate networks of interactions, which determine the behavior of the biological system. Indeed, the drug discovery field is beginning to focus on new concepts such as systems pharmacology [136, 137] and network medicine [138]. In this context, computational methods for the characterization and identification of interface hot-spot residues can help to discover small-molecules that can modulate protein-protein interactions of pathological interest and make an effect on the system at the phenotypic level. Traditional drug discovery has focused on targeting individual proteins. However, the number of proteins that are druggable, i.e. potentially used for rational drug discovery, is limited. The majority of FDA-approved drugs target a few families of proteins, basically GPCRs, ion channels and nuclear receptors [136]. The traditional goal of targeting a single protein has strong limitations, such as unforeseen side-effects (because the protein altered by the drug may be involved in other different functions) or limited effect at system level (because of homeostatic compensation through other pathways). For these reasons, the pharmaceutical industry has a strong necessity of expanding the current target space, considering the proteins involved in a given disease (traditional targets) in the context of an interaction network [139]. Understanding in detail the behavior of a protein interaction network in pathological conditions is especially critical for complex diseases such as cancer or rare conditions related to the malfunction of the Ras signalling pathway (RASopathies) [20]. In addition to understanding the systemic effect of targeting one protein, a detailed description of a pathological network can help to identify specific protein-protein interactions as potential targets to be analyzed and modulated according to the interaction profiling.

Targeting protein-protein interactions with small-molecules is highly challenging. Protein-protein interfaces are usually larger than traditional protein-ligand interactions, and mostly flat, showing a large variety of topologies [140, 141]. For this reason, known inhibitors of protein-protein interactions do not show high similarity to traditional inhibitors of enzymes and receptors [121], and appear to have physicochemical properties that may violate traditional rules such as the Lipinski's Rule of Five [142].

Perhaps the first difficulty for rational design is to find a correct spot to be targeted by small-molecule. Contrary to traditional targeting of enzymes and membrane receptors, in which the known active site is used as a starting point for drug discovery, identifying

suitable cavities in protein-protein interfaces is extremely difficult since natural protein-protein complexes have not evolved to host cavities for small ligands. Fortunately, protein-protein contact surfaces are not rigid in solution, and conformational motions at local level (basically side-chains and loops) can induce the transient opening of cavities in the protein surface [121, 143]. These small perturbations have been detected by molecular dynamics simulations, providing a potentially suitable target surface for binding of small-molecules [121]. Provided that transient cavities can be generated by molecular dynamics, the question is, how can we identify those potentially useful cavities in the different conformations generated during the dynamics? There are many free available computational tools for the identification of cavities in the protein surface, like Fpocket [144], PASS [145] or QsiteFinder [146]. Most of these tools were developed to locate surface pockets involving active sites or natural substrate binding sites. Whether they can also be efficient for the identification of suitable pockets for PPI inhibitors within protein-protein interfaces is yet to be evaluated.

Once transient pockets can be generated by molecular dynamics simulations, and identified by cavity predictor methods, the problem is now to select which of such transient cavities would be the most suitable one for small-molecule positioning. The starting position for further docking or virtual ligand screening studies is critical, since the small-molecule inhibitor needs to bind at the optimal site in order to efficiently compete with a large protein. In addition, many small-molecule inhibitors of protein-protein interactions are still larger than traditional enzyme inhibitors, so in order to identify a suitable cavity for these cases, more than one predicted pocket might need to be merged [147] or combined for a fragment-based drug design strategy [148]. This would involve different computational problems, such as efficiently clustering the many detected pockets, or selecting a suitable threshold size for the pocket (small pockets might not cover the entire inhibitor, while large pockets would not be adequate for a small-molecule inhibitor).

Protein-protein interfaces that have already been successfully targeted by small-molecule inhibitors tend to have extended binding grooves that may be split into different sub-pockets [149, 150]. Very often, hot-spot residues are located nearby these pockets and are complimentary on both sides of the interface. In addition, small conformational changes in the binding site make a pocket deeper when bound to a small-molecule than to the partner protein [148]. Finally, PPIs with known small-molecule inhibitors tend to have small, high-affinity interfaces and include a hot segment that is essential for the binding to the partner

protein [150]. In addition, it has been reported the existence of druggable hot-spots at protein-protein interfaces that have a general tendency to bind organic compounds. These druggable hot-spots show conformational flexibility to allow them to accommodate a ligand of drug-like dimensions [151].

1.7.1. Difficulties in targeting protein-protein interactions

Thus, targeting hot-spot residues at protein-protein interfaces with small-molecules seems a reasonable strategy to disrupt protein-protein interactions. By targeting hot-spots, a small-molecule could break critical interactions and be able to compete with a large protein-protein interface. This has been extensively applied to the discovering of PPI inhibitors, in some cases yielding potent and selective compounds [122, 152-154]. The knowledge of interface and hot-spot residues based on the complex structures of several inhibitors can be used to localize the binding pocket [155]. In one interesting example, predicted hot-spots are used for the rational design of small-molecule compounds capable of blocking the IFN-α / receptor interaction. Hot-spots were predicted with iPred, then pharmacophore search was performed with VirtualLigand, and final docking of candidate compounds was done with Gold [156]. The use of hot-spots can be complemented by other computational approaches based on the use of peptides and protein fragments [153, 157], fragment docking and coevolutionary analysis [158], or ligand docking to locate protein-protein interfaces and potentially druggable sites [159].

Computational docking can help to model protein-protein complexes of therapeutic interest and predict hot-spot residues, which can be helpful for developing drug discovery programs aiming to target protein-protein interactions with no available structure (Figure 1.6). Because, a major problem is that all of the discussed strategies to identify binding cavities in protein-protein interfaces need the structure of the complex, which is not available in the majority of complexes. Thus, identifying binding sites in the unbound forms of the interacting proteins is a desirable goal, but much more challenging than on the *holo* structures, because in many cases these pockets may show large conformational rearrangement and thus remain mostly hidden in the *apo* state (so-called cryptic sites). Recent studies have found that these cryptic sites tend to be conserved in evolution and could be identified by using machine learning methods [160]. The combination of molecular dynamics and docking predictions could also help to identify these sites. A sensible strategy would be to apply molecular dynamics to the unbound form of the protein,

identify all transient pockets generated in the different conformers with methods like MDpocket [161] or Caver [162], then apply protein docking and hot-spot predictions [96] to locate the interface and the important residues to target, and finally select the most suitable transient pockets to start virtual screening and ligand docking procedures. Then, it would be essential to integrate structural characterization, dynamics and hot-spot analysis, before attempting to rational design small-molecule inhibitors that could target such interactions.

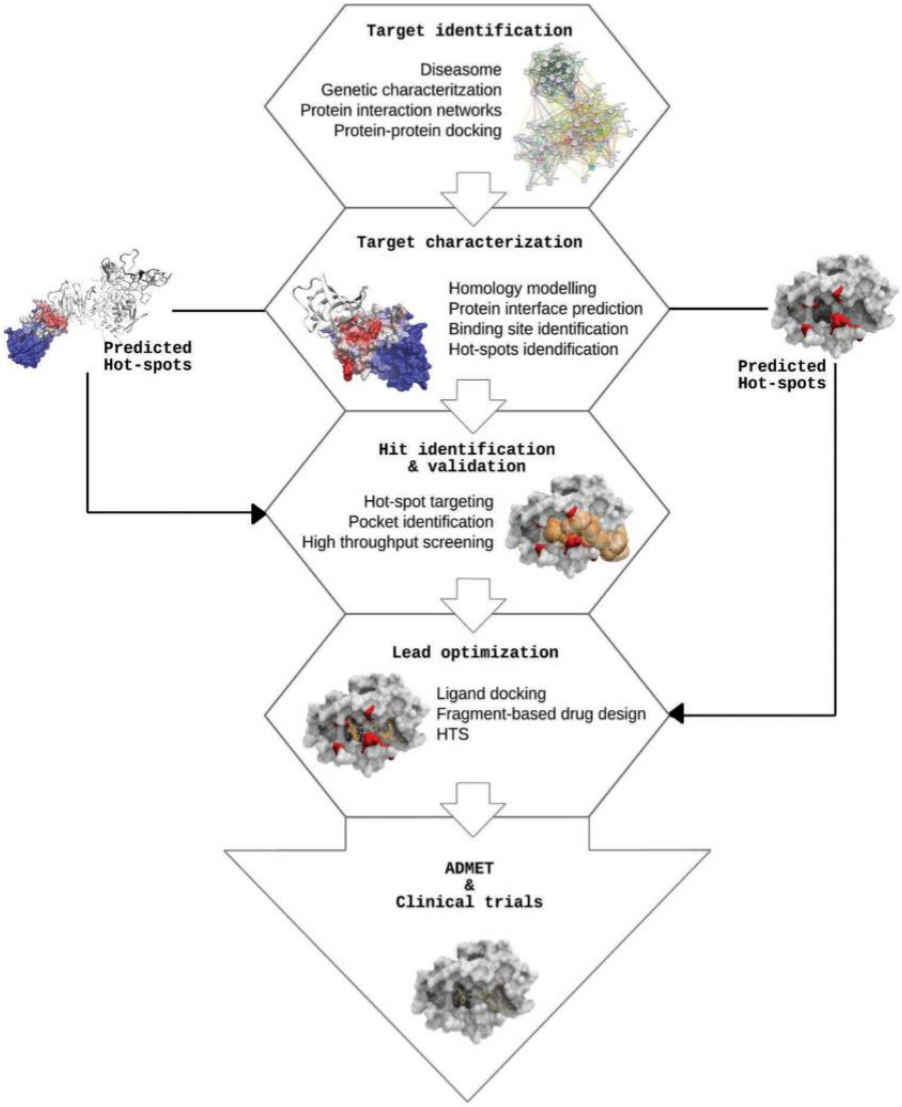


Figure 1.6 Computational approaches for rational drug discovery targeting protein-protein interactions. This shows a scheme of a general pipeline for drug discovery targeting protein-protein interactions, focusing on the different computational approaches that can help in each phase. An important part of target characterization is the identification of hot-spot residues in protein interfaces. Docking-based hot-spot prediction can help to locate binding cavities for hit identification, as well as in the process of lead optimization.

1.7.2. Small-molecules targeting PPIs of therapeutic interest

Modulating protein-protein interactions of therapeutic interest with small-molecules is a long-sought goal in drug discovery. Several examples of antibodies and peptides capable of inhibiting PPIs have been reported. Monoclonal antibodies can target the protein-binding surface of one of the proteins, and mimetic peptides can replace the interaction surface of one of the proteins. But the lower bioavailability of antibodies and peptides does not make them to be very attractive for therapeutic purposes. Small-molecules show many more advantages over large protein competitors, especially from a kinetic perspective [121]. Success in modulating PPIs with small-molecules for therapeutic purposes clearly depends on the target type [163], and this is why the characterization of protein-protein interfaces is so important. In the last decades, there is increasing knowledge on small-molecules capable of inhibiting protein-protein interactions, with a lot of information available in hand-curated databases such as TIMBAL [164], 2P2I [165] and iPPI-DB [166, 167]. TIMBAL database holds a wide diversity of PPI inhibitors, including peptides, useful for helping to characterize which type of molecules could be involved in protein interfaces [164]. The small-molecule compounds annotated in TIMBAL tend to be large and lipophilic, engaged in hydrophobic contacts and containing fewer hydrogen bonds [164]. Another hand-curated database combined with an automated extractor from ChEMBL is 2P2I, which holds all structural information involving small-molecule inhibitors of protein-protein complexes. Interestingly, the protein-protein interfaces of the complexes in 2P2I database are smaller, more hydrophobic, with less charged residues and more non-polar atoms than those of standard hetero-dimeric complexes [165]. Finally, iPPI-DB database aims to compile the diversity of profiles of inhibitors of PPI in order to enable a rational characterization of the PPI inhibitor chemical space [166, 167]. This database includes physicochemical and pharmacological data in addition to the profile of the PPI target.

Thus, the identification of small-molecule inhibitors of PPIs is not an easy task. As above mentioned, known PPI inhibitors do not have high similarity to traditional inhibitors of enzymes and receptors [121], showing physicochemical properties that may violate traditional rules such as the Lipinski's Rule of Five [142]. The molecular size of known small-molecule inhibitors of PPIs is around 500-900 Da, with K_i values of less than 1 µM. In many cases (IL-2, HDM2, HPV E2), this value is in the mid-nanomolar to low-nanomolar range, comparable to the binding affinity of the protein-protein complex [121]. There are currently some small-molecule inhibitors of PPIs in clinical trials, and a few of

them are approved by the FDA [140, 168]. Considering the large process to make a drug entering in clinical trials, a lot of attention is focused onto drugs that have passed clinical trials or been accepted. Table 1.1 shows all the small-molecule inhibitors of PPI that have already been approved by FDA [140, 168]. The case of Gabapentin is an interesting example of drug repositioning. This drug was originally designed to mimic the chemical structure of neurotransmitter GABA, and was used as a treatment for epilepsy. But later it was found that this drug significantly reduced PKCε translocation by competitively inhibiting the interaction with the pronociceptive peptides bradykinin and prokineticin 2, and it is now widely used to relieve neuropathic pain in patients with amyotrophic lateral sclerosis (ALS) [169].

Table 1.1 FDA-approved drugs that are targeting protein-protein interactions.

PPI target	Drug	Disease	PDB ID	Drug Bank
Inhibitors of protein-Protein interactions				
Bcl-2 family	Venetoclax	Cancer	-	DB11581
BIII	Gabapentin	Epilepsy 1st; Neuropathic pain	-	DB00996
c-Myc/Max	Nadroparin	Cardiovascular	-	DB08813
CCR5/gp120	Maraviroc	HIV	4MBS	DB04835
HIF-1a	Carvedilol	Cardiovascular	-	DB01136
IL-2/IL2-R	Apremilast	Psoriatic arthritis	-	DB05676
KEAP1/NRF2	Dimethyl fumarate	Multiple Sclerosis	-	DB08908
LFA1/CAM1	Lifitegrast	Dry eye	-	DB11611
PPAR-gamma/NCOA	Rosiglitazone	Diabetes	4EMA	DB00412
Rac1	Azathioprine	Asthma	-	DB00993
S100B/p53	Olopatadine	Itching eyes	-	DB00768
STAT5	Dasatinib	Cancer	-	DB01254
Tubulin	Griseofulvin	Tinea infections	-	DB00400
αIIbβ3	Tirofiban	Cardiovascular	2VDM	DB00775
Stabilizers of protein-protein interactions				
Cyclophilins	Cyclosporine	Gaft rejection	1CWA	DB00091
Immunoglobulin FKBP1A	Tacrolimus	Immunosuppressor (after transplant)	1FKJ	DB00864
Transthyretin/RBP	Diflunisal	Rheumatoid arthritis and osteoarthritis	3D2T	DB00861

1.8. The path for personalized medicine in the digital era

Personalized medicine, big data and artificial intelligence are emergent concepts in recent years, which will be almost routine in our immediate future. However, there is not an artificial intelligence without natural intelligence before. As well, there is not an accurate precision medicine without biomedicine advances. Neither there is a digital era without data. From this point of view, we must be conscient about which way the science advances, and it moves through a digital ecosystem in a common environment of data where there will be a necessity of the accessibility of such data.

The health of species is a global responsibility and since 2000, a vast number of countries have invested economically in genomic projects involving studies of its population. Now a large and diverse amount of genetic data is stored in databases that are updated continuously. Past medical history of the patient is the basis for personalized medicine, but the integration of additional descriptive and observational data is enlightening for a medical decision. Therefore, biomedical advances in genomics, metabolomics, proteomics, and interactomics can generate valuable and useful data to be integrated for personalized medicine. For this purpose, the most direct contribution to clinical practices is genome sequencing providing genetic antecedents of the patient. However, also other fields in omics, such as proteomics and interactomics, are making advances intending to generate valuable data such is the interactome network together with further characterization of variants affecting protein-protein interactions. In this sense, this thesis is based on proteomics and interactomic data, and aims to contribute to the structural interpretation of genomic data at the molecular level, as well, develop computational tools and protocols in order to cover missing genetic interpretations at a molecular level.

2. OBJECTIVES

The role of protein-protein interactions in biological systems is essential. This is especially relevant when PPIs can be perturbed by single amino acid variants (SAVs) leading to disease. Then, understanding the structural characterization of proteins and its multimeric assemblies is crucial for a better comprehension of the effect of specific SAVs in the organism. Moreover, protein-protein interfaces may constitute an important therapeutic target that can be targeted by small-molecules. However, identify small-molecules is very challenging since few PPI modulators are currently known.

One of the major difficulties is that only a tiny part of the estimated number of existing protein-protein complexes has available experimental structure. Trying to overcome the lack of structural knowledge, the main purpose of this PhD thesis is the application, development and assessment of computational tools for the structural, energetic and dynamic characterization of protein-protein interactions and their protein-protein interfaces for their modulation.

The objectives of the thesis can be grouped in these general aims:

i. Development and benchmarking docking-based computational tools to model and characterize protein-protein interactions.

ii. Application of computational tools for the structure- and energy-based identification of protein-protein interactions affected by pathogenic SAVs.

iii. Docking-based identification of transient cavities on protein-protein interfaces for their modulation by small molecules.

3. Integrative modeling of macromolecular assemblies: validation and new challenges

The work in this section has been published in *Proteins* [170, 171], *Curr Opin Struct Biol*.[172], *Encyclopedia of Life Sciences* [3], and *Methods Mol Biol.* [4].

3.1. Background

Given the importance of protein-protein interactions in virtually all biomolecular processes, their atomic-level knowledge would be useful for many biomedical and biotechnological applications, such as a better understanding of disease at the molecular level, the interpretation of genomic variants, or the identification of relevant molecules with therapeutic or biotechnological purposes. However, the current structural coverage of human interactome is very limited [44, 48]. For this reason, computational docking is becoming an essential tool in structural biology, with a growing community of developers providing new methods for the many challenges that the field faces. Central for this is the CAPRI assessment experiment [173, 174], which has been a catalyzer for the docking community, providing new challenges, validation tools, strong consensus on the evaluation of docking performance, and overall, motivating discussions. Over time, the CAPRI experiment has been progressively extended to all varieties of problems related to the structural modeling of protein interactions. As an example of this, the recent CASP12-CAPRI37 [175] and CASP13-CAPRI46 [170] joint rounds, which took part during the 7th CAPRI edition [176], imposed an additional difficulty level in the field of prediction of protein interactions, with a number of multimeric targets in which the docking approaches were fundamental to model many of the oligomeric interfaces.

Thus, the participation in CAPRI and CASP has an important impact in achieving new advances in the field as well as it is essential for validating our methodology. Here, we present the results of two blind experiments: the 7th CAPRI edition (comprising CAPRI rounds 38-45) and the joint CASP13-CAPRI46 round. In most of the cases, the existence of a template to model the assembly was critical. However, in some targets our pyDock scoring [80] played an important role in the selection of the correct template-based multimeric models among all possible ones. These CAPRI (also CASP-CAPRI) rounds included a number of multi-molecular assemblies in addition to the traditional protein-protein binary complexes, together with protein-peptide and challenging protein-saccharide interactions. In all proposed targets of CAPRI (except CASP-CAPRI), one or more of the components of the complex had no experimental structure and needed to be modeled. In some cases, there were available templates for at least part of the interaction, which encouraged us to integrate template-based modeling, docking, scoring, experimental restraints, and further refinement tools. Overall, this provided a realistic representation of the complexity of the problem of structural modeling of protein interactions. We present

here the main aspects of our participation in this CAPRI and CAPRI-CASP edition, and the new procedures we devised in response to the challenges proposed by the assessment experiment.

The 7th CAPRI edition comprised eight protein-protein complexes, i.e. four hetero-dimers (A:B), two hetero-trimers (A:B:C and A2:B), one hetero-hexamer (A2:B4) and one homo-decamer (A10), three protein-peptide complexes, i.e. one hetero-dimer (A:B) and two hetero-trimers (A2:B), and five protein-oligosaccharide complexes (A:B). On the other hand, the CASP13-CAPRI46 experiment comprised 20 protein-protein complexes: ten homo-dimers (A:A), one homo-trimer (A3), one homo-tetramer (A4), one homo-octamer (A8), four hetero-dimers (A:B), one hetero-tetramer (A2:B2), one hetero-18mer (A6:B6:C6), and one homo-dimer of 5-domain monomers (A_5D:A_5D). Table 3.1 describes all 38 CAPRI and CAPRI-CASP targets divided into difficulty levels, depending on template availability. It also provides information concerning the stoichiometry, size, number of evaluated interfaces for each target, and PDB experimental structures, if available.

Table 3.1 CAPRI/CASP targets divided into difficulty levels, depending on template availability.

CAPRI 7th experiment

Target	Level	Stoich.	#Int[1]	#Res[2]	PDB[3]	Description
Protein-protein complexes						
T122	Difficult	A1B1C1	1	198/328/330	5MZV	Human cytokine hetero-dimer/receptor complex IL23/IL23R
T123	Difficult	A1B1	1	174/121	-	PorM-Nt/nb(02)
T124	Difficult	A2B1	1	202/141	6EY6	PorM-Ct/nb(130)
T125	Difficult	A2B4	5	135/146	5MGT	Hetero-hexamer of LLT1/NKR-P1 (extra-cellular domains)
T131	Difficult	A1B1	1	108/404	6BGB	Human CEACAM1/HopQ-Type-I *H. pylori*
T132	Medium	A1B1	1	108/418	6BGH	Human CEACAM1/hopQ-Type-II *H. pylori*
T133	Easy	A1B1	1	69/95	6ERE	Redesigned Colicin E2 DNase/Im2 complex
T136	Easy	A10	3	751	6Q6I	LdcA *P.aeroginosa;* EM
Protein-peptide complexes						
T121	Difficult	A1B1	1	115/13	-	*P.aeroginosa* TolAIII domain/N-terminus *P.aeruginosa* TolB
T134	Easy	A2B1	1	88/50	6GZJ	DLC8 dimer/MAG 50-residue fragment
T135	Easy	A2B1	1	88/12	6GZL	DLC8 dimer (Rat)/MAG 12-residue fragment
Protein-oligosaccharide complexes						
T126	Difficult	A1B1	1	415/6	6RKH	Arabino-oligosaccharide binding protein, *G.stearothermophilus*, with AbnE/A6
T127	Difficult	A1B1	1	415/5	6RKX	Idem with AbnE/A5
T128	Medium	A1B1	1	415/4	6RL2	Idem with AbnE/A4
T129	Medium	A1B1	1	415/3	6RL1	Idem with AbnE/A3
T130	Easy	A1B1	1	315/5	6F1G	Arabino-oligosaccharide binding protein, *G.stearothermophilus*, with AbnB/A5

Joint CAPRI-CASP 13th experiment

Easy targets	CASP ID	Stoich.	#Int[1]	#Res[2]	PDB[3]	Description
T140	T0973	A2	1	146	-	Bacteriophage ESE058 coat protein
T143	T0983	A2	1	245	-	Cals10 protein
T144	**T0984**	A2	1	752	6NQ1	Two-pore calcium channel protein; EM
T152	T1003	A2	1	474	6HRH	ALAS2, 50-Aminolevulinate synthase 2
T153	T1006	A2	1	79	6QEK	Putative membrane transporter (C. desulfamplus)
T147	T0995	A2/A4/A8	3	330	-	Cyanide dihydratase (B. pumilus); EM
T158	T1020	A3	1	577	-	SLAC1 protein
T139	**T0961**	A4	2	505	6SD8	Acyl-CoA dehydrogenase from Bdellovibrio bacteriovorus
T142	H0974	A1B1	1	70/80	-	Repressor-antirepressor complex (lysogeny switch)
Difficult targets	**CASP ID**	**Stoich.**	**#Int[1]**	**#Res[2]**	**PDB[3]**	**Description**
T137	T0965	A2	2	326	6D2V	NADP-dependent reductase
T138	T0966	A2	2	494	5W6L	RasRap1 site-specific endopeptidase
T141	T0976	A2	1	252	6MXV	Rhodanese-like family protein, bacteria
T148	T0997	A2	1	228	-	LD-transpeptidase
T149	T0999	A2	5	1589	6HQV	Pentafunctional AROM polypeptide: five main enzymes of the shikimate pathway
T150	T0999	A2	5	1589	6HQV	Idem; with SAXS data
T151	T0999	A2	5	1589	6HQV	Idem; with crosslinking data
T154	T1009	A2	1	718	6DRU	Alpha-xylosidase
T155	H1015	A1B1	1	89/129	-	CDI_213 protein, bacteria
T156	H1017	A1B1	1	111/129	-	201_INDD4 protein, E. coli
T157	H1019	A1B1	1	58/88	-	CDI207t protein, E. coli
T146	H0993	A2B2	3	275/112	-	Lipid-transport, bacterial outer membrane
T159	H1021	A6B6C6	7	148/351/295	6RAP	18-mer heterocomplex; EM

[1] Number of protein-protein interfaces. [2] The number of residues for each monomeric unit is indicated in #Res column. [3] PDB ID of experimentally solved structures of the complex

3.2. Material and Methods

3.2.1. Processing input structures

For the blind CAPRI targets, the structures of one (or more) of the subunits were not available and needed to be modeled before docking. In most of the targets, we used MODELLER 9v19 with default parameters [177] based on the template/s suggested by the organizers or on other homologous proteins found by BLAST [178] search tools. The final selected model was the one with the lowest DOPE score [179]. In some cases (e.g. T131, T132), highly flexible modeled loops were removed before docking and they were rebuilt in the final models before the minimization phase. In target T133 the flexible C-terminal of the ligand was removed and then rebuilt in the final models. In some cases (T123, T124, T136), we also performed multiple template modeling of any of the interacting subunits with I-TASSER [180]. In the case of the peptides (T134-T135), they were modeled in a

similar way, based on available templates. All templates used in the 7th CAPRI edition are available in the supplementary Table 8.1.1 (see Appendix 1).

By contrast, the structures of the subunits were available in the CASP13-CAPRI46 experiment. We used the rank #1 predictions from ZHANG, ROSETTA, and QUARK CASP-hosted servers as starting models of the individual monomers for the docking-based assembly modeling. On the other side, we used the top five released predictions from the ZHANG, ROSETTA, QUARK, MULTICOM-CONSTRUCT, and RAPTORX-DeepModeller CASP-hosted servers as starting models of the individual monomers for the template-based assembly modeling. For this experiment, no other template was used to model the individual subunits.

3.2.2. Template-based docking

In half of the targets (T137-T144, T152-T154, T158) in CASP13-CAPRI46, the global assembly was also modeled based on available templates, which we extracted from the above mentioned CASP-hosted servers as well as from BLAST search. Concretely, we extract the possible templates from the five CASP-hosted servers used (ZHANG, ROSETTA, QUARK, MULTICOM-CONSTRUCT and RAPTORX-DeepModeller). The final templates were selected based on the biological unit of interest, the sequence identity with the targets, as well as on the structural similarity of conserved templates. All available templates were clustered to remove redundant structures. The monomeric models (see section 3.2.1.) were superimposed onto the corresponding subunits of every non-redundant template, and all generated assembly models were then scored based on energy (see next section 3.2.6.). In the multi-molecular assemblies, some of the interfaces were built based on available templates (see next section 3.2.5.). Regarding the blind 7th CAPRI rounds (T121-T136), the use of template-based modeling was especially relevant for the global assembly.

3.2.3. *Ab-initio* docking

In general, for the protein-protein and protein-peptide targets, we used FTDock 2.0 [68] (with electrostatics and 0.7 Å grid resolution) and ZDOCK 2.1 [181] to generate 10,000 and 2,000 rigid-body docking poses, respectively, in the same conditions as previously described, [182] with the exception of a few targets (T131, T132, T136, T149, T159), in

which ZDOCK was not used because of the computational cost of long proteins. For three targets (T131-T133), we used our LightDock [78] method to generate an additional set of flexible docking poses, which included explicit backbone flexibility by using Anisotropic Network Model (ANM) [183] during the sampling process, and the DFIRE [184] and pyDockLite [78] scoring functions. The number of poses generated by LightDock, which depended on the final step of clustering, was 6,093, 6,005 and 4,977 for targets T131, T132 and T133 respectively. As usual, cofactors, water molecules and solvent ions were not included in our docking calculations. In homo-oligomeric pairs, we kept only the docking poses with the expected symmetry (e.g. C2 for homo-dimers, C3 for homo-timers, etc.).

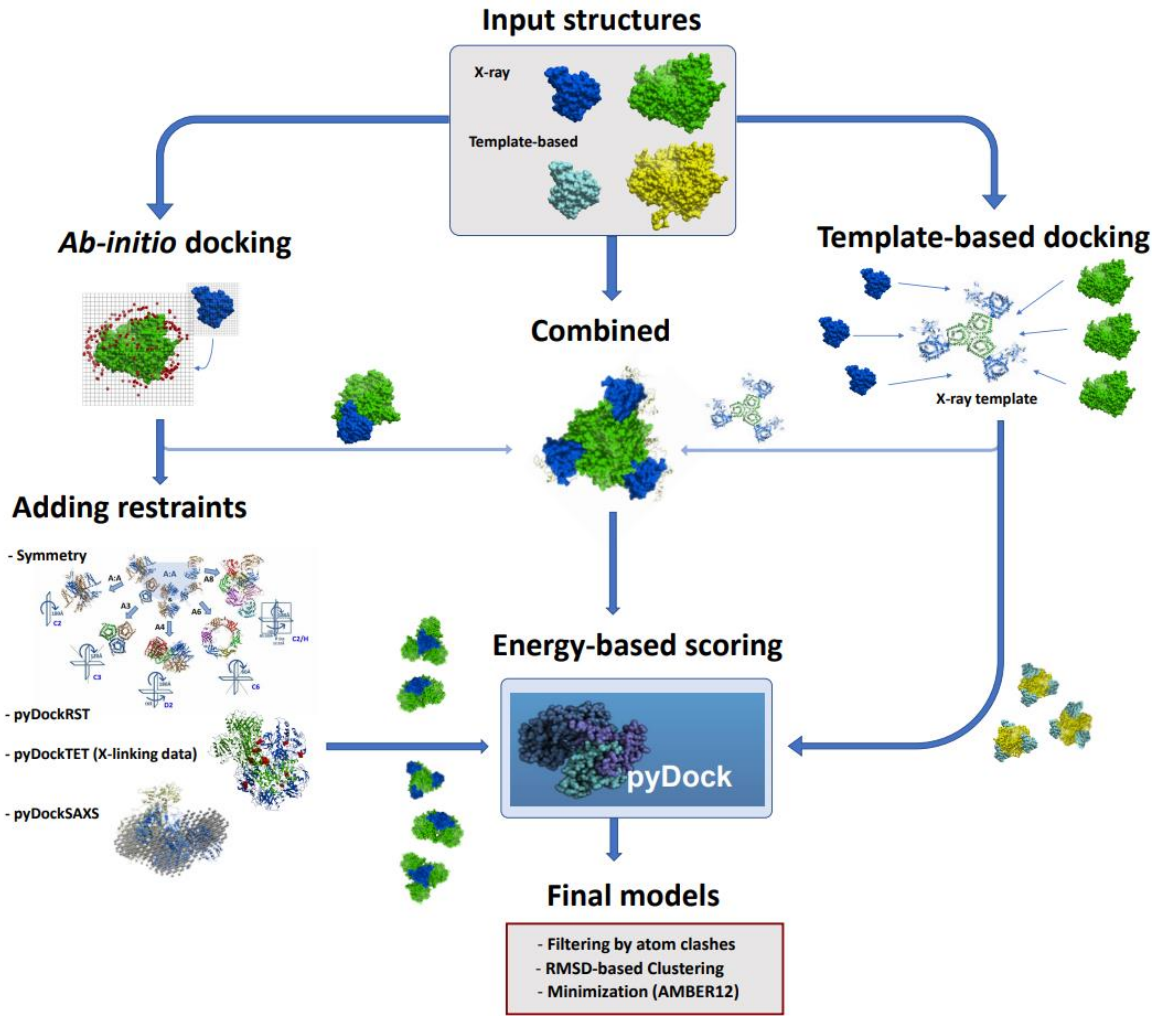


Figure 3.1 An example of the combination of template-based, ab initio docking and external data for integrative modeling of complexes.
The scheme is based on the strategy followed by our group (Fernández-Recio) as predictors in the recent CASP13-CAPRI and 7th CAPRI experiments.

3.2.4. Combined sampling method

In some targets, the global assembly was modeled by integrating template-based and *ab initio* docking (Figure 3.1). In target T136, the interfaces of the homo-decamer were modeled based on available template from BLAST, by superimposing the binary docking models on the global template (PDB 5FKZ). In target T146 (A2:B2), the homo-dimer interfaces were modeled based on all available templates from CASP-hosted servers, and the heteromeric interfaces by docking the homo-dimer sub-assemblies. In target T147 (A8), homo-dimer template-based models were docked to form tetramers, keeping only models capable of forming octamers with 2-fold helical symmetry. In target T159 (A6:B6:C6), the three homo-hexameric rings were independently modeled based on templates, and then one pair of rings were modeled based on available templates as well as by FTDock docking, keeping only models in which the hexameric symmetry axes of the interacting rings overlapped, and another pair of rings was modeled only by docking.

3.2.5. Integration of restraints

If experimental information was available, the global assembly was modeled by including distance restraints. We used a varied source of restraints (Figure 3.1). For instance, from structures of homologous proteins or with conserved/similar interactions, we can estimate the interface residues that will be used as restraints in docking or scoring. This was the case of T134-T135 (PDB 1F95 had 5 Å RMSD from protein-peptide complexes) and of target T153 (PDB 3W36 had 10 Å RMSD from protein-protein complex). The binding sites of protein-oligosaccharide targets T126-T129 were estimated from homologous PDB 5F7V (30% SI), and that of target T130 was estimated from PDB 3D5Z (100% SI) (see next section 3.2.8.).

On the other hand, for some targets the available knowledge of experimental residue contacts was used as distance restraints with pyDockRST [100]. In target T122, we found in the literature that Trp156 in IL-23A was important for the interaction [185], so we used it as a distance restraint, in the context of our standard docking protocol as servers, predictors and scorers. In target T125 we added distance restraints (10 Å inter-atomic cutoff) derived from residues identified in the literature as potentially involved in the interaction (LLT1 Lys169, and NKR-P1 Glu205) [186, 187]. For two targets consisted in

two different complexes of hCEACAM1 protein with HopQ Type I (T131) and HopQ Type II (T132), distance restraints were imposed based on two potential interface residues in hCEACAM1 (Tyr35 and Ile92) [188]. In target T136, the docking sampling was filtered by distance restraints derived from the estimated interface residues from template PDB 5FKZ (using 10 Å distance cutoff), and superimposing the binary docking models on the global template (PDB 5FKZ).

Symmetry restraints were applied for modeling homo-oligomers (either template-based or *ab initio*), since its unbound structure is rarely available. For practical purposes, in all homo-oligomeric targets (Figure 3.1), we assumed symmetric oligomerization, e.g. rotational symmetry C_2, C_3..., in order to filter the resulting docking models. Based on these considerations, pyDock has been applied in blind conditions to the modeling of a variety of homo-oligomers. In general, we annotated docking poses with C2 symmetry as those that have a rotation angle between 175° and 180°, and helical rise < 5 Å.

For target T149, involving the dimerization of a 5-domain protein, the challenge was not only to model the dimer orientation but also to describe the assembly of the 5 different domains (D) within each monomer. On the one side, we applied ab initio docking to the available CASP-hosted server models of the 5-domain monomers. On the other side, we applied an *ad-hoc* strategy as follows. First, each domain was independently modeled based on the rank #1 prediction from QUARK CASP-host server. Second, the intermolecular orientation between the first domains from each monomer (D1-D1') was modeled based on a template (PDB 1DQS). Then the interaction between D1 and the next domain (D2) of the same monomer was modeled by docking, imposing restraints derived from the inter-domain linkers with pyDockTET module [189]. For each D1-D2 model, a copy of it (D1'-D2') was superimposed on D1-D1' to generate D2-D2' pairs. This strategy was iteratively applied to the other domains (D2-D3 by docking, D2'-D3' by superposition, D3-D4 by docking, etc.). The final dimer models were selected by scoring the energy of intermolecular domain interactions with pyDock. When SAXS data was made available (target T150), models were re-scored with pyDockSAXS [98]. Finally, when cross-linking data was released (target T151), the number of residue pairs satisfying these experimental contacts was also evaluated.

3.2.6. Energy-based scoring

In general, using the pyDock *bindEy* module, we computed the pyDock docking energy for a given complex structure (either experimentally determined or modeled). When no clearly homologous templates were found, or when the unbound proteins could not be easily modeled, the challenge was to build suitable models among the possible docking orientations based on a variety of remote template structures. The use of pyDock scoring helped to identify the correct models. Based on this, we scored the docking models generated by the above described methods with our default pyDock protocol [80], based on energy terms previously optimized for rigid-body docking. Concretely, we scored the above described oligomeric models with pyDock, sorting them according to the total binding energy of all possible interfaces. The binding energy is basically composed of ASA-based desolvation, Coulombic electrostatics and van der Waals energy (with a weighting factor of 0.1 to reduce the noise of the scoring function). Electrostatics and van der Waals were limited to -1.0/+1.0 and 1.0 kcal/mol for each inter-atomic energy value, respectively, in order to avoid excessive penalization from possible clashes derived of the rigid-body approach. For the target T133, we also applied our new IRaPPA scoring server [63] to the docking models generated by FTDock. It was an opportunity to test in a redesigned complex from the already evaluated target T47. In last round involving T47 we got high (***) quality predictions in submission rank 1 for predictors and rank 2 for scorers. This redesign suggests more efforts in the prediction with differences with respect T47. Then, we changed the traditional pyDock scoring (successful in previous round T47) for integrating new methodologies such as lightDock [78] getting more flexibility and IRaPPA increasing the interface parameters to being evaluated in the scoring function. In half of the targets we were able to define possible interface residues based on experimental information available in the literature (T122, T125, T126-T130, T131, T132, T149-T151) or from homologous complex structures (T121, T134, T135, T136, T143, T144, T153). This information was usually included in the final scoring as distance restraints with pyDockRST [100], pyDockSAXS [98] or/and pyDockTET [189]. In some cases (T125, T146), we filtered out the docking models incompatible with the estimated position of the membrane [187, 190]. In the case of the peptides, we forced docking models to adopt antiparallel β-strand orientation (which turned out to be correct for targets T134, T135, but incorrect for target T121). Cofactors, water molecules and solvent ions were not considered for scoring. After scoring, we eliminated redundant predictions by using a BSAS algorithm

[191] with a distance cutoff of 4.0 Å, as previously described [192]. In the case of template-based or symmetry-based models, we eliminated those ones with strong clashes that would be difficult to solve with minimization.

The final ten selected docking poses were minimized by using different versions of AMBER (AMBER12[193] or AMBER17[194]), with AMBER ff99SB and gaff force field [195] with implicit solvent in order to improve the quality of the docking models and reduce the number of interatomic clashes, as previously described [196]. The minimization protocol consisted in a 500-cycle steepest descent (SD) minimization with harmonic restraints applied at a force constant of 25 kcal/(mol·Å^2) to all the backbone atoms in order to optimize the side-chains, followed by another 500 cycles of conjugate gradient (CG) minimization without restraints. In some cases, due to time constraints or because of earlier convergence, the minimization protocol varied (e.g. in T134 we used 200-cycle SD and 300 cycle-CG; in T135 500-cycle SD and 100-cycle CG; in T136 some models were not minimized, or were minimized in vacuum; the largest CASP targets T149-151 and T159 were minimized in vacuum). In targets T131 and T132, the loops previously removed for docking were rebuilt by MODELLER before the final minimization step.

The scoring and minimization protocol that we used for the docking models generated as predictors was basically applied in the same way to the set of models provided for scorers (exceptions: no distance restraints as scorers in T121 and T136; IRaPPA was not used as scorers in T133). The number of available templates and their reliability determined the percentage of template-based complex models included in the final 5 submitted models in CASP (10 for CAPRI). Finally, we eliminated the redundant predictions and minimized the final ten selected docking models. In the scorers experiment, we eliminated all the docking models with a percentage of secondary structure significantly lower than the one observed in the corresponding set of structures previously selected as predictors. Models with more than 250 clashes (i.e. intermolecular pairs of atoms closer than 4 Å) were also removed. Then, the same protocol used in predictors was applied to score the docking models (favoring models structurally similar to reliable available templates).

For protein-oligosaccharide targets (T126-T130), the electrostatics energy was calculated as in standard pyDock [80], using the atom charges obtained as above described. The van der Waals energy was calculated with the Lennard-Jones parameters corresponding to the AMBER types of the saccharide atoms. Regarding the desolvation

energy term, the saccharide C atoms were assigned the same atomic solvation parameters (ASPs) as "C aliphatic" in pyDock [80], while the O atoms were assigned the same ASPs as "O hydroxyl" in pyDock. For scorers, we used another charge model due to time constraints, the Gasteiger-Marsili empirical atomic partial charges [197], and the final scoring was only based on this new pyDock version adapted to glucide interactions.

3.2.7. Docking and scoring for the servers experiment

We participated in all rounds of the 7th CAPRI edition (Round 38-45). More specifically, we only participate in all protein-protein and protein-peptide targets with our pyDockWeb server (https://life.bsc.es/servlet/pydock) [83]. The generation of docking poses and further scoring were done in a fully automatic manner by FTDock and pyDock, as previously described. In some targets, we used the same interface residue restraints (T122, T131, T132), or the same membrane-based filtering (T125) as in predictors, but in general, we used much less external data to process the models than in predictors due to time constraints. In multi-molecular assemblies, additional modeling steps were performed based on the docking models provided by the server (more details for each target in the Results section). Finally, the best-scored server predictions were clustered and minimized according to our default protocol before submission to CAPRI (with the exception of T122, in which final models were not minimized, and T131 and T132, in which the loops removed before docking were rebuilt by MODELLER with no further minimization).

3.2.8. Modeling of protein-saccharide complexes

For the protein-saccharides targets (T126-130), we used rDock [198] (http://rdock.sourceforge.net/) to generate and score the models. Additionally, we developed a new pyDock module specially adapted for the scoring of saccharide molecules. This new module can read topology and coordinate files from AMBER, using a dictionary of atom types based on AMBER94. This makes it possible to load the data from the oligosaccharide in order to compute the energy-based scoring function. To obtain these files, we used *antechamber* with the AM1-BCC charge model [199], setting the net charge to 0, and then *parmchk2* to obtain the charges, the energetic angle parameters, and a *mol2* file. Then we used *LEaP* to load the general forcefield of AMBER (GAFF) and followed the procedure to generate a library with the information obtained from *antechamber*. As a

final step, we used *LEaP* to load each docking pose and obtain its coordinates (*.incrd*) and topology (*.prmtop*) files, in order to be scored with pyDock. The center of the cavity used in rDock docking was defined as the center of masses of known ligands bound to homologous proteins (PDB 5F7V for T126-T129; PDB 3D5Z for T130). From PDB 3D5Z (100% SI with T130), we used SCWRL to mutate Gly9 residue to Asn as in WT, and kept water molecules 501 and 658, which were found to be important for the catalytic activity [200]. Thus, we submitted two sets of models. The models 1-5 were obtained as scored by rDock, and the models 6-10 were obtained as scored by pyDock. For scorers, we used another charge model due to time constraints, the Gasteiger-Marsili empirical atomic partial charges [197], and the final scoring was only based on this new pyDock version adapted to glucide interactions.

3.3. Results

For the 7th CAPRI edition, we participated in all the proposed targets, as predictors, servers and scorers (with the exception of the protein-saccharide cases, in which our server did not participate since it was not ready for this type of interactions). Our participation in the two CASP-CAPRI joint rounds is described elsewhere [86, 170]. Here we report about our results for the 39 targets proposed in 7th CAPRI and CASP13-CAPRI editions (considering hetero-meric and homo-meric interfaces in target T125 as two separate targets), which are summarized in Table 3.2 and Table 3.3.

3.3.1. Seventh CAPRI edition

When considering our top 10 models, as in past editions, we submitted acceptable models (or better) for 10 targets as predictors, four targets as servers, and 13 targets as scorers (Table 3.2). The actual number of evaluated targets was 19, because some of the interfaces in these multimeric assemblies were considered as independent targets This represents a success rate of 53% as predictors, 21% as servers, and 68% as scorers, which is consistent with our trajectory in CAPRI. The performance as scorers was especially good, actually the best of all participants (for more details see Appendix 1: Table 8.1.2, Table 8.1.3 and Table 8.1.4). If we consider only our top 5 models, as it is now the consensus in CAPRI in order to be more realistic for practical applications, we submitted acceptable or better

models for ten, three and eight targets as predictors, servers and scorers, respectively. When comparing our performance on the top 5 submitted models with respect to that based on the traditional top 10, the results as predictors and servers did not change too much, but we had a significant drop in performance as scorers.

Predictions involving protein-protein complexes consist of six difficult, one medium, and two easy targets. Protein-peptide complexes were involved in one difficult and two easy targets. Protein-oligosaccharide complexes were assigned by two difficult, two medium and one easy targets (Table 3.1). The summary of our performance is represented in Table 3.2 that shows the quality of the submitted top 10 predictions (high***, medium** and acceptable* [176]).

3.3.1.1. Successful predictions

Most of the successful predictions for predictors also got a good quality for scorers. Server predictions got an additional acceptable quality for the T122 target (assigned as a difficult target), while did not get success for T134. For scorers, besides the targets correctly predicted in predictors, four more targets (T122, T125, T131, T132) were successfully predicted, which most of them classified as difficult targets (T132 medium) (Table 3.2).

Table 3.2 Quality of submitted predictions for the 7th CAPRI experiment predictions.

Target	Level	Stoich.	Submission quality for: Predictors[1]	Servers[1]	Successful groups[2]	Submission quality for Scorers[1]	Successful groups
Protein-protein complexes							
T122	Difficult	A1B1C1	-	**M08***	11/35	**M08** (M05*)**	9/20
T123	Difficult	A1B1	-	-	0/32	-	0/19
T124	Difficult	A2B1	-	-	0/29	-	0/18
T125h[5]	Difficult	A2B4	-/-	-/-	15/1 of 30	**M02**** / -	18/0 of 19
T125	Difficult	A2B4	**M01***** / -	**M01***** / -	26/0 of 30	**M02*** (M01**)** -	17/0 of 19
T131	Difficult	A1B1	-	-	1/30	**M06****	4/19
T132	Medium	A1B1	-	-	3/30	**M08****	6/19
T133	Easy	A1B1	**M01****	**M05***	30/35	**M02****	18/20
T136	Easy	A10	**M01**/M02*/ M07***	- [M01**]/- [M01*]/-	26/29/26 of 30	**M02** (M01*)/M02**/ M08** (M02*)**	16/16/16 of 16
Protein-peptide complexes							
T121	Difficult	A1B1	-	-	5/33	-	8/17
T134	Easy	A2B1	**M03***	-	15/31	**M09****	15/18
T135	Easy	A2B1	**M01***	**M02***	23/30	**M01*****	17/17
Protein-oligosaccharide complexes							
T126	Difficult	A1B1	**M08** (M03*)**		19/29	**M08***	16/17
T127	Difficult	A1B1	**M01***		26/28	-	13/14
T128	Medium	A1B1	**M02***		28/29	**M10** (M01*)**	17/17
T129	Medium	A1B1	**M02***		27/30	**M09** (M02*)**	16/16
T130	Easy	A1B1	**M02** (M01*)**		26/29	**M06****	17/17

[1] In general, we indicated the best quality models within the top 10 submitted models (in parenthesis, we indicated if there were another successful model within the top 5 or top 1). T136:

in square brackets indicate quality models that were disqualified due to clashes. ***=high-quality models; **=medium-quality models; *=acceptable-quality models for the results in the performance [176].

[2] In general, docking servers are included in predictors groups.

The availability of templates and experimental restraints facilitates the modeling of monomeric subunits and complexes. In T122 (reference PDB 5MZV is now available), despite finding an available homologous template (PDB 1I1R, 25% SI), which had three conserved disulfide bonds, this was a difficult target for which only 9 predictor groups had acceptable models within the top 5 submitted models. The T125 (reference PDB 5MGT is now available) target had four different interfaces (LLT1:NKR-P1 / LLT1-LLT1 / alternative LLT1:NKR-P1 / NKR-P1:NKR-P1), and the heteromeric and homomeric interfaces were independently assessed, as two different targets (Table 3.2). We had high accuracy models for the LLT1-LLT1 interface as servers, predictors, and scorers, which were actually built by standard template-based modeling (PDB 4QKH, 100% SI). Interestingly, we identified a medium model for LLT1:NKR-P1 interface as scorers. Target T136 (homo-decamer, reference PDB 6Q6I is now available) is a clear example of a successful application of template-based modeling for the global assembly. Distance restraints were derived from the estimated interface residues from template PDB 5FKZ (using 10 Å distance cutoff), and superimposing the binary docking models on the global template results in a successful strategy. In protein-peptide targets (T134-T135, references with PDB ID 6GZJ-6GZL are now available), binding residues at 5 Å from the peptide were used as distance restraints for docking applying the same distance restraint in scorers. For modeling protein-oligosaccharide complexes (references for T126-T130 are now available and annotated in Table 3.1) define the center of the cavity from homologous proteins for docking was crucial to get successful predictions.

For the re-designed complex in T133 (reference PDB 6ERE is now available), despite PDB 3U43 (CAPRI target T47) being a good template, we did not use it for modeling the protein-protein complex. There were acceptable models (as servers) and medium ones (as predictors and scorers). Interestingly, the medium model submitted as rank 1 for predictors was obtained with the new IRaPPA scoring.

pyDock, as a scoring tool, has been able to get successful predictions for 13 out of 19 exposed targets (mostly with medium quality). Energy-based scoring shows excellent performance for the evaluation of protein-protein complexes and blind prediction of

protein-protein and protein-peptide interactions. In addition, the protein-saccharide models were all selected according to the new adapted pyDock version, obtaining acceptable and medium models for T126-T130. For T131-T132, we did not submit any acceptable model within our best-ranked 5 models, either as servers, predictors or scorers, but we had medium models as scorers for both targets within our top 10 models. For servers and predictors in T131-T132 (references with PDB ID 6BGB-6BGH are now available), the missing loops in HopQ type I or II were rebuilt for the best-ranked 100 docking poses before the minimization step, thus providing implicit flexibility to the docking procedure. We could speculate that a possible reason for failure is the modeling of these highly flexible loops.

3.3.1.2. Unsuccessful predictions

Difficult targets such are T123 and T124 (did not get successful predictions for any participating group. For predictors and servers experiment, few groups got successful predictions for T131, T132, T125 (hetero-complex) and T121.

In T122 we do not get successful predictions in predictors. The main difficulty was to identify which of the topologically similar three domains of IL-23R was interacting with IL-23. We could speculate that a possible reason for failure is the nanobody needed for the crystallization of the complex, since it was not considered in our modeling. However, it is unlikely that the nanobody is affecting the complex orientation in this case, since it is far from the interface, and there are other structures in which the molecules show similar conformation with or without the nanobody (e.g. PDB 4GRW, 5MXA). We believe that the major problem, in this case, is actually related to the difficulties in modeling IL-23R structure. In relation to T123 and T124 (reference PDB 6EY6 is now available) targets, we could speculate that complexes involving nanobodies are particularly challenging for current docking methods, but the major reason of failure in this case seems to be the difficulties in modeling the PorM domains. For T125 target, as a general strategy, we docked LLT1 homodimer vs. NKR-P1 homo-dimer, and then built the second NKR-P1 by symmetry (this assumption turned out not to be correct, since there are two different LLT1:NKR-P1 interfaces in the complex structure, now available (PDB 5MGT)). In T131 and T132 targets, the distance restraints imposed based on two potential interface residues in hCEACAM1 (Tyr35 and Ile92) [188], was a wrong strategy. For the protein-peptide T121 target, only docking decoys with antiparallel β-strand orientation were selected, which turned out to be a wrong decision according to the data presented and discussed

during the 7th CAPRI evaluation meeting in Hinxton (UK). As a consequence, no acceptable models were found in any of our submissions. In any case, this was a difficult target, with only 3 predictor groups with acceptable models within the top 5 submissions.

3.3.2. CASP13-CAPRI experiment

The recent CASP13-CAPRI challenge comprised a total of 20 oligomeric protein assemblies, including 14 homo-complexes and 6 hetero-complexes, which could be classified into 15 dimers and 5 multimeric assemblies [170]. This experiment is involving a total of 9 easy and 13 difficult targets (T149-T151 related with the same oligomer) (Table 3.1). When considering our top 10 models, we submitted acceptable models (or better) for 13 targets as predictors (considering T149-T151 as a unique target) and 12 targets as scorers out of 20 blind targets (Table 3.3). This is a success rate of 65% (predictors) and 63% (scorers) with respect to all targets in which we participated (Table 3.3). The performance was especially good, actually the second-best of all participants for predictors and the best for scorers (for more details see Appendix 1: Table 8.1.5 and Table 8.1.6). When comparing our performance on the top 5 submitted models with respect to that based on the traditional top 10, the results as predictors and scorers are the same [170].

3.3.2.1. Successful predictions

Our participation in CASP13-CAPRI46 blind experiment was highly rewarding since we got the maximum number of targets successfully predicted among all groups [170]. All easy targets were successfully predicted (except for T147, in which we could not predict all protein-protein interfaces in the predictors experiment). For the difficult targets, involving more than three protein-protein interfaces (T149-T151 and T159), we were able to predict some of the interfaces. In homo-dimer T141 we forgot to send the predictions on time as predictors, but we were successful as scorers. We were the unique group predicting the hetero-dimer T154, and one of the few ones that send acceptable models for the target T157 in predictors (Table 3.3).

As described in Methods, in half of the targets (T137-T144,T152-T154,T158), the global assembly was modeled based on available templates, which we extracted from the CASP-hosted servers (ZHANG, ROSETTA, QUARK, MULTICOM-CONSTRUCT and RAPTORX-Deep Modeller) as well as from BLAST search. Template-based strategy was

successful for most of the targets as predictors, in which pyDock scoring identified the correct models (T139, T140, T142, T143, T144, T152, T153, T158) (Table 3.3**).** The exceptions were T137, T138, and T154, where we got wrong modeling predictions with this strategy. Despite of the low reliability of the identified templates for the hetero-dimer T142, based on a conserved structure between templates (PDB codes 1Y7Y, 1UTX, and 2B5A), we got medium models for the two template-based submitted models according to pyDock scoring (incidentally, in this target, the submitted models built by pyDock docking were not successful).

Table 3.3 Results of pyDock in the recent CAPRI 46 round and CASP 13th for Predictors and Scorers.

Easy Targets	Stoich.	Submission quality for Predictors[1]	Successful groups[2]	Submission quality for Predictors	Successful groups
T140	A2	**	22/30	**	17/18
T143	A2	***	25/29	**	16/18
T144	A2	**	27/29	**	17/18
T152	A2	***	26/31	***	18/18
T153	A2	***	28/33	***	18/18
T147	A2/A4/A8	**/-/-	19/17/16 of 26	**/**/***	15/14/12 of 16
T158	A3	**	18/25	**	16/16
T139	A4	***/***	27/26 of 28	***/***	17/16 of 17
T142	A1B1	**	12/30	**	12/18
Medium-Diff. Targets	**Stoich.**	**Submission quality for Predictors**	**Successful groups**	**Submission quality for Predictors**	**Successful groups**
T137	A2	-/-	1/0 of 28	-	cancelled
T138	A2	-/-	0/1 of 28	-/-	0/17
T141	A2	-	7/29	*	9/18
T148	A2	-	0/32	-	0/17
T149	A2	*/-/-/-/-	12/3/4/0/2 of 21		
T150_SAXS	A2	**/-/**/-/-	9/3/3/0/0 of 21		
T151_XL	A2	***/-/-/-/-	10/1/2/0/0 of 21	***/-/-/-/-	16/1/2/0/0 of 16
T154	A2	*	1/30	-	1/18
T155	A1B1	-	1/30	-	0/17
T156	A1B1	-	3/30	-	1/17
T157	A1B1	*	5/29	-	1/17
T146	A2B2	-/-/-	7/0/2 of 29	-/-/-	7/0/4 of 18
T159	A6B6C6	**/-/-/-/-/-	18/0/0/0/0/9/8 of 22	**/-/-/-/-/**/-	15/0/0/0/0/14/12 of 15

[1] ***=high-quality models; **=medium-quality models; *=acceptable-quality models for the results in the performance [170].

We remark that our *ab initio* docking was successful as predictors in CAPRI (considering 10 submitted models instead of the top 5 models in CASP) in two difficult cases: T154 (no other group succeeded) and T157 (only 5 successful groups). In T154 (homo-dimeric reference PDB 6DRU is now available), from 8 docking-based and 2 template-based submitted models, we got 2 acceptable docking-based models. In T157, we

got acceptable docking-based models in the top 5 (Table 3.3). Despite the available template (PDB 3W63) for the homo-dimer in T153 (reference PDB 6QEK is now available), *ab initio* docking was also able to predict a high quality model in predictors.

In difficult targets involving more than three protein-protein interfaces, we used a combined strategy (*ab initio* and template-based docking), in which pyDock scoring was critical to identify the best modeled interfaces. If a suitable template is found, the complex can be directly modeled based on the template, or alternatively, the structures (or models) of the unbound proteins can be directly superimposed onto the corresponding ones of the available template. In case of closely homologous templates, both approaches can provide reasonable docking models. However, when no clearly homologous templates are found, or when the unbound proteins cannot be easily modeled, the challenge is to build suitable models among the possible docking orientations that can be obtained from a variety of remote template structures. The use of pyDock scoring can help to identify the correct models.

For targets T149-151 (reference PDB 6HQV is now available), we got acceptable models for the homo-dimerization of domain 1 (D1-D1) (1-400 res) based on a template (PDB 1DQS). The inter-domain interactions were modeled based on *ab initio* docking. In target T150, restraints derived from the inter-domain linkers were imposed with pyDockTET. This was a successful strategy for predicting the interdomain interaction between D1 and D2, since we got medium models (Table 3.3). In T151, the re-scoring of models by pyDockSAXS improved the quality of the D1-D1 interaction in all submitted models (Table 3.3). In general, this strategy was not optimal for this target (due to the limited time and its complexity), but that does not mean that it is not a suitable approach for further experiments. The successful prediction of initial interacting domains seems to be crucial for predicting the consecutive domains in the correct order. On the other side, in target T147 (A8), homo-dimer template-based models were docked to form tetramers, keeping only models capable of forming octamers with 2-fold helical symmetry. This strategy was only successful for predicting the homo-dimeric interfaces based on templates (PDB codes 2W1V, 2GGL and 5H8I), while the application of docking for predicting the other interacting interfaces was not successful). In target T159 (reference PDB 6RAP is now available), the three homo-hexameric rings were independently modeled based on templates, and then one pair of rings were modeled based on available templates as well as by FTDock docking, keeping only models in which the hexameric symmetry axes of the

interacting rings overlapped, and another pair of rings was modeled only by docking. Here the only successful strategy was the template-based prediction (based on PDB ID: 1Y12, 3EAA, 4HE1 and 3V4H) of the specific interface for the-subunit 1 ring (Figure 3.2).

In scorers, our performance was the best among all the groups. We predicted medium quality models (or better) for all easy targets (Table 3.3). For the difficult targets, we predicted acceptable models for the T141 and high quality models for the homo-dimeric interface of T149-T151. Interestingly, in target T59, in addition to achieving a medium model for the interface of ring 1, we also got a medium model for the interface of ring 2 (Figure 3.2). Despite the unsuccessful prediction of the interaction between subunits 1 and 2, the general topology was correctly modeled (Figure 3.2).

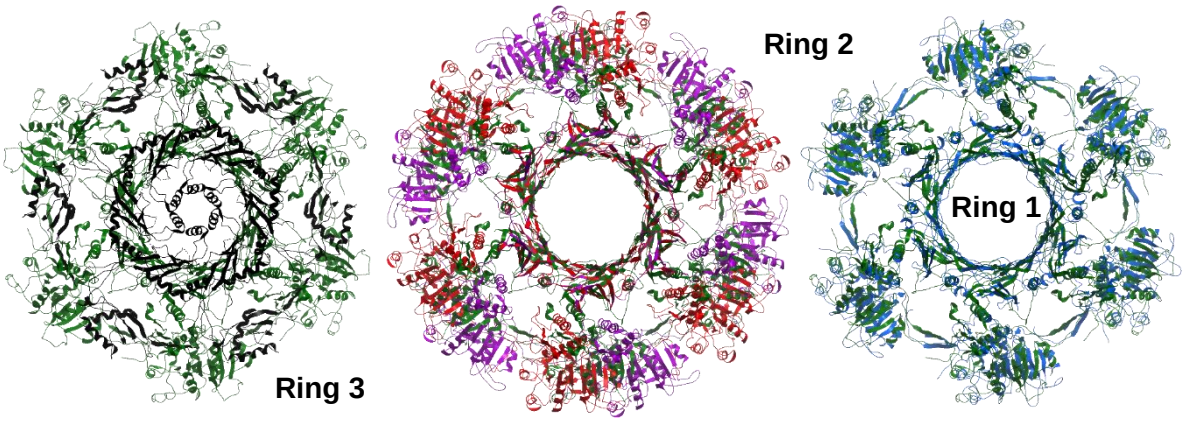


Figure 3.2 Successful models for specific protein-protein interfaces of target T159.
At left, the now available Cryo-EM structure (6RAP PDB) with subunit 1 and 2 rings in green and subunit 3 ring in black. In the middle, the rank 5 (red) and the rank 1 (magenta) submitted models in predictors with success for ring 1 formation. At right, the rank 4 (blue) submitted model in scorers with an additional success for the interaction of subunits 2 forming the ring 2.

3.3.2.2. Unsuccessful predictions

In general, unsuccessful predictions came from difficult targets. However, for the easy target T147, we were not able to predict all the protein-protein interfaces of the homo-octamer (Table 3.3). The problem was that, even though the dimeric interaction was correctly predicted, we did not identify the correct helical rise value for the formation of an octameric fiber. For the difficult targets, no participant group (either as predictors or scorers) got successful predictions for the homo-dimer T148. For the targets T137, T138, T146, T155, T156), we did not get successful predictions either for predictors or scorers. Only a few groups got successful predictions for these targets for both predictors and

scorers experiments (Table 3.3). We forgot to send models for T141 as predictors. And for targets T154 and T157, both hetero-dimers, we failed in scorers. For targets involving more than three protein-protein interfaces (T149-T151 and T159), we could not successfully predict all interfaces (Table 3.3). We could not either model all the interfaces in T147 (A8) as mentioned above. On the other hand, for the target T146 (A2:B2) the integration of template-based, *ab initio* docking and SAXS data was not successful probably because of significant conformational rearrangements.

3.3.3. General results of the participating groups in the last editions

As mentioned in the Introduction, a*b initio* computational docking can provide acceptable models within the top 10 predictions in up to 40% [49, 63, 64]. Traditionally CASP has been focused on the prediction of the structure of individual proteins. However, very often proteins are found as oligomeric assemblies, which adds complexity to the modeling effort. To evaluate the applicability of docking methodologies for the prediction of protein oligomeric assemblies, the last three CASP editions included a joint CASP-CAPRI experiment focused on multimeric assemblies, which are independently evaluated by CASP and CAPRI communities. In the most recent CASP13-CAPRI [170], there were good structural templates for the (partial or full) assembly of the 9 "easy" targets, and it was possible to find remote templates for part of the assembly of some of the remaining 11 "difficult" targets. The availability of templates in each case is critical to explain the predictive success of the groups. Focusing on the results for the top 10 predictions (to facilitate comparison with the reported performances of different docking methods in the literature), the best-performing group submitted acceptable (or better) models for 13 targets (65% of the cases) (Figure 3.3). In the "easy" targets, the best-performing group submitted acceptable models for all these cases, while in the "difficult" targets, the best-performing group submitted acceptable models for only 4 of such targets (36% of the cases). Regarding the quality of the models, high-quality models [170] were submitted by any group in 78% of the "easy" targets (with template), but only in 9% of the "difficult" targets (no template).

On the other side, the recent 7th CAPRI edition had more heterogeneous targets, which could be classified in 10 dimers and 6 multimeric assemblies [176]. For the 19 evaluated targets, there were structural templates for a total of 13 target interfaces (6 protein-protein, 2 protein-peptide, and 5 protein-saccharide). This was determinant for the

overall predictive success of the groups as well as for the quality of the predicted models. Overall, the maximum number of target interfaces successfully predicted by a single group was 13 (i.e. success in 68% of the cases) (Figure 3.3). But in cases with no available template, the best-performing groups submitted acceptable models for only 2 target interfaces (i.e. success in 33% of the cases). Regarding the quality of the models, high-quality models [176] were submitted by any group in 31% of the "easy" targets (with template) and in 17% of the "difficult" targets (no template). The 7th CAPRI edition showed that *ab initio* docking in cases for which there is no available template is still highly challenging, and progress is actually coming from the efficient procedures to combine template-based modeling and other docking methodologies.

3.4. Discussion

3.4.1. Combination of template-based and ab initio docking

The CASP and CAPRI experiments show that template-based modeling approaches are clearly the tools of choice when good templates are available. However, very often only remote templates are available, which might not be good enough to provide reliable models, as above discussed [47]. In unclear situations, a relevant question is which method to choose, or how to efficiently combine these protein-protein docking approaches depending on each specific case [49]. This is even more relevant when modeling multimeric complexes, in which some interfaces might be modeled based on homologous structures, while others would need *ab initio* docking, as above mentioned. An updated version of the InterEvDock2 server [201] can perform template-based docking or *ab initio* docking with evolutionary constraints, depending on the case. But the question is still open about how to efficiently combine template-based and *ab initio* docking when the reliability of the template is unclear. We can obtain some hints from the recent CASP and CAPRI experiments.

In the recent CASP13-CAPRI joint assembly prediction experiment, one of the most efficient approaches was that of Fernández-Recio, based on a combination of template-based and *ab initio* docking followed by pyDock scoring [170], which ranked 2nd and 1st among all the CAPRI predictors and scorers groups, respectively. Models for the subunits were built by CASP-hosted servers. Then, *ab initio* docking was applied in all cases, using

appropriate symmetry constraints or interface restraints from literature. Additionally, when reliable templates were found, template-based models were built by superimposing all possible models of the monomers onto them. After sorting all built models by pyDock scoring, the proportion of template-based and *ab initio* docking models in the final set of submitted models depended on the reliability of the templates (Figure 3.1). The difference with other methodologies was more evident on the "difficult" cases for which no clear template was available. For instance, in T154 *ab initio* docking by pyDock produced the only acceptable models among all participants. In T157, pyDock also produced some of the few successful models of all groups. For scorers, pyDock was used to evaluate all the proposed models, and in case of reliable templates, consistency between energy-based scoring and template-based data was sought.

In 7th CAPRI, predictions using template information were in general successful. Indeed, failing to use available templates, as Fernández-Recio did in T122, T125 interface 1/4, and T133 targets, led to much worse predictions (although interestingly, this group was successful in the latter target, using only *ab initio* docking). This shows that it is critical to choose the optimal docking approach for each case, depending on the template availability. In the rest of targets, templates were used indirectly. In the two protein-peptide targets with good templates (T134, T135), *ab initio* docking with pyDock with restraints from the available templates was successful. In the six protein-saccharide targets (T126-130), *ab initio* docking on the cavity identified from the available templates was also successful. These represent alternative strategies to combine *ab initio* docking with template information. Finally, in the scorers experiment, pyDock got the best performance when considering top 10 predictions, which shows its capabilities to evaluate complex models derived from combined approaches (template-based, *ab initio*, refinement) [176] (Figure 3.3).

3.4.2. Novel methodological developments in protein docking

The most successful approach as predictor in CASP13-CAPRI was that of Venclovas group. They basically used template-based models when reliable templates were found, and free docking with HEX [71] otherwise**.** One of the reasons of their success could be the use of VoroMQA [202] for the evaluation and selection of the final models. However, they were less efficient in the scorers experiment (rank 7th), which might indicate that this function seems mostly optimized for their own pipeline for template-based and docking generation,

while its application to models generated by other sources represents a challenge to be solved. Other successful approach was the use of CONSRANK [203, 204] for the ranking of docking models. CONSRANK is based on the most frequent inter-residue contacts in the ensemble of decoys, and has been updated to Clust-CONSRANK with the addition of a recently developed clustering procedure [205]. The best-performing server in CASP13-CAPRI was HDOCK [206], from Huang's group, who developed a new pairwise shape-based scoring function (LSC) for protein-protein docking to take into account long-range interactions between protein atoms [207].

Other recent new developments in protein docking are RosettaDock 4.0, which shows improved predictions for flexible cases [208], LightDock, using glowworm swarm optimization with NMA-based flexible search [78], or CIPS, a new scoring procedure [209] based on interface propensities from docking calculations. Docking interface propensities have interesting applications, such as interface prediction [95], and more recently, characterization of multi-protein complexes in combination with other evolutionary and physico-chemical properties [210].

3.4.3. Use of external information for integrative docking

The identification of correct docking poses often fails due to intrinsic errors in current scoring functions, incorrect consideration of oligomerization states, or because of multiple interfaces that are not usually included in docking calculations. For all these reasons, the use of external information on a given complex is often critical for successful docking predictions. The pioneering HADDOCK [76], as well as other protein-protein docking methods, such as pyDock [100], ZDOCK [211] or LightDock [212] have developed procedures to include distance restraints to improve the docking calculations. In this line, evolutionary information can be a relevant source of information for docking [213]. Indeed, the most successful docking approach in the recent 7th CAPRI edition was that of the Andreani and Guerois group. The challenging cases of this CAPRI edition encouraged them to go beyond their traditional rigid-body and InterEvScore approach, so they applied different strategies for the inclusion of evolutionary constraints, such as template-based modeling with RosettaCM-based protocol [214], identification of conserved anchoring interface motifs when only remote homologs were available, and covariation-based modeling of interacting subunits in cases in which traditional homology-based modeling would fail [79].

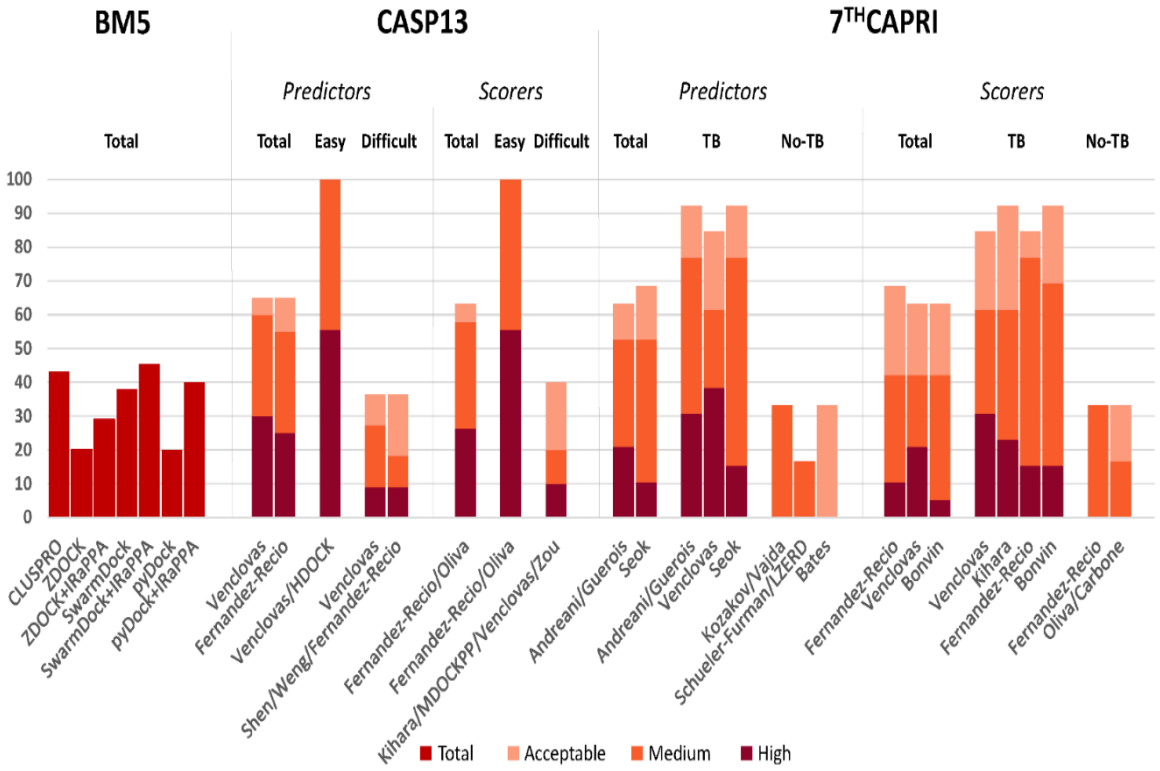


Figure 3.3 Predictive success rates of state-of-the-art docking approaches on different benchmark sets.
ClusPro performance on Protein-Protein Docking Benchmark 5.0 (BM5) is taken from [49]. Performance of other docking methods on BM5 [63] is taken from the rest of results are taken from CASP13-CAPRI and 7th CAPRI blind experiments. Sampling and scoring strategies included, but were not limited to: FFT-based sampling (ZDOCK, FTDock, ClusPro, Weng, Kozakov/Vajda, Shueler-Furman, Venclovas, pyDock, Fernandez-Recio, HDOCK, MDockPP, Zou, Shen, Seok), geometric hashing (Kihara, LZerD), particle swarm optimization (Bates), NMA-based sampling (Shen, Bates), information-driven sampling (Bonvin), energy-based scoring (pyDock, Fernandez-Recio), machine learning-based scoring (IRaPPA, Shen), statistical potentials (Kihara, MDockPP, Zou), evolutionary-based scoring (Andreani/Guerois), Voronoi-based scoring (Venclovas), shape-based scoring (HDOCK), docking-based contact consensus and residue propensities (Oliva, Carbone), and flexible refinement (Schueler-Furman, Seok).

In a broader sense, integrative computational approaches that aim to efficiently use experimental structural data and additional information from a variety of sources for the structural modeling of complexes are becoming increasingly popular [215]. One example is the integration of Small-Angle X-ray Scattering (SAXS) experimental data in *ab initio* docking methods such as pyDock [97-99], HADDOCK [216], PatchDock [217, 218], ATTRACT [219] or ClusPro [220]. And chemical cross-linking data has also been integrated in protein docking methods such as ZDOCK [221]. In the 7th CAPRI

experiment, the use of integrative modeling approaches was blindly evaluated. Targets T150 and T151 were the same complex as T149, a challenging multi-domain dimer, for which SAXS and chemical cross-linking data were provided, respectively. Interestingly, the inclusion of restraints from SAXS data improved the models submitted by pyDock for the original target (with few successful groups), and the cross-linking data further improved pyDock submissions [171].

3.5. Conclusions

The most recent community-wide blind tests on the structural prediction of multi-molecular assemblies and heteromeric protein complexes (including interaction with peptides and saccharides) clearly showed that template availability, as well as any additional information on the complex, are critical for the modelling success. Several groups are focusing their efforts on developing new procedures for efficient integration of template-based and evolutionary information with *ab initio* docking methods, which are producing more accurate and realistic models. Additional methodological developments on protein docking include improvement of scoring functions, and better treatment of conformational flexibility during docking search, but the field is clearly moving towards an integrative analysis and modeling of protein complexes.

The two joint CASP-CAPRI rounds in this CAPRI edition imposed new challenges for our docking and scoring approaches, which were integrated in a broader modeling scheme, including docking, template-based modeling, flexible refinement and experimental restraints. The scoring function from pyDock was particularly successful for the multimeric targets of the last CASP13 edition. As predictors, we should note the good performance of *ab initio* docking in predicting difficult targets in which template-based was not helpful (T154, T157), as well as the improvement in the quality of the submitted models from the integration of template-based and energy-based docking. For the purely CAPRI rounds, we have submitted models for all targets as predictors and scorers, and our server pyDockWeb has participated in all protein-protein and protein-peptide targets. Our scoring scheme pyDock has been used to rank models generated by different approaches (mostly FTDock, ZDOCK, and occasionally LightDock or homology-based models). As predictors, we failed in the most difficult targets but succeeded in the majority of the remaining ones. Interestingly, the use of IRaPPA scoring in T133 was the reason of the

successful predictions. However, when we retrospectively applied this scoring scheme to the rest of protein-protein targets for which standard pyDock scoring did not work well as predictors (T122-125 and T131-132; excluding T136, a challenging multi-molecular assembly, in which the devised *ad-hoc* modeling strategy made it impractical to include IRaPPA scoring of binary complexes), we would have obtained successful predictions within top 10 models only in one target (T131). In most of the other targets, the main problem was the poor quality of the models built for one or both interacting subunits, which IRaPPA scoring cannot help to overcome. The performance as servers for the protein-protein and protein-peptide targets was only slightly worse than that as predictors. As scorers, we had excellent performance when considering top 10 submitted models (best of all participants). However, the results were slightly worse when considering top 5 models. This could be due to the fact that pyDock scoring was optimized in the past for the scoring of rigid-body docking poses, and perhaps it is not fine-tuned for the increasingly refined protein-protein docking models generated by the community. As usual, our participation in CAPRI and the analysis of our results and those of the other groups is an excellent source of inspiration for our next developments in the protein-protein docking field.

4. Computational analysis of protein-protein interactions affected by pathological mutations

4.1. Structural characterization of disease-related mutations involved in protein-protein interfaces

The work of this section has been published in *Int J Mol Sci* [222].

4.1.1. Background

In this chapter of the thesis, we are interested in understanding the molecular-level impact of genetic variants and their relationship to disease. Here, we will discuss about the gap between genotypes and phenotypes. Considering that single nucleotide variants (SNVs) are the predominant source of genetic variations detected in the human population, in this chapter, we will analyze the single amino acid variants (SAVs) caused by the result of these nucleotide substitutions. The main reason to analyze these amino acid changes is because they might be associated with genetic diseases since they can alter protein stability, modulate protein-protein interactions [40, 223, 224], eliminate catalytic activity, affect protein folding [225] or lead to aggregation [226].

Therefore, studying the effects of SAVs on molecular function is crucial, but experimental methods are costly, time-consuming, and challenging, making it infeasible to analyze a large number of amino acid substitutions. Hence, computational tools that rely on conservation-related attributes reflecting structural and functional relevance, as well as on protein structure and stability-related properties following the relationship between structure and function, are used to estimate the phenotypic effect of these variants. Some examples of such reported methods are SIFT [227], CADD [228], PolyPhen-2 [229], PON-P2 [230] or PMut [231]. However, pathogenicity predictors do not accomplish the requirements of clinical applications for standalone tools, not only because they show success rates of around 80% as average, but also because prediction rates for some specific diseases are dramatically lower [232, 233]. Current predictors are not able to capture all the possible effects of mutations at the molecular level. For this reason, a more detailed description of these variants, including information such as their potential involvement in protein-protein interactions (PPIs), would help to improve the prediction of their pathogenic character, providing a more accurate representation of the association between genetic variants and its phenotype by complementing general predictive methods.

Recent studies show that mutations in protein-protein interfaces are over-represented among disease-causing mutations [41, 234, 235]. While common variants from healthy individuals rarely affect interactions, almost two-thirds of disease-associated SAVs perturb PPIs. Half of these pathogenic SAVs are ‘edgetic’ mutations, which impair only a subset of interactions while leaving most others unperturbed [17]. Consequently, within the context of PPI networks, knowledge about the molecular mechanisms by which genetic

variants affect interaction networks can elucidate how mutations on the same gene might cause different phenotypes [224].

Regarding disease-causing mutations at PPI interfaces, they can induce geometrical and physicochemical changes at interaction sites that may affect interface stability, interface conformation dynamics through disruption or stabilization of specific conformational states, and the direct interactions between partner protomers [235]. Thus, the structural location of PPI interface mutations is important concerning pathogenicity. It has been demonstrated that disease-causing SAVs are preferentially located in the solvent-inaccessible interface zones ('core'), as opposed to the interface regions that remain partially solvent accessible ('rim') and are enriched in polymorphisms in the same way as the non-interacting surface. Moreover, energetic hot-spot residues, which play a crucial role in the free binding energy of the complex, tend to be enriched in disease-causing mutations regardless of the interface location [41]. All these findings highlight the importance of understanding the effects of SAVs in protein structure to grasp the genotype to phenotype relationships.

Therefore, in this study, we have characterized protein-protein interactions involving 58 proteins with pathogenic mutations related to diseases detected in newborn screening. We used the experimentally solved structures of the protein complexes when available, and when not, the protein-protein interface was predicted by an *ab-initio* docking approach. The distribution of disease-causing and neutral SAVs across the different interface regions, as well as the substitution susceptibility of distinct amino acids, was discussed.

4.1.2. Materials and Methods

4.1.2.1. Protein interaction and mutational data

Human PPI data and structural information for both protein complexes and their individual components were retrieved from Interactome3D [25]. The Interactome3D database also provides the experimentally solved structures of protein-protein complexes when they are available in the Protein Data Bank (PDB) [236].

Human pathological SAVs data were extracted from UniProt [237] by retrieving variants that corresponded to those ones labeled as "Disease" in the downloadable file

humsavar.txt (see Appendix 2: Table 8.2.1). For neutral variants (see Appendix 2: Table 8.2.2), we used the homology-based model described in Riera et al. [232, 238, 239], where variants were obtained from a multiple sequence alignment (MSA) for each protein family and corresponded to mismatches between the human protein and its close homologs (more than 95% sequence identity with respect to the human protein sequence).

4.1.2.2. Interacting protein analysis

As a further analysis, protein structures involved in each interaction were characterized in more detail regarding sequence identity, structural coverage, domains, and biological assembly (see Appendix 2: Table 8.2.2). Sequence identity and structural coverage were calculated using the UniProt canonical sequence as a reference. Missing loops were not considered in the structural coverage calculations (as opposed to the structural coverage value given by Interactome3D, which includes the missing loops). To identify the protein domains, HMMER3 [240] was used to search against Pfam database [241], based on the canonical sequence. Based on the structural coverage, PPIs could be defined depending on whether the interacting proteins had: (i) global structural coverage greater than 80% in a single PDB file, (ii) global structural coverage < 80% and at least one domain with more than 80% structural coverage, and (iii) global or domain structural coverage < 80% (see Appendix: Table 8.2.2).

4.1.2.3. Experimental protein-protein interfaces

Protein-protein complex structures, when available, were retrieved from PDB based on Interactome3D information. Protein-protein interfaces were defined in a similar way as previously described [120]. Prior to the interface calculation, the sequence and numbering of the PDB structures were extracted and aligned with the corresponding canonical sequence fetched from UniProt database to ensure a correct residue numbering.

Residues were defined as buried if they have a relative Accessible Surface Area in the uncomplexed structure ($rASA_u$) < 0.1, or surface if they have $rASA_u \geq 0.1$. Surface residues were classified as interface residues when the difference in rASA between the uncomplexed and complexed form ($rASA_u$ - $rASA_c$) was > 0, or non-interface surface otherwise. Interface residues were further divided into core and rim. Core was formed by interface residues that were buried in the complex ($rASA_c < 0.1$), and rim was formed by

interface residues that remained exposed in the complex ($rASA_c > 0.1$). The value rASA was computed as the ratio between the Accessible Surface Area (ASA) of a given residue, and the ASA of the corresponding residue type in the extended conformation of the Gly-X-Gly peptide. All (ASA) calculations were done with ICM-Browser (http://www.molsoft.com).

4.1.2.4. Predicted protein-protein interfaces

For selected protein-protein interactions without available protein complex structure, we applied a computational procedure to estimate the interface residues. For this, the uncomplexed structures were retrieved from PDB, taking into account the oligomeric state as defined in the biological unit in the PDB. In this work, *ab initio* protein-protein docking was used to model the PPI when both proteins forming the complex had more than 80% structural coverage.

First, the sequence and numbering of the PDB structures were extracted and aligned with the corresponding canonical sequence fetched from UniProt database, to ensure a correct residue numbering. Then, docking simulations were run with the FFT-based program FTDock 2.0 [242], and the resulting 10,000 rigid-body orientations were rescored by pyDock scoring function, which includes electrostatics, desolvation energy, and a limited van der Waals contribution [80] (Figure 4.1).

From the resulting docking poses, a normalized interface propensity (NIP) was obtained per residue with the built-in *patch* module in pyDock, implementing the pyDockNIP algorithm [96]. A normalized interface propensity (NIP) value of 1 indicates that the corresponding residue is involved in all predicted interfaces of the 100 lowest energy docking solutions, while a value of 0 means that it appears as expected by random. On the other hand, a negative NIP value implies that the residue appears at the low-energy docking interfaces less often than expected by random [96]. Concretely, NIP is based in the calculation of the Averaged Buried Surface (ABS) for each residue from the 100 lowest-energy solutions (N) as follow:

$$ABS_i = \frac{1}{N} + \sum_{k=1}^{N} \left(\frac{ASA_i^{Unb} - ASA_i^{Bnb}}{ASA_i^{Unb}} \right) \quad (4.1)$$

where ASA_i^{Unb} is the solvent-accessible surface area for the receptor residue *i* before ligand binding, ASA_i^{Bnd} is the solvent-accessible surface area for the same residue after ligand

binding according to the docking pose *k*. These *ABS* values were normalized in order to obtain a Normalized Interface Propensity:

$$NIP_i = \frac{ABS_i - \langle ABS \rangle}{ABS^{MAX} - \langle ABS \rangle} \quad (4.2)$$

where $\langle ABS \rangle$ is the average *ABS* value; and ABS^{MAX} is the maximum expected ABS value (maximum is 1) [96].

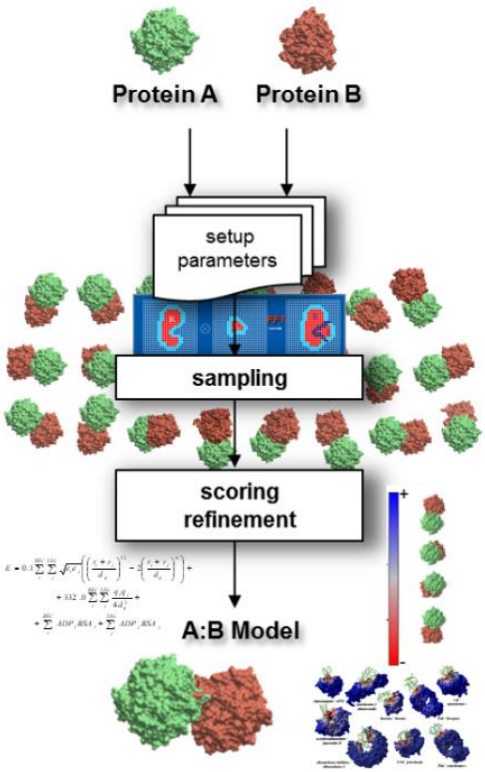


Figure 4.1 Schematic representation of protein-protein docking protocol of pyDock.

Usually, residues with NIP ≥ 0.2 are considered as hot-spot residues when using FTDock, but given the large size of the proteins analyzed here, we used a cutoff of NIP ≥ 0.1 to define the predicted hot-spot residues. These constituted the predicted interface core residues. Then, predicted interface rim residues were built by surface residues located within 10 Å distance from the predicted hot-spot (core) residues [243].

4.1.2.5. Energetic characterization of protein-protein interfaces

The energetic characterization of protein-protein interfaces was performed with the pyDock *bindEy* and *resEnergy* modules. The *bindEy* module computes the total binding energy for

a given protein-protein interaction, based on the complex structure or a model. The r*esEnergy* module calculates the contribution of each individual protein residue to the binding energy for a given protein-protein complex structure.

4.1.2.6. Statistical analysis

The statistical analyses were performed using version 3.4.4 of the R statistical package [244]. The probability of observing a SAV in the protein region *i* is calculated as shown in equation 4.1, where n_i is the number of SAVs observed in the protein region *i*, and N_i is the total number of residues in that region. The likelihood of a SAV to be in region *i* rather than in region j in the protein was expressed then in terms of odds ratio (OR_{ij}). The χ^2 test was used to compare the observed number of SAVs in each region with the expected one if SAVs were distributed according to the number of residues in the different regions. A two-tailed *p*-value < 0.05 indicated statistical significance of the preference for SAVs to be in one region over another. Bonferroni correction was used to adjust the *p*-value for multiple comparisons.

$$OR_{ij} = \frac{x_i/(1-x_i)}{x_j/(1-x_j)} \tag{4.3}$$

where

$$x_i = \frac{n_i}{N_i} \tag{4.4}$$

The amino acid substitution susceptibility to disease-causing SAVs or neutral ones at protein interfaces was calculated, and a two-tailed *p*-value < 0.05 implied statistical significance according to an "*N-1*" χ^2 test.

4.1.3. Results

4.1.3.1. Structural characterization of proteins and interactions in diseases detected in newborn screening

A total of 58 proteins with pathogenic mutations involved in diseases detected in newborn screening were analyzed (see Appendix 2: Table 8.2.1). As many as 56 of these proteins have structural information in Interactome3D, from which 42 have more than 80%

structural coverage (see Appendix 2: Table 8.2.2). Only 16 of these proteins are monomers; 35 are homo-oligomers, and 5 hetero-oligomers. There are experimental structures for 62% of these 42 proteins, while for the remaining 38% of proteins, the Interactome3D database provides a homology-based model.

Among the 58 analyzed proteins, 50 of them have interaction data available in Interactome3D database (three of them form only self-interactions). A total of 389 PPIs were found (<12% with available complex structure), involving 351 protein partners (75% with available 3D structure). Among the partners with known 3D structure, 37% of them have more than 80% structural coverage, while 40% of them have their structure split in separated PDB files.

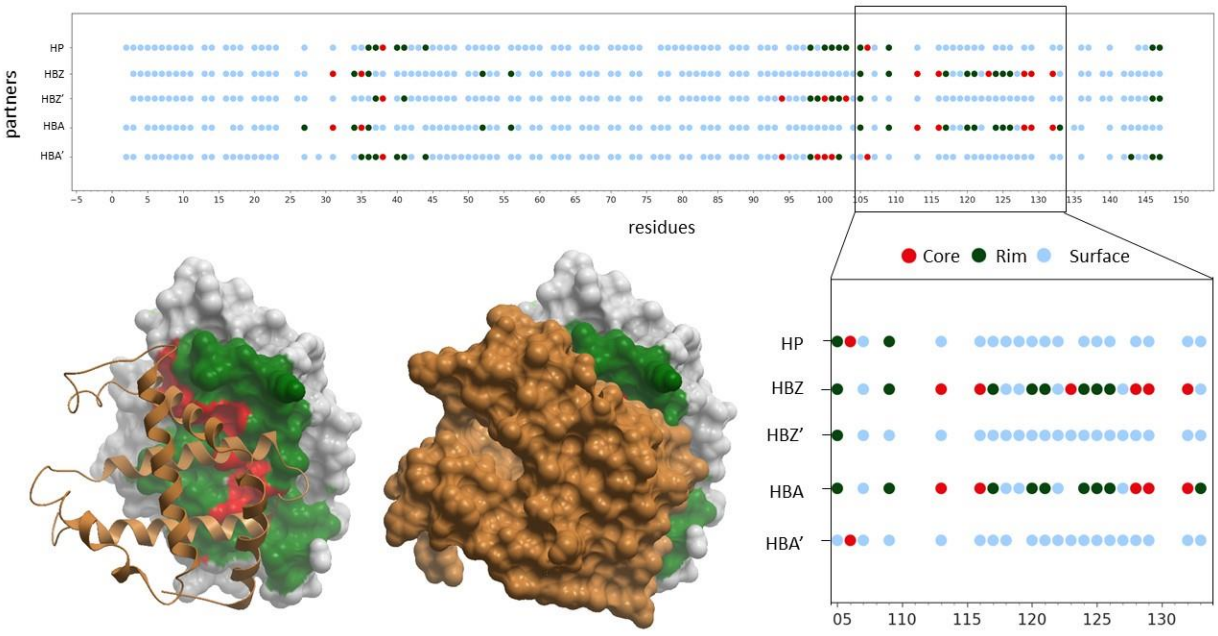


Figure 4.2 Structural characterization of hemoglobin subunit beta (HBB) interactions. The graphic represents the binding interface of HBB with different partners for which there is available complex structure: haptoglobin (HP, hemoglobin subunit zeta (HBZ) and hemoglobin subunit alpha (HBA). As HBB interacts with both HBA and HBZ forming a heterotetramer, two different interfaces are formed with each of the HBB subunit (HBB-HBA and HBB-HBA', or HBB-HBZ and HBB-HBZ'). The graphic represents as dots the non-interacting surface residues (in blue), the interface rim residues (in dark-green) and the interface core ones (in red). The complex structure between HBB (white skin) and HBZ (gold ribbon or skin) is represented, with HBB interface rim residues (in green) and interface core ones (in red).

Protein-protein interfaces were divided into core and rim residues (described in the section 4.1.2.3. of Material and Methods). For a given protein, residues can be defined as surface, interface core or interface rim depending on the considered partner. Figure 4.2 shows an example in which residues have been annotated according to their interaction with five different partners. Interface patches in one protein can be the same for some partners and different for others. This is important since SAVs in these regions could disrupt only a subset of interactions, possibly leading to 'edgetic' effects [17].

In all the annotated PPIs involving the 58 analyzed proteins, there are a total of 11,199 residues, of which 6,019 are found to be buried in at least one structure (Table 4.1). Of the remaining non-buried (surface) residues, 2,062 are located at the interface with at least one protein partner, and of these, 1,146 residues have been found at the interface core at least in one complex.

Table 4.1 Distribution of residues along the different protein regions and odds ratio for disease-causing and neutral SAVs.

Disease-causing SAVs

Region	All residues[1]	Observed[2]	Expected[3]	O/E[4]	Regions	OR[5]	95% C.I.	*p*-value	Adjusted *p*-value
Buried	6019	1842	1.548.96	1.19	**Buried versus Surface**	2.05	1.83-2.28	<0.00001	<0.00001
Surface	3118	552	802.40	0.69	**Core versus Buried**	0.94	0.82-1.09	0.441	1
Rim	916	151	235.73	0.64	**Core versus Rim**	2.11	1.69-2.64	<0.00001	<0.00001
Core	1146	337	294.92	1.14	**Core versus Surface**	1.94	1.65-2.27	<0.00001	<0.00001
Total	11199	2882			**Rim versus Surface**	0.92	0.75-1.12	0.428	1
					Rim versus Buried	0.45	0.37-0.54	<0.00001	<0.00001
					Interface versus Surface	1.44	1.25-1.54	<0.00001	<0.00001

Neutral SAVs

Region	All residues[1]	Observed[2]	Expected[3]	O/E[4]	Regions	OR[5]	95% C.I.	*p*-value	Adjusted *p*-value
Buried	6019	524	834.14	0.63	**Buried versus Surface**	0.29	0.25-0.33	<0.00001	<0.00001
Surface	3118	767	432.10	1.78	**Core versus Buried**	0.82	0.63-1.04	0.105	0.738
Rim	916	178	126.94	1.40	**Core versus Rim**	0.32	0.24-0.43	<0.00001	<0.00001
Core	1146	83	158.82	0.52	**Core versus Surface**	0.24	0.19-0.30	<0.00001	<0.00001
Total	11199	1552			**Rim versus Surface**	0.74	0.61-0.89	0.001187	0.008209
					Rim versus Buried	2.53	2.08-3.06	<0.00001	<0.00001
					Interface versus Surface	0.44	0.38-0.52	<0.00001	<0.00001

[1] Total number of residues in each protein region.

[2] Observed number of residues involving pathogenic (or neutral) SAVs in each protein region.

[3] Expected number of residues involving pathogenic (or neutral) SAVs in each protein region, according to a random distribution based on the total number of residues.

[4] Ratio of observed to expected residues involving pathogenic (or neutral) SAVs in each protein region.

[5] Odds ratio for different possibilities is calculated with a 95% confidence interval and a *p*-value for a two-tailed Fisher's exact test. This *p*-value is adjusted using Bonferroni correction. A *p*-value < 0.05 is considered indicative of statistical significance.

4.1.3.2. Residues energetically relevant for the interaction are more likely to be at the interface core

The energetic contribution to protein complex stability is not uniform across the interface, and for instance, interface residues that are estimated to be energetically relevant for the interaction (i.e. those with residue binding energy < -2 a.u., as calculated by pyDock) tended to be located at the interface core region more often than expected by random (Table 4.2). This is consistent with previous studies reporting that core amino acids contribute significantly more than rim amino acids to the binding free energy of the complex [41, 245]. For this analysis, we have used pyDock module *resEnergy*, implemented in pyDockEneRes web server (https://life.bsc.es/pid/pydockeneres), which provides pyDock binding energy partitioned by residue [128]. There are more details about the methodology and further energetic analysis in the next sections 4.2 and 4.3.

Table 4.2 Distribution of all interface residues and those energetically relevant for the interaction.

Interface region	All residues [1]	Observed low-energy residues [2]	Expected low-energy residues [3]	O/E [4]
Rim	916	201	298.08	0.67
Core	1146	470	372.92	1.26
Total	2062	671		

[1]Total number of residues in the set of PPIs analyzed here in each interface region (core and rim).

[2] Residues with binding energy < −2 a.u., as calculated by pyDock, in each interface region.

[3] Expected number of low binding energy residues in each interface region according to a random distribution based on the total number of residues.

[4] Ratio of observed to expected residues.

4.1.3.3. Pathological and neutral SAVs are differentially distributed in protein-protein interfaces

A total of 2,882 disease-causing mutations and 1,552 neutral variants were mapped onto the 3D structures of the protein-protein interactions involving the 58 genes analyzed here (Table 4.1). Around 47% of all SAVs occurred in solvent-accessible residues, which included non-interacting regions (surface) and interacting ones (interface).

Regarding the disease-causing SAVs, 36% of them (1,040) occurred in solvent-accessible residues, of which 488 were found at the interface with at least one partner (being 337 of them at the core region in at least one complex). The odds of being located at the

interface rather than at non-interacting surface is 1.44 higher for pathogenic SAVs compared to the rest of residues (OR 1.44, 95% CI 1.25-1.54, p=<0.0001), which is consistent with previous studies [39-41, 234]. More specifically, the odds of being located at the interface core region rather than rim is 2.11 higher for disease-causing SAVs compared to the rest of residues, similar to the odds of being located at the interface core rather than at non-interacting protein surface (1.94). On the other side, there is no significant difference between the location of disease-causing SAVs at the interface rim region and the non-interacting protein surface (Table 4.1). These results show clearly the different impact of interface core and rim mutations in human disease.

Regarding the neutral variants, 66% of them occurred in solvent-accessible residues, of which 261 were found at the interface region (being 83 of them at the core, and 178 at the rim), and 767 at the non-interacting protein surface. Contrarily to disease-causing SAVs, the odds of being located at the interface rather than at non-interacting protein surface is smaller for neutral SAVs compared to the rest of residues (OR 0.44, 95% CI 0.38-0.52, p<0.00001). Moreover, for these neutral SAVs, the odds of being located at the interface core rather than the rim or the non-interacting surface is 0.32 and 0.24, respectively. As in the case of disease-causing variants, there is no significant difference between the location of neutral SAVs at the interface rim region and the non-interface protein surface.

The division of the interface into core and rim regions showed that the core is enriched in disease-causing variants, while the rim is enriched in neutral variants. As in the case of the non-interacting protein surface, amino acid changes in the rim region can be easily accommodated without significant distortions in the overall fold of the protein and without affecting the PPIs. This is consistent with their lower evolutionary conservation and higher side-chain flexibility [246]. Figure 4.3 shows the distribution of neutral and pathogenic SAVs in a case example (interactions of the protein HBB). There are five PPIs annotated in Interactome3D for HHB. Four of these PPIs have available structure (or reliable model): HBB-HP, HBB-HBA1, HBB-HBZ, and HBB-HBB. As can be seen, the proportion of pathogenic SAVs located in the interface rim and the non-interacting surface is much lower than the proportion of neutral SAVs in the same regions.

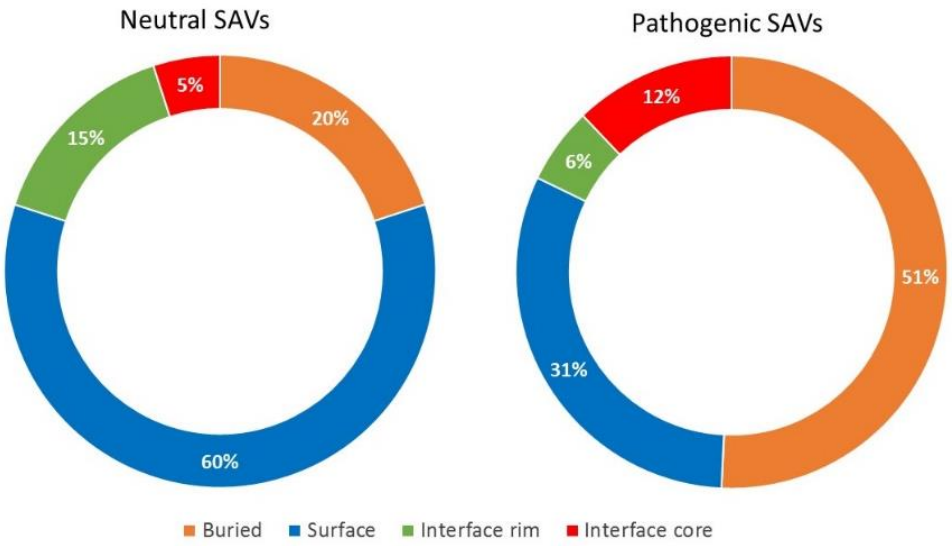


Figure 4.3 Structural characterization of residues in HBB affected by neutral or pathogenic variants.
The graphics show the percentage of HBB residues affected by either neutral or pathogenic variants, as a function of their location in the available protein-protein complex structures.

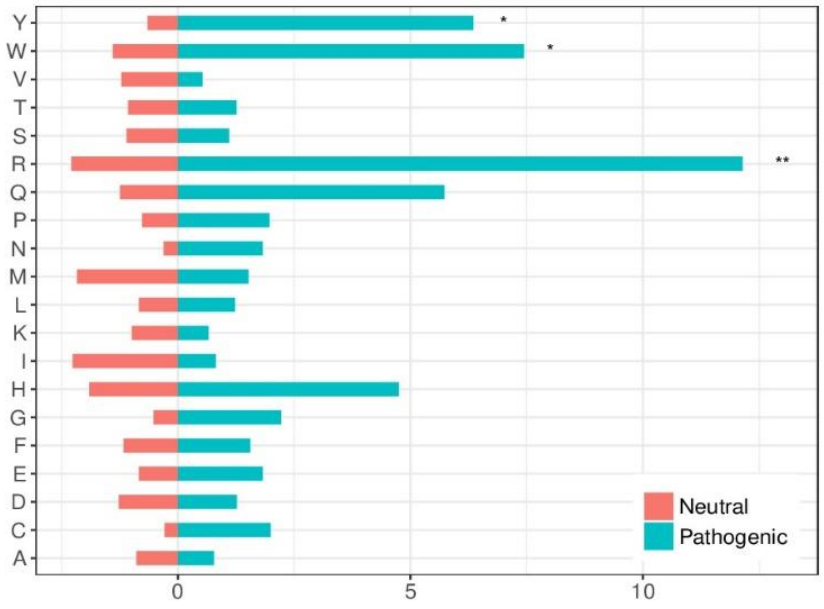


Figure 4.4 Amino acid substitution susceptibility to disease-causing or neutral SAVs within protein interfaces.
The bars show the percentage of residues of a given type that are found mutated in disease-causing SAVs (in blue) or neutral SAVs (in red). Asterisks show statistical significance according to a 'N - 1' X^2 test (* $0.01 < p < 0.05$; ** $0.001 < p < 0.01$).

4.1.3.4. Amino acid substitution susceptibility in the interface is larger in pathogenic SAVs

The amino acids mutability susceptibility was analyzed to determine whether it could be relevant for the molecular characterization of disease-causing SAVs. Arginine (R) was the most mutated residue in both neutral and pathogenic SAVs in protein-protein interfaces. This high mutability can be explained by the fact that four out of the six codons for R include CpG dinucleotides, which tend to mutate at rates 10-15 times higher than other dinucleotides in the DNA [247]. Arginine (R), tryptophan (W) and tyrosine (Y) were significantly over-represented among mutated residues leading to disease (Figure 4.4), which is coherent with previous findings [248].

4.1.3.5. Docking-based interface prediction for further characterization of SAVs: a case study

Many of the proteins analyzed here are involved in protein-protein interactions for which there is no available complex structure. In these cases, we could apply docking simulations to identify potential interface residues, using pyDock approach and NIP interface prediction module. First of all, in order to apply docking simulations, we should check whether we have a complete structure or a reasonable model for the interacting proteins. Available databases such as Interactome3D provide this information. However, there are several issues to consider here. For example, in many cases either a model or an experimental version of the overall structure of the target protein is available but with incomplete structural coverage. Or, it may also happen that the overall structure is split between different PDB files, and then we would need to infer the global structure from these different parts, a non-trivial problem. In Table 8.2.1 on Appendix 2 we identified those proteins that have >80% structural coverage in a single PDB file. This global structure would be suitable for docking. If this global structure is not available because of incomplete coverage or because it is split between different PDB files, we could still use for docking the individual domains that have >80% structural coverage, and then we would try to rebuild the whole protein in the context of the docking models. Since this is out of the scope of automatic docking, we have not used these cases here for docking. In addition to the previous issues, it is important to identify the oligomeric state of the interacting proteins, so that we use for docking the correct assembly form. Table 8.2.1 provides such extended

information of genes and its partners, and Table 4.3 shows the structural information of all 58 genes that is needed for protein-protein docking, including the oligomerization state provided by Eppic [105].

Table 4.3 Genes with pathogenic single amino acid variants involved in diseases detected in newborn screening.

Gene [1]	UniProt [1]	PDB [2]	Glob. Cov. [3]	Domain Cov. [4]	State [5]	Oligom. state [6]	Chains [7]	Eppic Assess.	Chains [8]
ABCD4	**O14678**	6jbj_A	96	98/100	Bound	Homo-2-mer	A:B		
ACADM	**P11310**	4p13_A	92	100/100/100/100	Unbound	Homo-4-mer	A:B:C:D		
ACADS	**P16219**	2vig_E	92	100/100/100/100	Unbound	Homo-4-mer	E:F:G:H		
ACADSB	**P45954**	2jif_A	88	100/100/100/100	Unbound	Homo-4-mer	A:B:C:D		
ACADVL	**P49748**	2uxw_A	87	100/100/100/100	Unbound	Homo-2-mer	A1:A2		
ACAT1	**P24752**	2ib8_D	92	100/100/100	Unbound	Homo-4-mer	A:B:C:D	Homo-2-mer	A:B
ACSF3	**Q4G176**	(3nyq_A)	89	100/93	Unbound	Homo-2-mer (Monomer)	A1:A2 (A)	Homo-2-mer	A_0:A_3
ARG1	**P05089**	3gmz_A	98	100	Unbound	Homo-3-mer	A1:A2:A3		
ASL	**P04424**	1k62_B	99	100/100	Unbound	Homo-4-mer	A1:B1:A2:B2		
ASS1	**P00966**	2nz2_A	98	99	Unbound	Homo-4-mer	A1:A2:A3:A4		
BCKDHA	**P12694**	1olx_A	88	98	Bound	Homo-2-mer	A1:A2	Hetero-4-mer	A_0:A_4:B_0:B:4
BCKDHB	**P21953**	2bev_B	86	100/100	Bound	Homo-2-mer	B1:B2	Hetero-4-mer	B_0:B_4:A_0:A_4
BTD	**P43251**	(4cyf_A)	90	100	Unbound	Monomer	A		
CBS	**P35520**	4l3v_B	90	100/100	Unbound	Homo-2-mer	B1:B2		
CFTR	**P13569**	6msm_A	80	100/91/6/91/100	Unbound	Monomer	A		
CPT1A	**P50416**	(2fy5_A)	77	100/0	Unbound	Monomer	A		
CPT1B	**Q92523**	(1nm8_A)	77	0/100	Unbound	Monomer	A		
CPT1C	**Q8TCG5**	2m76_A	6	100/0	Unbound	Monomer	A		
CPT2	**P23786**	(2deb_A)	95	100	Unbound	Monomer	A		
CYP11B1	**P15538**	6m7x_B	93	100	Unbound	Monomer	A		
CYP17A1	**P05093**	6ciz_C	93	100	Unbound	Monomer	C	Homo-2-mer	C:B
CYP21A2	**P08686**	4y8w_C	89	96	Unbound	Monomer	A		
DBT	**P11182**	1k8m_A / 2coo_A / (2ii3_A)	48	100/100/100	Unbound	Monomer / Monomer / Homo-24-mer	A / A / A1:B1:C1:D1:E1:F1:G1:H1:A2:B2:C2:D2:E2:F2:G2:H2:A3:B3:C3:D3:E3:F3:G3:H3		
ETFA	**P13804**	1efv_A	94	100/100	Bound	Monomer	A	Hetero-2-mer	A:B
ETFB	**P38117**	1efv_B	99	100	Bound	Monomer	B	Hetero-2-mer	A:B
ETFDH	**Q16134**	(2gmh_A)	94	100/100/100	Unbound	Monomer	A		
FAH	**P16930**	(1hyo_A)	99	100/100	Unbound	Homo-2-mer	A:B		
FCGR2A	**P12318**	1h9v_A	54	100/100/100/100	Unbound	Monomer	A		
GALT	**P07902**	6gqd_A	91	98/99	Unbound	Homo-2-mer	A1:A2		
GCDH	**Q92947**	2r0n_A	89	100/100/100	Unbound	Homo-4-mer (Monomer)	A1:A2:A3:A4 (A)	Homo-4-mer	A_0:A_3:A_7:A_10
HADHA	**P40939**	6dv2_G	95	100/100/100/100	Bound	Monomer	G (H)	Hetero-4-mer	A_0:B_0:G_0:H_0
HADHB	**P55084**	6dv2_A	91	100/100/100	Bound	Homo-2-mer	A:B	Hetero-4-mer	A_0:B_0:G_0:H_0
HBB	**P68871**	1dxt_D	100	100	Bound	Homo-2-mer	B:D	Hetero-2-mer	A:B
HCFC1	**P51610**	4go6_D	8	0	Bound	Hetero-8-mer (Hetero-4-	A1:B1:C1:D1:A2:B2:C2:D2 /	Hetero-2-mer	A:B

						mer) (Hetero-2-mer)	A:B:C:D / A:B / C:D		
HMGCL	**P35914**	2cw6_A	91	100	Unbound	Homo-12-mer (Homo-6-mer) (Homo-2-mer)	A1:B1:C1:D1:E1:F1:A2:B2:C2:D2:E2:F2 (A:B:C:D:E:F) (A:B)	Homo-2-mer	A:B
HPD	**P32754**	3isq_A	96	100/100/100	Unbound	Homo-2-mer	A1:A2		
HSD3B2	**P26439**	(3wj7_A)	95	100	Unbound	Homo 3-mer	A:B:C	Monomer	A
IVD	**P26440**	1ivh_A	91	100/100/100/100	Unbound	Homo-4-mer	A:B:C:D		
LMBRD1	**Q9NUN5**	-	-	-	-	-			
MCCC1	**Q96RQ3**	(3u9s_A)	92	100/100/100/100	Bound	Homo-3-mer	A:C:E	Hetero-12-mer	A_0:L_0:K_0:J_0:I_0:H_0:G_0:F_0:E_0:D_0:C_0:B_0
MCCC2	**Q9HCC0**	(3u9r_B)	95	100	Unbound	Homo-6-mer	B1:B2:B3:B4:B5:B6	Homo-6-mer	B_1:B_2:B_3:B_4:B_5:B_0
MLYCD	**O95822**	4f0x_A	92	100/100	Unbound	Homo-4-mer	A:B:C:D	Homo-2-mer	A:B
MMAA	**Q8IVH4**	2www_D	75	89	Unbound	Homo-2-mer	A:B		
MMAB	**Q96EY8**	2idx_C	72	98	Unbound	Homo-3-mer	A:B:C		
MMACHC	**Q9Y4U1**	3sc0_A	84	100	Unbound	Homo-2-mer (Monomer)	A1:A2 (A)	Monomer	A
MMADHC	**Q9H3L0**	5cv0_A	57	62	Unbound	Homo-2-mer	A:B		
MTHFR	**P42898**	6fcx_A	92	100	Unbound	Homo-2-mer	A:B		
MUT	**P22033**	3bic_A	95	99	Unbound	Homo-2-mer (Monomer)	A:B (A)	Homo-2-mer	A:B
PAH	**P00439**	(5den_A)	95	100	Unbound	Homo-4-mer	A:B:C:D	Homo-2-mer	A:D
PAX8	**Q06710**	2k27_A	35	100/0	Unbound	Monomer	A		
PCCA	**P05165**	(3n6r_A)	91	100/100/100/100/100/100/100	Bound	Homo-3-mer	A:C:E	Hetero-12-mer	A_0:L_0:K_0:J_0:I_0:H_0:G_0:F_0:E_0:D_0:C_0:B_0
PCCB	**P05166**	(3n6r_B)	93	100	Bound	Homo-6-mer	B:D:F:H:J:L	Hetero-12-mer	A_0:L_0:K_0:J_0:I_0:H_0:G_0:F_0:E_0:D_0:C_0:B_0
SLC22A5	**O76082**	-	-	-	-	-			
SLC25A13	**Q9UJS0**	4p5w_A	46	0/0/0	Unbound	Homo-2-mer	A:B		
SLC25A20	**O43772**	(1okc_A)	95	100/100/94	Unbound	Monomer	A		
TAT	**P17735**	3dyd_A	85	0/100	Unbound	Homo-2-mer	A:B		
TGFB1	**P01137**	5ffo_H	83	95/94	Bound	Homo-2-mer	G:H	Homo-2-mer	G_0:H_0
TSHR	**P16473**	2xwt_C	31	0/100/100	Bound	Monomer	C	Hetero-3-mer	A_0:B_0:C_0

[1] Name and UniProt code of the corresponding gene.

[2] PDB ID and chain ID of available structure. If no available structure, a model is built (in brackets, PDB and chain ID of the template suggested by Interactome3D). There are two cases without available structure or template. DBT gene has available structure or template for three domains separately.

[3] Total % of protein with structure in the indicated PDB.

[4] % structural coverage of each Pfam-defined domain.

[5] Structures are available for the proteins in their unbound states or bound to other proteins and biomolecules.

[6] Oligomerization state.

[7] Chain IDs by Protein Data Bank (PDB) in the biological unit: A1, A2… corresponds to MODEL1, MODEL2....

[8] Chain IDs by Eppic in its annotation of Biological Unit of PDBs.

Given all the above considerations, we have selected one example case, HADHA, involved in 13 different protein-protein interactions for which there is no complex structure available. There are 116 neutral and 31 pathogenic mutations described in HADHA**.** Figure 4.5-a shows the location of these mutations in HADHA structure. However, they cannot be located at any protein-protein interface due to the lack of this structural information. We explored whether docking-based estimation of interface residues could help to further characterize such mutations. In six of these interactions, interacting partners have sufficient structural coverage (i.e. >80%) for docking. We used pyDock to run docking in these cases, and based on that, we estimated the interface rim and core residues from the NIP calculations. Figure 4.5-b:g shows the predicted interface core and rim residues for each of these interactions, which can be used to visually check whether any of the known SAVs in HADHA are located at the predicted interfaces. Table 4.4 shows in detail the neutral and pathological SAVs in HADHA that are located at the different predicted interfaces. Disease-related mutations R399*, Y740* and V412L, involved in mitochondrial trifunctional protein deficiency, are found in all predicted interfaces. Pathological mutations with more specific effects are Q358K, involved in hemolysis, elevated liver enzymes, and low platelets, and found at the predicted interface with Q14134 (Figure 4.5-d), or R610K, involved in long-chain 3-hydroxy acyl-CoA dehydrogenase deficiency and found at the predicted interface with Q14134 and Q99714 (Figure 4.5-d,e). Additionally, the latest updates in Interactome3D provide the structure of the interaction of HADHA with P55084 (HADHB), PDB 6DV2, which had not been annotated as a protein interacting partner at the time of publishing the article with the structural characterization Figure 4.5-h shows the real interface core and rim residues defined from such structure). Disease-related mutation R235W, involved in mitochondrial trifunctional protein deficiency, is located at the rim region of HADHA/HADHB interface.

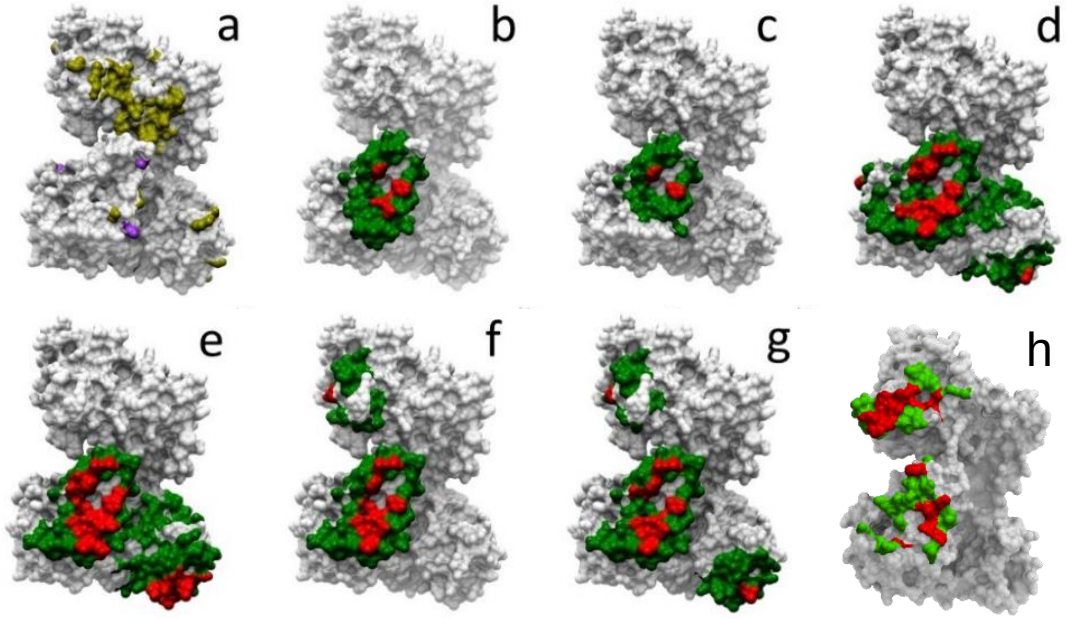


Figure 4.5 Docking-based characterization of HADHA mutations related to protein interactions.
(a) Neutral (yellow) and disease-related (purple) mutations mapped on HADHA structure. In panels (b-g), docking-based predicted interface core (red) and rim (green) residues in HADHA for the interaction with the following partners: (b) O95166, (c) P60520, (d) Q14164, (e) Q99714, (f) Q9GZQ8, and (g) Q9H0R8. (h) Experimentally determined interface of HADHA bound to P55084, available only after publishing the article.

Table 4.4 Docking-based characterization of HADHA mutations.

UniProt [1]	Neutral mutations		Pathogenic mutations [2]	
(partner)	Core [3]	Rim [3]	Core [3]	Rim [3]
O95166	D398G	A396G, K406R	-	**R399***, **V412L**
P60520	-	D398G, K406R	-	**R399***, **V412L**
Q14164	D398G	A396G, K406R, K519R, A596V, S654N, K734Q	-	Q358K, **R399***, **V412L**, R610G, **Y740***
Q99714	D398G, S654N	A396G, K406R, A596V, R645S, R645N, L661I, K734Q	-	**R399***, **V412L**, R610G, **Y740***
Q9GZQ8	D398G	V52I, V526I, N142S, L221I, E223T, I237M, A396G, K406R	-	**R399***, **V412L**
Q9H0R8	D398G	N142S, L221I, E223T, I237M, A396G, K406R, S654N, L661I	-	**R399***, **V412L**
P55084	Q220A,L221I,V222L,L225A,G226R,D398G	E223T	R235W	-

[1] UniProt code of the corresponding interacting partner.

[2] The symbol "*" indicates stop codon. Mutations R399*, Y740* and V412L (in bold) are associated with mitochondrial trifunctional protein deficiency. Mutation Q358K (in italics) is associated to haemolysis, elevated liver enzymes, and low platelets. Mutation R610G is associated to long-chain 3-hydroxyacyl-CoA dehydrogenase deficiency.

[3] Interface core and rim estimated from the docking calculations.

4.1.4. Discussion

To better understand the functional influence of genetic variants at the protein level, structural characterization of single amino acid variants and their interactions is one of the basic steps. In this regard, several structural databases of protein interaction data can be used, such as Interactome3D [25]. However, a major limitation is the low availability of 3D structures for protein-protein complexes. Consequently, a large fraction of SAVs cannot be precisely located at protein interfaces. For this reason, using docking models to estimate whether SAVs can be involved in PPIs may be auspicious. In this regard, a potential problem for the application of docking at large-scale is that most of the available interaction databases essentially provide sets of binary interactions (i.e. protein-protein interacting pairs), while for this type of experiment we would need more detailed data, such as the identification of the contacting domains, or the oligomeric state of the interacting proteins. We have collected all this information for the proteins and interactions analyzed here (Table 8.2.2**).** This can be valuable information in order to run docking simulations in the most realistic conditions. To test this in a real example, we chose HADHA interactions, in which interacting partners had available 3D structure with > 80% structural coverage. We applied protein-protein docking using the available structures of the interacting partners and their biological units in order to predict the binding residues for these interactions. Concerning the mutations found in the newest HADHA/HADHB complex x-ray crystallographic structure, they will be introduced in more detail in the next chapters.

Mutations in the same gene can affect different phenotypic traits (pleiotropy). In this context, the number of interactions and interfaces in a protein is key to understand pleiotropic effects in disease genes. Recent studies show that SAVs located at distinct protein-protein interfaces of the same protein are prone to produce different disease phenotypes [17, 39, 249]. Moreover, it has been demonstrated that one-third of the SAVs produce an 'edgetic' effect, by impairing only a subset of the interactions [17].

In this line, structural analysis of the case example hemoglobin subunit beta (HBB) showed that the same SAV could affect the interaction with different partner proteins if their interface patches are the same, and different SAVs could perturb different partner proteins if these have distinct interface patches. Figure 4.6 shows some of the pathological mutations found in HBB as well as the interaction they impair. For instance, F123S only affects the interaction between HBB and hemoglobin subunit zeta (HBZ); E27A perturbs the interaction between HBB and hemoglobin subunit alpha (HBA); E44Q hampers the

interactions between HBB and both HBA and haptoglobin (HP); C113R affects the interactions of HBB with HBA and HBZ, same as R105W, which also hinders the interaction between HBB and HP.

The extensive analysis of protein interface residues shown here, combining complex structures and docking predictions, demonstrates that pathogenic SAVs are more likely to be located at the interface rather than at the non-interacting surface. More precisely, we found that they are more probable to occur at the interface core region rather than at the rim, in agreement with previous studies [40, 41, 234, 243]. On the contrary, neutral SAVs occur significantly more often in the interface rim as well as in the non-interacting surface, as compared with the interface core region. Furthermore, the residues that contribute the most to the binding free energy of the protein-protein complex (hot-spots) are more likely to be located at the interface core. This is in line with previous studies [41, 245, 250], which revealed that hot-spot residues are not equally distributed among interface regions, but they tend to be clustered within the interface core. Thus, this core region is critical for the stabilization of PPIs; this is reflected in the fact that core residues show a higher level of conservation and coevolution among homologous proteins as compared to those in the rim [246, 251]. This energetical relevance of the core region also explains why SAVs are not as likely to be tolerated there as in the interface rim or the non-interacting surface [41]. We found that arginine, tryptophan and tyrosine are over-represented among disease-causing mutated residues. This is consistent with previous studies reporting that the most frequent hot-spot residues are tryptophan (21%), arginine (13.3%) and tyrosine (12.3%) [110, 252, 253]. Indeed, arginine mutations in interface core residues are not likely to be tolerated and tend to have a profound effect in phenotype [110, 252].

The present study has some limitations, such as the low availability of 3D protein structures or the global consideration in the analysis of both transient and permanent PPIs, which are known to show very different mechanistic, structural and energetic properties. Despite these limitations, this study shows that the structural characterization of PPIs and the analysis of the location of pathogenic and neutral SAVs, together with the identification of the interface residues that are more prone to be mutated and lead to disease, can provide novel information on disease-causing variants. This can be useful in order to characterize SAVs in future studies, interpret them at the molecular level, improve the accuracy of

pathogenicity predictors on new mutations, and help to advance toward precision medicine by helping clinicians to provide adequate diagnosis and treatments.

Further studies with more docking simulations will need to be undertaken. For partners with 3D structures split in different PDBs, a template could be used to model the missing amino acids and join the distinct protein fragments in a correct global 3D structure. Moreover, if these PDBs contain at least one complete domain, docking simulations could be done at the domain level. Homology models could be generated for those proteins without available 3D structure, so that docking can be run afterward in order to find a possible protein-protein interface. This would help to achieve a better understanding of disease at the molecular level since more PPIs could be characterized and more disease-causing and neutral SAVs could be mapped on the structural models.

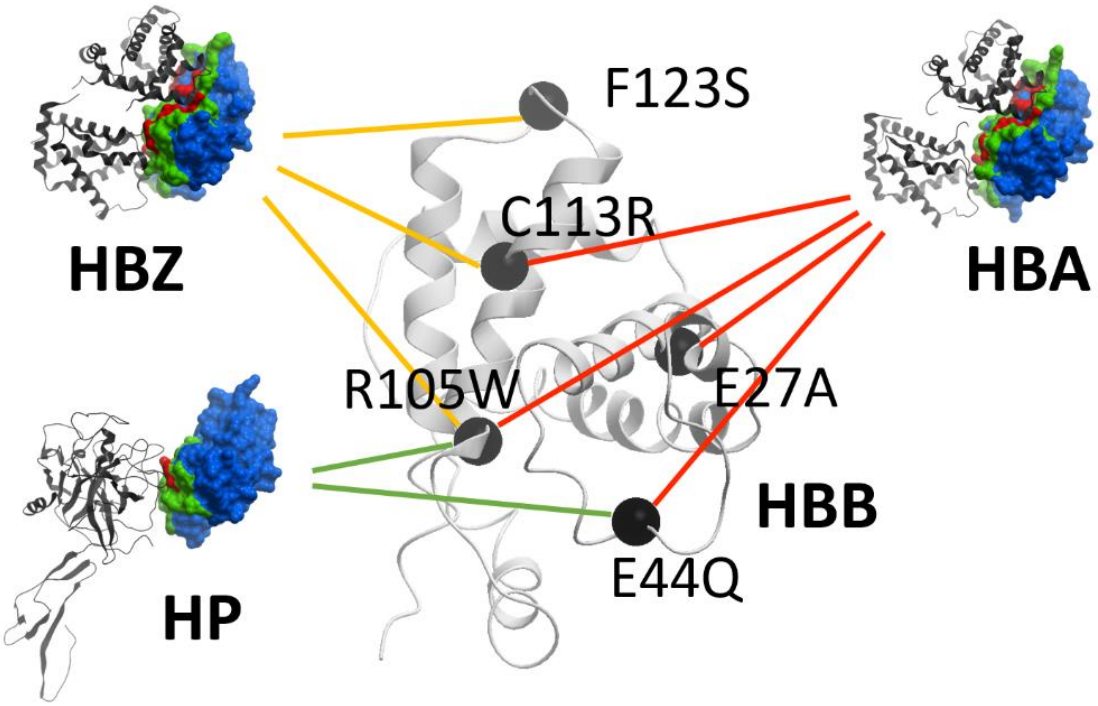


Figure 4.6 Estimation of the effect of disease-causing SAVs in the HBB interaction network based on experimentally solved complex structures.
HBB is represented in white ribbon, with selected pathogenic SAVs in black, linked to the protein-protein interactions in whose interface they are located. The structures of such interactions are represented, showing HBB in white skin with interface rim residues (in green) and interface core residues (in red), and HBB partners (HBZ, HBA, HP) either as ribbon or skin in two different views.

4.1.5. Conclusions

With this study we bring to the table the present limitations facing the structural characterization of SAVs involving PPIs. One is the availability of 3D protein structures is limited, and the knowledge of the interacting domains as well as the oligomeric state of the interacting domains is crucial to understand the structural, mechanistic, and energetic properties of the protein-protein interactions. Despite of such limitations, we have collected all this information for the proteins and their interacting proteins involved in diseases detected in a newborn screening program. Another present limitation is the global consideration in the analysis of both transient and permanent PPIs. Despite these limitations, we characterized PPIs and we analyzed pathogenic and neutral SAVs with the regard to provide novel information on disease-causing variants through locating SAVs on protein-protein interface and identifying interface residues that are more prone to be mutated and lead to disease.

The analysis of protein interface residues, together with protein-protein docking procedures it has been demonstrated that pathogenic SAVs are more likely to be located at the interface rather than at the non-interacting surface. More precisely, we found that they are more probable to occur at the interface core region rather than at the rim. On the contrary, neutral SAVs occur significantly more often in the interface rim as well as in the non-interacting surface, as compared with the interface core region.

The hot-spot residues are more likely to be located at the interface core, then SAVs located at core region have energetic relevance in the interaction. We found that arginine, tryptophan, and tyrosine are over-represented among disease-causing mutated residues.

This work aims to be useful as a path to the characterization of SAVs in future studies, interpret them at the molecular level, improve the accuracy of pathogenicity predictors on new mutations, and help to advance toward precision medicine by helping clinicians to provide adequate diagnosis and treatments.

4. Computational analysis of protein-protein interactions affected by pathological mutations

4.2. Estimation of binding energy changes upon mutation (I): development and validation

Part of the work in this section has been presented as an oral communication at the *SEBBM Congress (2019).*

4.2.1. Background

In the previous section, we discussed the importance of interpreting the phenotypic effects of mutated proteins at the molecular level [222]. To interpret the edgetic effects of mutations on specific protein interactions, we need to know the structural details at atomic resolution of the mutated proteins and all involving protein interactions. In the absence of protein structure for some of the interactions, we can use docking calculations to predict the interface residues.

However, knowing the location of the mutation at the protein-protein interface might not be sufficient to estimate its impact on the interaction. For example, variants H27A and H180A in hPRL/hPRLr interaction, already mentioned in the Introduction (see section 1.6), show a completely different impact on the free binding energy, despite being both located at the protein-protein interface (see Figure 1.5). While hPRL H180A has a strong impact on binding affinity (+2.45 kcal/mol), the hPRL H27A mutation has virtually no effect (-0.07 kcal/mol) [127]. Energetic characterization of the protein interface can indicate the contribution of each residue to the binding affinity, as well as the expected effect when mutated to alanine [128]. If the structure of the protein-protein complex is not available, we can use docking simulations to predict hot-spot residues (a practical application of this will be discussed in more detail in the next section 4.3). But to estimate the energetic impact of a specific mutation, we need to apply computational modeling and energetic analysis. With this purpose, we have analyzed the potential effect of interface mutations by applying diverse *in silico* methods to model the mutations and compute the change in binding affinity, using techniques such as energy-based, machine learning, and empirical forcefields. Then we validated the prediction methods on experimental data from the SKEMPIv2.0 database [88].

4.2.2. Material and Methods

4.2.2.1 Structural and mutational data

For the validation, we selected experimental affinity mutational data from the SKEMPIv2.0 database (https://life.bsc.es/pid/skempi2/), specifically using the interface core and rim residues [88]. SKEMPIv2.0 also provides the experimentally solved structures of protein-protein complexes. A total of 958 human interface core and rim SAVs involved in 103

proteins were evaluated. The binding free energies (ΔG_{bind}) were determined both for wild-type (WT) and mutant by $\Delta G_{bind} = RT \ln Kd$ equation according to the temperature (T) and the experimentally determined equilibrium dissociation constant (Kd) in each case, where R is the ideal gas constant. The change in binding affinity upon mutation ($\Delta\Delta G_{bind}$) was computed as the difference between the ΔG_{bind} of mutant and that of WT.

4.2.2.2 Modeling mutations and estimating changes in binding affinity

Using the cleaned PDB files provided by SKEMPIv2.0, we applied three different computational approaches to model a given mutation and compute its change in binding affinity ($\Delta\Delta G_{bind}$) for each annotated SAV in the database, as follows.

The first method is based on pyDock, a program developed in our group for *ab initio* docking and energy-based scoring. The program uses SCWRL3 [254] to rebuild interacting molecules in case of missing or incomplete side-chains. We implemented here SCWRL3 in an automatic pipeline to model the side-chain of a specific mutated residue within a protein-protein complex structure, while keeping the backbone and the rest of the residue side-chains in the complex as rigid. Then, pyDock can compute the binding affinity of the wild-type (WT) and mutated protein-protein complexes with the *bindEy* module. We also used the pyDock module *resEnergy*, recently implemented in pyDockEneRes web server (https://life.bsc.es/pid/pydockeneres), which computes pyDock docking energy partitioned at the residue level, thus providing a much more detailed description of the docking energy landscape [128]. There are some differences between these two pyDock modules in the way desolvation energy term is computed. Both use atomic desolvation parameters (ADPs) optimized for rigid-body protein-protein docking [255], but while the *bindEy* module (from now on called pyDock for simplicity) is based on the changes in the atomic accessible surface areas (ASAs) between the unbound and the complex, the *resEnergy* module (from now on called pyDockEneRes) calculates ΔASA by a fast contact-based term [128].

The second computational method used is mCSM, a machine learning technique based on the integration of coevolution data, inter-residue interactions, and graph-based signatures [129]. It is only available as a web server, which we used to obtain the $\Delta\Delta G_{bind}$ by uploading the cleaned coordinate file of the protein complex for each given mutation (http://biosig.unimelb.edu.au/mcsm/protein_protein).

The third method is the free available software FoldX, which uses an empirical force field [91]. Default parameters were applied to model the mutation and compute the change in binding affinities.

4.2.2.3 Evaluation of predicted energies

The statistical analyses were performed using version 3.4.4 of the R statistical package [244]. Different cutoff values were applied to the experimental $\Delta\Delta G_{bind}$ data in order to describe the stabilizing and destabilizing effects of SAVs. We used $\Delta\Delta G_{bind}$ cutoffs of -1.5, -1.0 and -0.5 kcal/mol to classify the mutations as stabilizing, and cutoffs of 1.0, 1.5 and 2.0 kcal/mol to classify them as destabilizing.

As for the predicted energies for each mutation, we applied a range of cutoff values to the score provided by each method in order to predict the mutations as stabilizing or destabilizing. Then, for each cutoff value, the predicted classes for the set of mutations were compared with the classes defined by the experimental $\Delta\Delta G_{bind}$ values, from which we computed the True Positives (TP), i.e. experimentally defined stabilizing (or destabilizing) mutations correctly predicted as stabilizing (or destabilizing), False Positives (FP), i.e. experimentally defined as not stabilizing (or not destabilizing) mutations incorrectly predicted as stabilizing (or destabilizing), True Negatives (TN), i.e. not stabilizing (or not destabilizing) correctly predicted as not stabilizing (or not destabilizing), and False Negatives (FN), i.e. stabilizing (or destabilizing) incorrectly predicted as not stabilizing (or not destabilizing).

Then, for each experimental cutoff and each assigned prediction over the entire range of cutoff values, the following metrics were computed: the Precision or Positive Predicted Value (PPV), the Negative Predicted Value (NPV), the Sensitivity or True Positive Rate (TPR), the Specificity or True Negative Rate (TNR), the False Positive Rate (FPR), the False Negative Rate, the Positive Rate (PR) or positives on the whole population, and the Accuracy (ACC) as follow:

$$\textbf{Precision or Positive Predicted Value (PPV)} = \frac{TP}{TP + FP} \tag{4.5}$$

$$\textbf{Negative Predicted Value (NPV)} = \frac{TN}{TN + FN} \tag{4.6}$$

$$\textbf{Sensitivity or True Positive Rate (TPR)} = \frac{TP}{TP + FN} \tag{4.7}$$

$$\textbf{Specificity or True Negative Rate (TNR)} = \frac{TN}{TN + FP} \quad (4.8)$$

$$\textbf{False Positive Rate (FPR)} = 1 - TNR \quad (4.9)$$

$$\textbf{False Negative Rate (FNR)} = 1 - TPR \quad (4.10)$$

$$\textbf{Positive Rate (PR)} = \frac{P}{P + N} = \frac{TP + FN}{TP + FN + TN + FP} \quad (4.11)$$

$$\textbf{Accuracy (ACC)} = \frac{TP + TN}{TP + FN + TN + FP} \quad (4.12)$$

The ROC curves were represented from the already calculated FPR and TPR values.

4.2.3. Results and Discussion

4.2.3.1 Energetic distribution of protein-protein interface mutated residues from SKEMPIv2.0

The 958 mutations from SKEMPI that we analyzed here were classified as stabilizing if $\Delta\Delta G_{\text{bind}} \leq$ -1.0 kcal/mol, or as destabilizing if $\Delta\Delta G_{\text{bind}} \geq$ 1.0 kcal/mol. While these definitions are ultimately arbitrary, these cutoff values are frequently used in other studies to define stabilizing or destabilizing mutations, and they are often reported to experimentally define hot-spot residues. According to this classification, 39.67% of the SAVs in SKEMPI are destabilizing, while only 5.57% are stabilizing (Figure 4.7). In total, at least 45% of the analyzed mutations have an energetic impact on protein-protein interactions. Interestingly, when we analyze the location of those mutations in the complex structures, most of the stabilizing (92%) or destabilizing (80%) mutations are located at the interface core (Figure 4.7).

If we use other $\Delta\Delta G_{bind}$ cutoff values, the number of stabilizing or destabilizing mutations will vary. For instance, if we define destabilizing as those mutations with experimental $\Delta\Delta G_{bind} > 2.0$ or > 1.5 kcal/mol, the % of destabilizing mutations in SKEMPI would be 19.14% or 28.49%, respectively. And if we define stabilizing as those mutations with experimental $\Delta\Delta G_{bind} <$ -1.5 or < -0.5 kcal/mol, the % of stabilizing mutations in SKEMPI would be 4.06% or 10.19%, respectively.

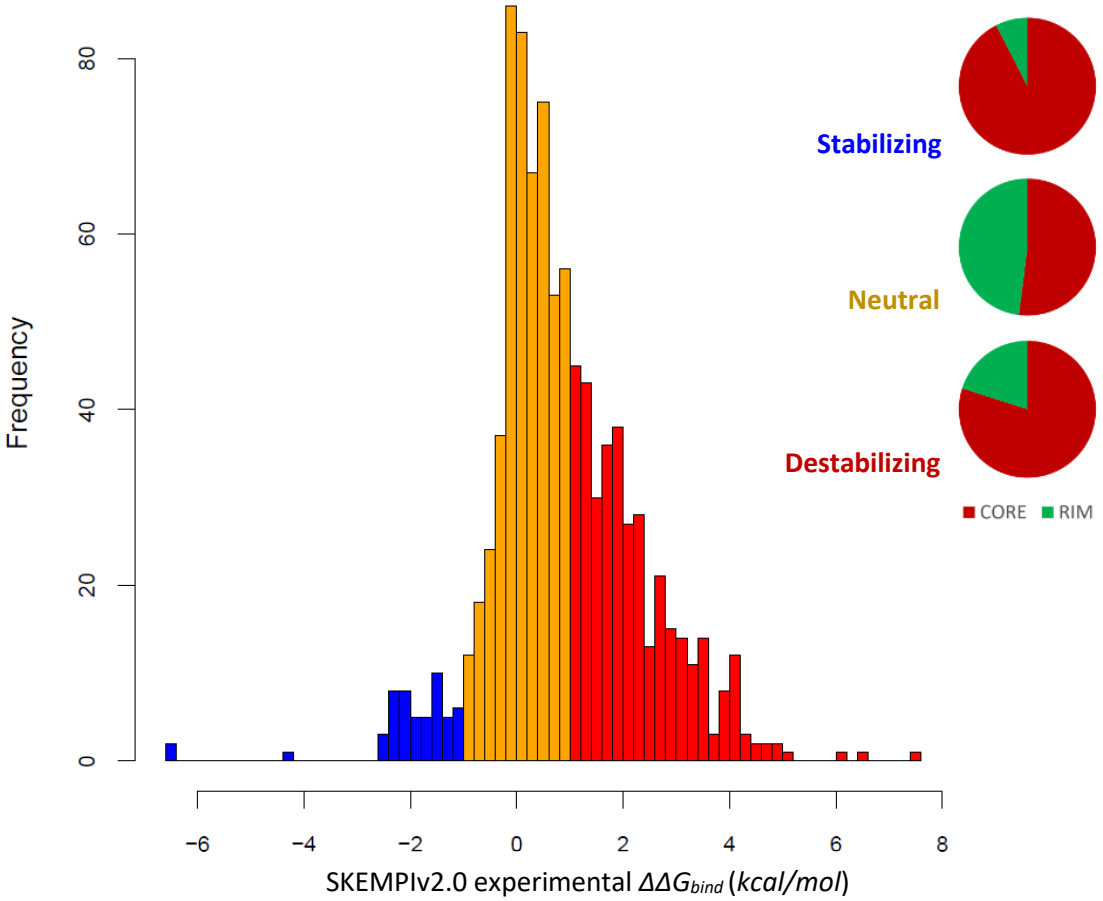


Figure 4.7 Distribution of mutations in SKEMPIv2.0 according to their experimental binding affinity changes
The histogram of frequencies of the mutations according to their experimental binding affinity changes is shown, with stabilizing mutations in blue, and destabilizing in red (neutral are those neither stabilizing or destabilizing, in yellow).
On the right, the percentage of each type of mutation that is located at the interface core or rim regions.

4.2.3.2 Prediction of binding energy changes: correlation analysis

Then, we computed the binding affinity changes for all mutations using different predictors (see Methods). Figure 4.8 shows the comparison between the experimental binding affinity changes in mutations from SKEMPIv2.0 and the predicted $\Delta\Delta G_{bind}$ values obtained by the different methods. Pearson Correlation Coefficients (PCC) are in general low, ranging from $r = 0.31$ (pyDockEneRes) to $r = 0.34$ (FoldX). In a recent study, using a set of mutations to alanine in binary complexes from SKEMPIv2.0 not included in SKEMPIv1.0, the prediction of changes in binding affinity by the different methods showed slightly better correlations: $r = 0.44$ (FoldX); $r = 0.34$ (mCSM); $r = 0.48$ (pyDockEneRes, based only on electrostatics + van der Waals) [128].

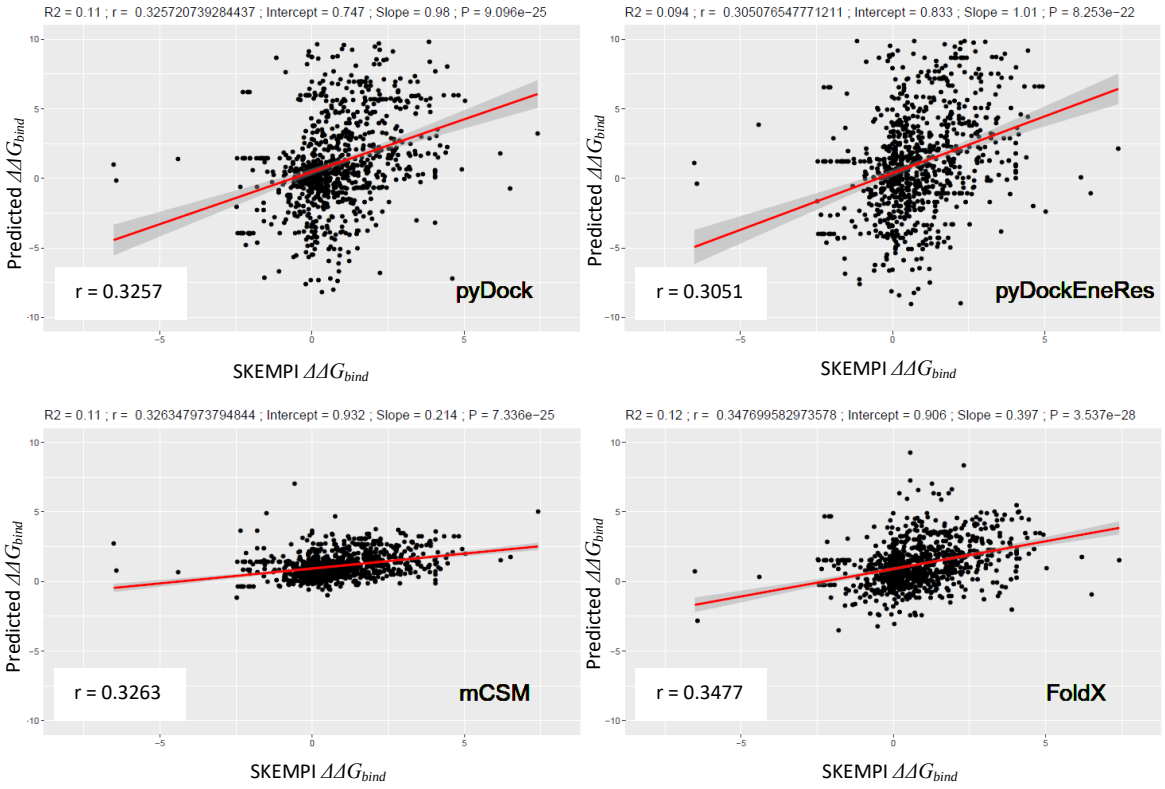


Figure 4.8 Distribution of experimental vs. predicted *ΔΔGbind* for all mutations.
The predicted *ΔΔGbind* values obtained by each method are shown in their corresponding plots. Pearson correlation coefficient and p-value are shown above each plot.

Figure 4.9 shows the distribution of scores (*predicted* $\Delta\Delta G_{bind}$) obtained by the different methods for the experimentally defined mutation classes (stabilizing, destabilizing or neutral). In general, the distributions for the different mutation types are very similar in all methods, indicating poor separation between classes. Nonetheless, some of the destabilizing mutations (represented in red) have higher scoring values for all predictors, and a few of the destabilizing mutations (in blue) have smaller values for some of the predictors, especially for pyDock, pyDockEneRes and mCSM. Thus, despite the low general correlation observed for all methods, we aimed to further explore the possibility of using the scoring values as a classifier to predict the stabilizing or destabilizing mutations.

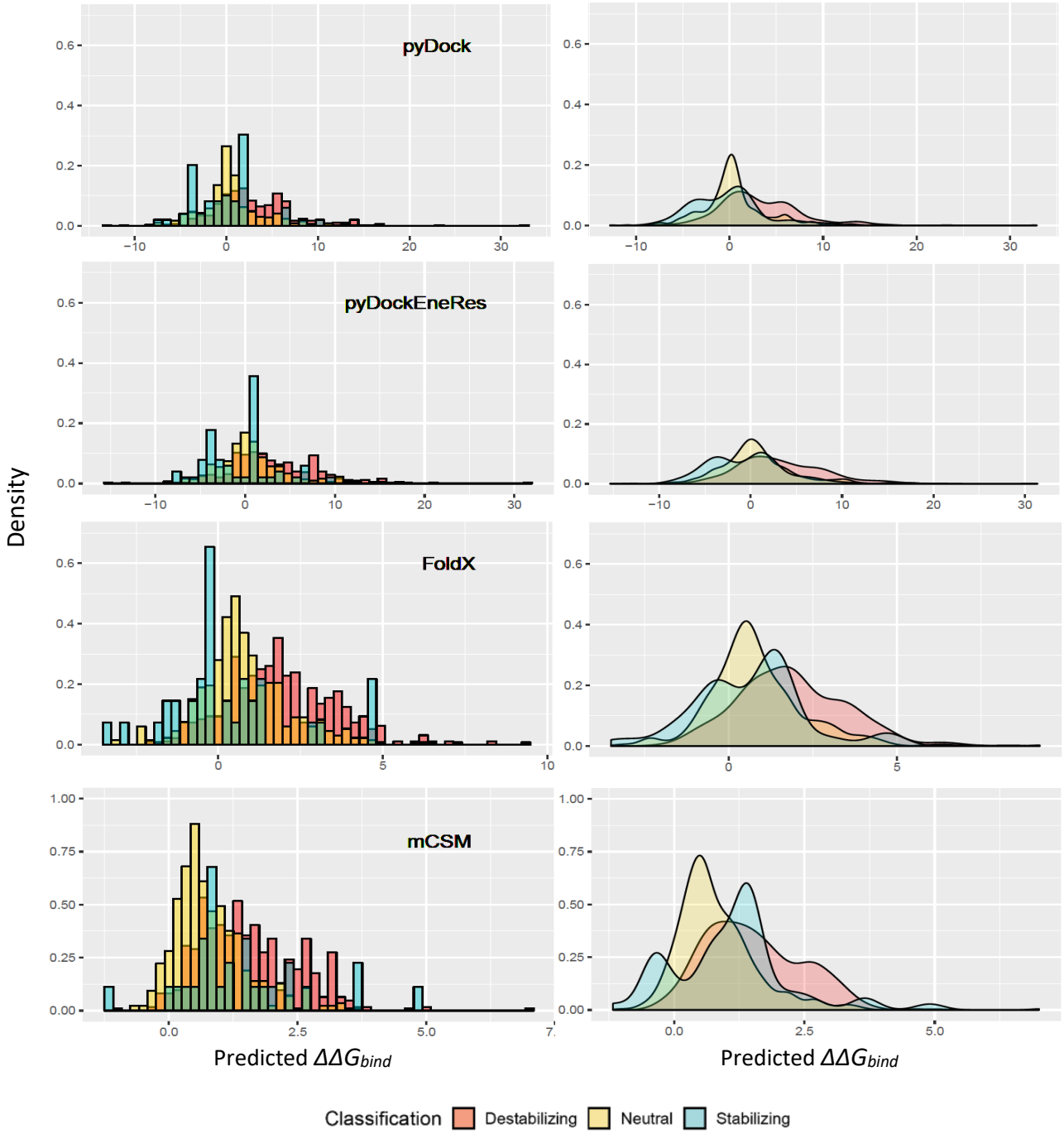


Figure 4.9 Distribution of predicted energy changes for the different mutation types.
Left: histogram of frequencies of the predicted binding energy changes obtained by each method, for the different experimentally determined mutation types: stabilizing (blue), neutral (yellow) and destabilizing (red). Right: smooth line fitted to the histograms only for visualization purposes.

4.2.3.3 Prediction of binding energy changes: a new classifier

Based on these observations, we next explored whether we could use these predictors as a classifier to identify the destabilizing or stabilizing mutations. We explored many different cutoff values applied to the predicted energies from each method to define the predicted destabilizing mutations. These were compared with the experimentally defined destabilizing mutations, and several evaluation metrics were computed, such as the Positive Predicted Value (PPV) and Sensitivity (Equations 4.5 and 4.7) (Figure 4.10, see also Appendix 2: Figure 8.2.1 and Figure 8.2.2). The cutoff value that seems to provide the best balance between PPV and sensitivity is 2.0 a.u. (arbitrary units) for all predictors (Figure 4.10). With this cutoff, pyDock shows a good trade-off between PPV and sensitivity (62.3% and 54.5%, respectively), slightly better than the values obtained by pyDockEneRes (PPV 57.8%, sensitivity 55.8%), FoldX (PPV 64.9%, sensitivity 41.8%), or mCSM (PPV 70.1% PPV, sensitivity 29.1%). The ROC curve indicates that the overall performance of all methods are quite similar (Figure 4.10). From the analysis of each predictor, pyDock achieves the highest PPV of 84.2% with a cutoff of 13.2 a.u., but at very low sensitivity (4.3%). A cutoff of 1.6 a.u. provides more balanced predictions, with 63.5% PPV and 60.9% sensitivity. The highest PPV obtained with pyDockEneRes is 72.4%, with a cutoff of 11.8 a.u., also at very low sensitivity (5.7%). FoldX achieves the highest PPV of 70.7% with a cutoff value of 3.4 a.u., at a better sensitivity (18.9%). Finally, mCSM achieves the highest PPV of 76.0% PPV with a cutoff of 2.4 a.u., at sensitivity 21.3%. Compared to other studies, similar PPV/Sensitivity values are achieved by other predictive methods (i.e., PPV 41%, sensitivity 70% for UEP; PPV 38%, sensitivity 27% for BeAtMuSiC) [256], as well as by the predictive methods used in this thesis (i.e., FoldX shows PPV 37% with 54% of sensitivity; lower values by mCSM with PPV 35% and sensitivity 12% ; pyDock (using FoldX to mutate residues) had PPV 34% and sensitivity 73%) (data from Fig 2 of [256]). All methods have similar values of PPV/sensitivity. The difference between the rates obtained from the compared study is the form to assign positive/negative predictions. We tested different experimental cutoffs for assigning stabilizing/destabilizing mutations, while the mentioned study is based on whether mutations increase/decrease experimental binding affinity.

Regarding the predicted energy changes for the stabilizing variants defined from SKEMPIv2.0 data, the PPV and sensitivity values are, in general, much worse for all predictors (Figure 4.10). Based on that, the best predicted energy cutoff for FoldX and

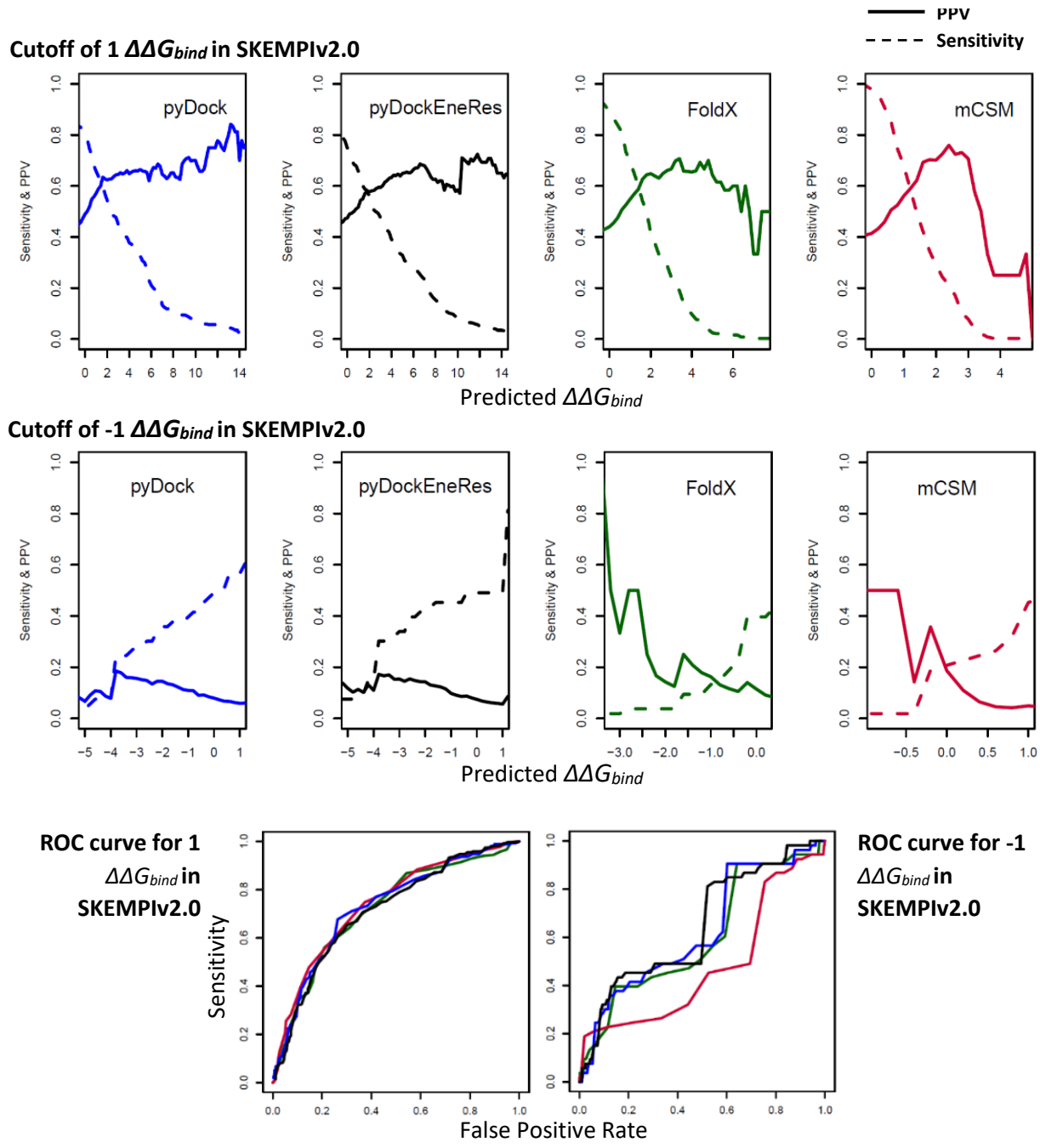


Figure 4.10 Predictive performance of the different methods as mutation classifier. PPV (line) and sensitivity (dashed line) for the prediction of destabilizing (top) or destabilizing (middle) mutations are calculated for a range of cutoff values for the different predicted binding energy changes obtained by each method. Bottom: Predictive performance for the different methods as evaluated by ROC curves for destabilizing (left) and stabilizing (right) mutations.

pyDock/pyDockEneRes seems to be -2.0 a.u. Again, with this cutoff, pyDock and pyDockEneRes show the best balance between PPV and sensitivity (PPV 14.5%, sensitivity 35.8% for pyDock; PPV 14.3%, sensitivity 43.4% for pyDockEneRes), better than FoldX (PPV 14.3%, sensitivity 3.4%). The best cutoff for mCSM seems to be 0 a.u., showing PPV 18.6%, and sensitivity 20.7% (Figure 4.10, see Appendix 2: Figure 8.2.1 and Figure 8.2.2). The ROC curve shows that pyDockEneRes is the best predictive method, and mCSM has the worst predictions (Figure 4.10). From the analysis of each predictor, pyDock achieves the highest PPV of 18.6%, still a low value, with a cutoff of -3.8 a.u., at a sensitivity of 24.5%. Similar results are obtained with pyDockEneRes. FoldX achieves the highest PPV of 100% with a cutoff of -3.4 a.u, but at a very low sensitivity (1.9%). Finally, mCSM achieves the highest PPV of 50.0% with a cutoff of -1 a.u., at a very low sensitivity (1.9%). In any case, the prediction of stabilizing mutations seems to be more difficult than the prediction of destabilizing mutations. Many destabilizing effects are obvious when bulky residues replace smaller ones, strong electrostatic repulsion is introduced, or specific interactions are broken. These effects are easily described even in simple models of the mutation, and this is the reason why most of the methods are able to provide reasonable predictions for destabilizing mutations. However, stabilizing effects can be due to the formation of new interactions, whose correct description requires accurate modeling of the new conformation induced by the mutation, a bit limitation of current modeling methods.

As a further analysis, we wanted to explore whether the predicted stabilizing/neutral/destabilizing classes are enriched in the real type of mutation that corresponds with the prediction. With this purpose, we classified the mutations according to the selected cutoff values for each predictor, and the distribution of experimental values for the mutations within each predicted class was plotted in Figure 4.11. There is significant overlapping between the distributions of the predicted classes by the different methods (Figure 4.11), similar to what we observed when the mutations in the experimentally defined classes were distributed according to their predicted values (Figure 4.9). Nevertheless, a large part of the most destabilizing mutations (as experimentally defined) are correctly predicted as destabilizing. And many of the most stabilizing mutations are correctly predicted as stabilizing. This is reflected in the large sensitivity shown for most of the methods in Figure 4.10. Overall, despite the large overlapping between the predicted classes, there is an important number of predictions correctly classified. The selected

cutoffs are not able to clearly separate mutations in an accurate way, but helped to improve the classification.

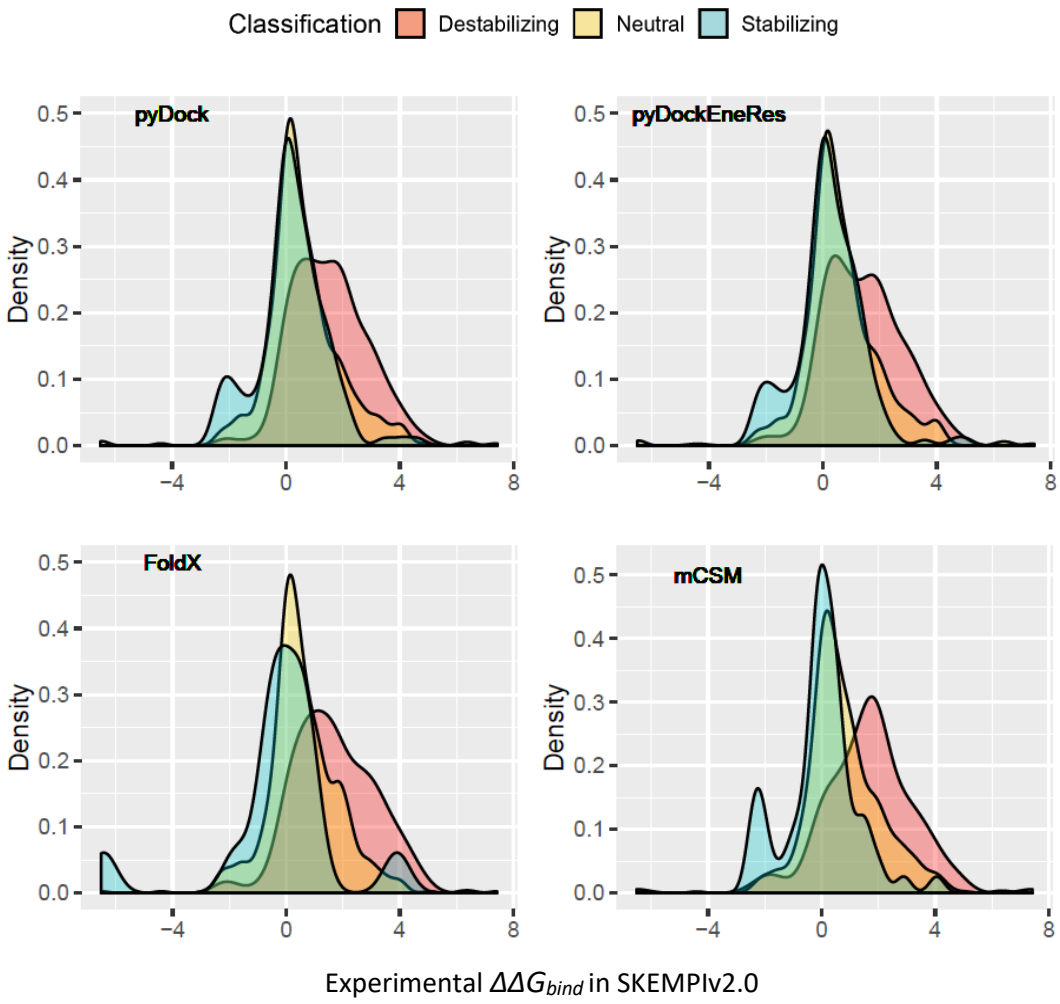


Figure 4.11 Distribution of experimental energy changes for the predicted mutation types by each predictor.
Density plots of experimental energy changes for the predicted mutation types by each method (predicted stabilizing in blue, predicted neutral in yellow, and predicted destabilizing in red), according a cutoff of ΔΔGbind > 2 a.u. for destabilizing mutations, and ΔΔGbind < -2 a.u. (for pyDock and FoldX) or ΔΔGbind < 0 (for mCSM) for stabilizing mutations.

4.2.4. Conclusions

The analysis of experimental binding affinity changes of interface mutations shows a larger proportion of destabilizing mutations in comparison with the stabilizing ones. Both stabilizing and destabilizing mutations are more enriched in the interface core region compared to the interface rim region, while neutral SAVs are indistinctly distributed on the interface.

After evaluating the prediction of binding energy changes in mutations in comparison with their experimental values, all of the studied predictive methods provide a reasonable prediction of destabilizing effects of the mutations, but more limited prediction of stabilizing mutations.

4. Computational analysis of protein-protein interactions affected by pathological mutations

4.3. Estimation of binding energy changes upon mutation (II): application to mutational data

Part of this section has been presented as oral communication at the *SEBBM Congress (2019)*

4.3.1. Background

The phenotypic effects of non-synonymous genetic variations leading or predisposing to disease can be rationalized based on the functional and structural impact in the mutated protein and its interactions. We have demonstrated the crucial impact of pathogenic SAVs in protein-protein interfaces, as well as the important role of oligomerization for understanding such effects at the molecular level. We observed that pathogenic SAVs are more likely to be at the interface core region rather than at the rim (see section 4.1), which is consistent with the fact that the different interface residues may have a varied impact in the interaction, depending on their location and their character. Therefore, as we already discussed in the previous section 4.2, knowing that a given mutation is located at a protein-protein interface is not enough to describe their impact on the interaction. It is also necessary to describe their effect on the binding affinity. In this sense, protein-protein interactions that are clearly stabilized or destabilized by mutations can be potential targets for therapeutic intervention. With this aim, we have explored the use of binding affinity prediction methods on available experimental mutational data in SKEMPIv2.0 [88], and we have devised a protocol to classify SAVs according to their stabilizing/destabilizing energetic effect on specific protein-protein interactions (see section 4.2).

Here, we have applied the protocol developed and validated in the previous section 4.2 to estimate the potential impact of know pathogenic and neutral SAVs on protein-protein interfaces for which there is experimental structural data but no energetic information**.** We analyzed two different sets of mutations potentially affecting protein-protein interactions for which there is an available structure. The first set is formed by SAVs involved in proteins related to diseases detected in newborn screening programs, which were already structurally characterized (section 4.1). The second set is formed by mutations from ClinVar/Humsavar [13, 237] and gnomAD (http://gnomad.broadinstitute.org/) databases that are involved in proteins of the protein-protein docking benchmark version 4.0 (BM4) [257], for which there is available structure both for the complexes and the individual proteins. As we previously discussed, we carefully considered in our energetic analysis the structural details and oligomeric state of the interacting domains, which are crucial to understanding the structural, mechanistic, and energetic properties for the protein-protein interactions.

4.3.2. Materials and Methods

4.3.2.1. Mutational data of protein interaction datasets

Two different sets of mutations were analyzed, annotated in proteins involved in protein-protein interactions for which there is structural data. The first set is formed by SAVs annotated in proteins related to diseases detected in newborn screening programs, which were already structurally characterized (section 4.1). In that study, we identified pathogenic/neutral SAVs located at known protein-protein interfaces [222]. We also included a new released x-ray crystallographic structure (PDB 6DV2) of the interaction between HADHA and HADHB, involved in diseases detected in newborn screening.

The second set is formed by SAVs involved in protein-protein interactions from a subset (only between human proteins or between human and virus proteins) of the protein-protein docking benchmark version 4.0 (BM4) [257]. This set of mutations was compiled during a short stay at the laboratory of Mark Wass at the University of Kent (2017). Human pathological SAVs data were extracted from ClinVar [13] (https://www.ncbi.nlm.nih.gov/clinvar/) and Humsavar (https://www.uniprot.org/docs/humsavar) from UniProt [237], both using the release of June 2017. In addition, we extracted unclassified variants from gnomAD version 1.0 (http://gnomad.broadinstitute.org/) aligned on GRCh37 (discarding pathogenic variants already contained in ClinVar/Humsavar).

4.3.2.2. Experimental protein-protein interfaces

Protein-protein complex structures for the first mutational set were previously retrieved from PDB based on Interactome3D information for proteins related to the newborn screening dataset. Only the available x-ray protein complexes were analyzed, using their annotated biological assemblies [222]. Concerning the second mutational set, the protein-protein complex structures were directly retrieved from the already mentioned BM4. This benchmark set provides the cleaned complex structures, which were mapped and aligned onto the corresponding canonical sequence, as well as that of the unbound proteins.

Interface residues of the first dataset were previously defined from the complex structures, as those ones with difference in relative ASA between the uncomplexed and complexed form ($rASA_u$ - $rASA_c$) higher than 0. Interface residues were further divided into core and rim (see section 4.1.2.3) [222]. We used the same ASA-based criteria for

defining residues at protein-protein interfaces in the second dataset (BM4). As we previously defined, the interface core region was formed by interface residues that were buried in the complex ($rASA_c < 0.1$), and the interface rim region was formed by interface residues that remained exposed in the complex ($rASA_c > 0.1$) (see section 4.1.2.3).

4.3.2.3. *In silico* prediction of binding free energies

Following the protocol developed and validated in section 4.2, we computed the changes in binding affinity ($\Delta\Delta G_{bind}$) upon mutation with FoldX, mCSM, pyDock (*bindEy*) and pyDockEneRes (*resEnergy*). For SAVs in homo-oligomeric complexes with more than two subunits (i.e. involving more than one protein-protein interface), we computed the $\Delta\Delta G_{bind}$ upon mutation independently for each subunit. These values were analyzed independently, since we assumed that the same mutation could have different energetic effects in different interfaces, for instance, in case of asymmetric oligomerization.

4.3.2.4. Energetical classification of predicted $\Delta\Delta G_{bind}$

We used different predicted $\Delta\Delta G_{bind}$ cutoff values in order to predict stabilizing and destabilizing variants, according to the method used. As previously assessed (section 4.2), to predict stabilizing variants we used a cutoff of -2 a.u for the predicted $\Delta\Delta G_{bind}$ values from pyDock and FoldX, and a cutoff of 0 a.u. for mCSM. To predict destabilizing variants, we used a cutoff of 2 a.u for the predicted $\Delta\Delta G_{bind}$ values from all methods.

4.3.3. Results

4.3.3.1. Structural characterization of proteins with SAVs located at protein-protein interfaces

The first mutation set involves 29 homo-oligomeric and 9 hetero-oligomeric protein-protein complex structures. This set was already described in section 4.1, but since new additional x-ray structures were made available after we published the mentioned work [222], we have updated here the structural analysis. These are all the available x-ray structures for the homo- and hetero-oligomers of proteins involved in diseases detected in newborn screening programs (Table 4.3 of section 4.1, see also Appendix 2: Table 8.2.2). A total of 99 protein interfaces were analyzed, including single interfaces for homo-dimeric

and hetero-oligomeric complexes, and multiple interfaces for homo-oligomeric complexes with more than two subunits. As can be seen in Table 4.5, 560 disease-related SAVs were found at the protein-protein interfaces (55% of them at homo-oligomeric interfaces, and 45% of them at hetero-oligomeric interfaces). From the total of disease-related mutation, 67% were found at the interface core, while 33% were found at the interface rim region. The above described analysis counts every mutation only once. However, in cases of multiple oligomeric interfaces, a single mutation can be found repeated at several subunits of the homo-oligomer. If these repeated mutations are considered as different instances (they might be in different environments and have different energetic impact), we would have a total of 1162 disease-related SAVs. On the other hand, 288 neutral variants were found at the protein-protein interfaces (69% of them at homo-oligomeric interfaces, and 31% at hetero-oligomeric interfaces). From the total of neutral mutations, 33% were located at the interface core, and 67% at the interface rim region (Table 4.5). If repeated mutations at multimeric interfaces are considered as different instances, we would have a total of 570 neutral SAVs.

Table 4.5 Distribution of SAVs at the protein-protein interfaces.

	SAVs found at the protein-protein interfaces*								
				In homo-oligomeric complexes			**In hetero-oligomeric complexes**		
Type	Total	Core	Rim	Sub-total	Core	Rim	Sub-total	Core	Rim
Pathogenic	560	377	183	308	212	96	252	166	83
Neutral	288	95	193	199	60	139	89	38	51

* SAVs in multimeric complexes are considered only once

The second mutation set involves hetero-oligomeric complexes from BM4. We have compiled a total of 23,610 single amino acid variants from ClinVar and Humsavar (those labeled as disease-causing SAVs), and gnomAD databases (unclassified SAVs). We selected only variants involving interface core or rim residues, which were found in 38 out of the 54 protein-protein complexes (the subset of BM4 complexes between human proteins, or between human and virus proteins) (see Appendix 2: Table 8.2.3). A total of 67 disease-causing SAVs were found at protein-protein interfaces. 20% of these mutations are located at core regions, and 80% of them at rim regions. Contrary to what we expected, this set of disease mutations have a tendency to be located at the rim region rather than at the core. Regarding unclassified variants, 457 SAVs were found at protein-protein

interfaces, with 25% of these mutations involving core residues and 75% of them involving rim residues.

4.3.3.2. Distribution of the predicted energy values for different types of mutations

We computed the change in binding energy of the interface mutations in the first set, involving proteins related to diseases detected in newborn screening, using the predictors previously validated in section 4.2. Figure 4.12 shows the normalized distributions of these predicted $\Delta\Delta G_{bind}$ values obtained by the different methods for this set of mutations. A high concentration of neutral variants can be observed around 0 value for most of the predictors. Although the distributions of the disease-related and neutral variants show strong overlap, we observed that the disease-related variants had a little bit more abundance of extreme values, especially for mCSM and FoldX predictors.

When calculating the change in binding energy of mutations in hetero-oligomeric complexes, we considered the oligomerization state of each interacting protein as annotated in their biological unit, so the binding energy of the mutated protein in its oligomerization state is computed vs. the rest of proteins in the complex. In the case of homo-oligomers, the binding energy of one subunit is computed vs. the rest of the subunits in the complex. Perhaps the oligomerization type of each complex could affect the calculations. To check this, we further analyzed the predictive results according to the complex oligomerization. The first subset is formed by complexes with unique binary interfaces, i.e. those involved in homo-dimers or hetero-oligomeric complexes, comprising a total of 18 homo-dimers, 1 hetero-trimer, 2 hetero-dimers, 4 hetero-tetramers, 1 hetero-hexamer, and 1 hetero-pentamer. The second subset is formed by complexes with repeated interfaces, i.e. those involved in homo-oligomeric complexes with more than two interacting subunits, comprising a total of 2 homo-trimers, and 9 homo-tetramers. We observed some differences between the energetic distribution of the subsets. The predicted energy distributions of variants involved in complexes with one interface are more similar to those of the global set, while the distributions of variants involved in homo-oligomeric interfaces show some

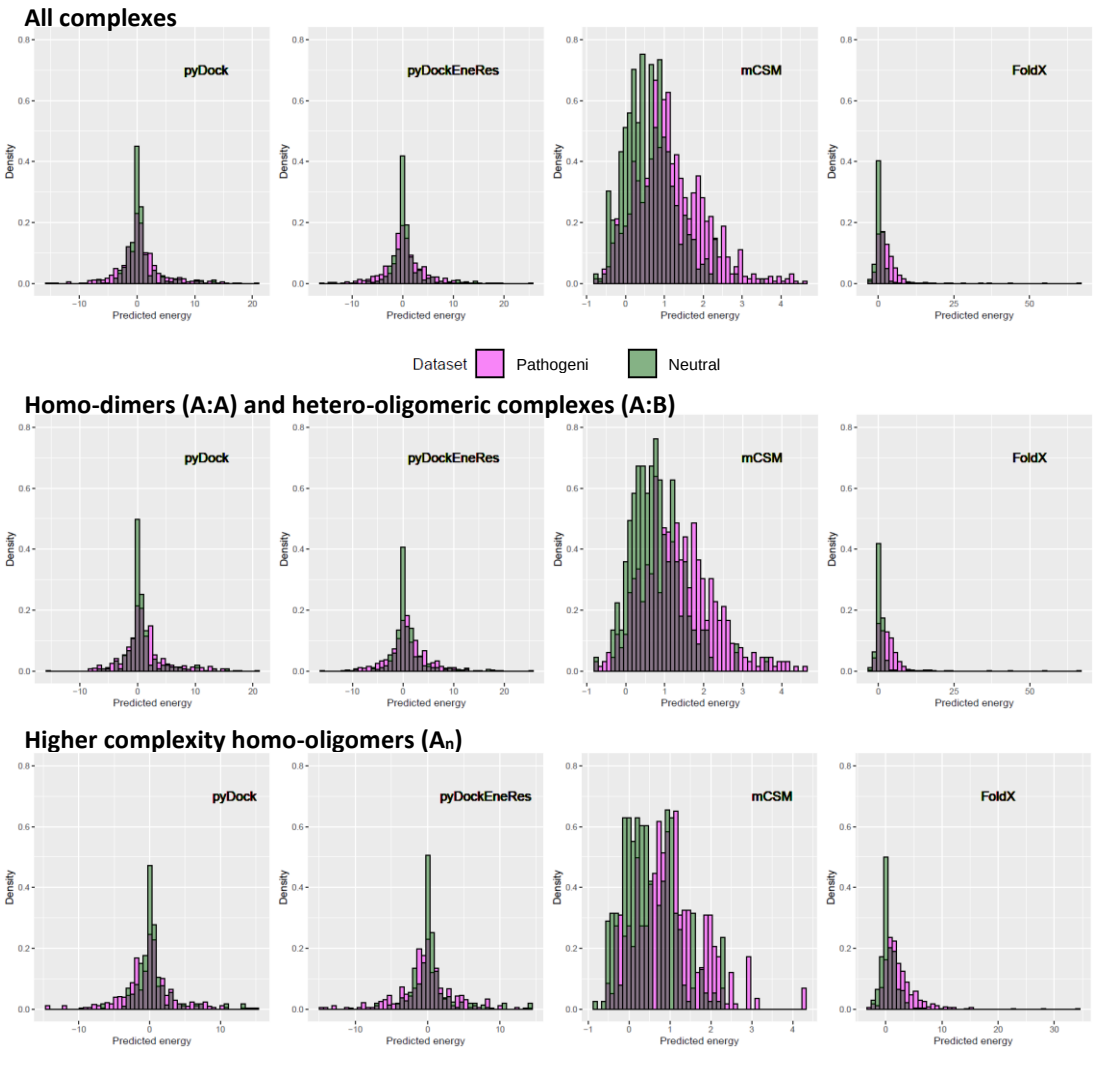


Figure 4.12 Distribution of predicted energy values of SAVs in newborn screening disease-related complexes.
Normalized distributions of predicted energy changes obtained by each method for neutral variants (green) and disease-causing variants (pink). Top panel shows distribution of variants in all analyzed protein-protein complexes. Middle panel shows distribution of variants affecting only single interface complexes. Bottom panel shows distribution of variants affecting complexes with more than one interface (homo-oligomers).

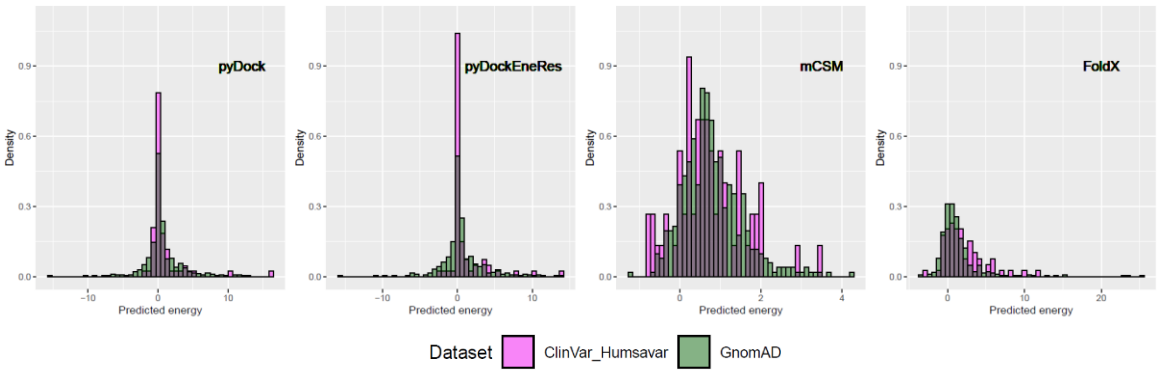


Figure 4.13 Distribution of predicted energy values of SAVs in the interface of proteins involved in BM4.
Normalized distribution of predicted energy changes obtained by each method for unclassified variants (green) and disease-related variants (pink).

changes. We observed that the disease-related variants had more extreme values in homo-oligomeric interfaces (Figure 4.12). The distributions of pyDock and pyDockEneRes energy values for the disease-related variants show more frequent extreme values in the homo-oligomeric interfaces.

Regarding the second mutational set (BM4 complexes), the distributions of the predicted energy changes obtained by each method for the pathogenic ClinVar/Humsavar and the unclassified gnomAD variants (Figure 4.13) have more overlap than the above described distributions for the disease-related and neutral variants of the first mutational set. This might indicate that the pathogenic ClinVar/Humsavar variants cannot be explained in their effects on the protein-protein interactions studied here or that the unclassified gnomAD might include pathogenic variants that have not yet been annotated.

The high overlap existing between the predicted energy values for the disease-related and the neutral mutations could be in part related to inaccuracies in the predicted values, but also could indicate that the pathogenicity of these mutations might arise from energetic effects on other interactions different from the ones studied here. If we consider that there is available structure for only a small fraction of all possible protein-protein interactions, it is highly likely that the mutated proteins studied here are involved in other interactions for which there is no experimental structure and therefore they are not included in the analysis.

4.3.3.3. Energetical characterization of pathogenic SAVs

In the previous section, we aimed to find global differences between disease-related and neutral mutations regarding their predicted energetic effects in known protein-protein interactions, but the results suggest that perhaps only a small fraction of disease-related mutations can be explained in terms of their predicted binding energy changes due to our limited structural information on protein-protein complexes.

Nevertheless, we have explored in more detail how many of the disease-related or neutral mutations are predicted as stabilizing/destabilizing or neutral, according to the different energetic predictive methods evaluated in section 4.2. The results show differences between the neutral and pathogenic variants. (Figure 4.14). A larger proportion of the pathogenic variants related to newborn screening diseases are predicted to be stabilizing or destabilizing in comparison to the neutral variants (mutational set 1).

However, when comparing the pathogenic ClinVar/Humsavar and the unclassified gnomAD variants on the cases of BM4 (mutational set 2), the proportion of these that are predicted to be stabilizing or destabilizing are both similar to that of the neutral variants in the mutational set 1. Perhaps in this second mutational set, the studied pathogenic mutations cannot be explained in terms of their effect on the protein-protein interactions analyzed, which are taken from the BM4, a benchmark that was compiled to test docking methods, not because of their functional implications. It could happen that most of these pathogenic mutations might have an effect in other protein-protein interactions that are not included in our analysis.

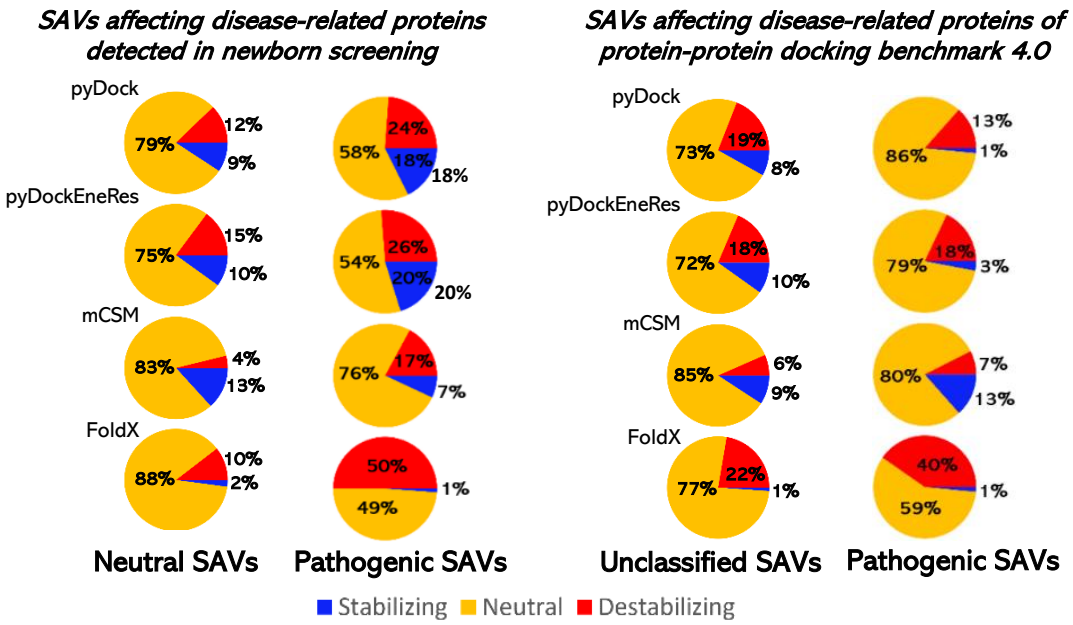


Figure 4.14 Characterization of the energetic effects of SAVs on protein-protein interactions.
Energetic characterization of neutral and pathogenic SAVs in proteins involved in diseases detected in newborn screening (at left) and unclassified and pathogenic SAVs in proteins of protein-protein docking benchmarck 4.0 (at right), according to the predicted energetic effects (stabilizing, neutral or destabilizing) by the different studied methods.

Furthermore, when comparing the predictive capacity of each predictor against the number of annotated pathogenic mutations we observe interesting predictions. For each predictor, Table 4.6 shows the predicted energetic impact of mutations and its distribution according to neutral/unclassified and pathogenic annotated mutations. For mutations with predicted energetic effect on PPI, the % of mutations that are pathogenic in the mutational set 1 (newborn screening set) are 89% (FoldX), 70% (mCSM), 79% (pyDock) and (80% pyDockEneRes). In comparison, 53-66% of mutations that do not have a predicted energetic impact on PPI are pathogenic, which is lower than expected by random (67%).

For stabilizing predictions, 80% of mutations predicted to be stabilizing by pydockEneRes are pathological. For destabilizing predictions, we observe a higher percentage of pathological mutation on predictions affecting energetically PPI (91% FoldX, 90% mCSM, 78% pyDock and 80% pyDockEneRes) (Table 4.6). On the other hand, for the mutational dataset 2 (benchmark 4 set), 21-7% pathogenic mutations have predicted energetic effect on PPI (21% FoldX and 17% mCSM) while 10-15% pathogenic mutations do not have predicted energetic impact on PPI, which is almost than expected by random (13%). These low percentages observed in the latter set could be caused by that most of these pathogenic mutations might have an effect in other protein-protein interactions that are not included in our analysis (see section 4.3.3.2 and Discussion).

Table 4.6 Predicted energetic effects on PPI and its distribution by pathogenicity.

Newborn screening set

	Predicted energetic effect					Non-predicted energetic effect			
	FoldX	**mCSM**	**pyDock**	**pyDockEneRes**		**FoldX**	**mCSM**	**pyDock**	**pyDockEneRes**
Neutral	72	129	142	123	Neutral	498	441	428	447
Pathogenic	594	304	539	483	Pathogenic	568	858	623	679
	Stabilizing effect:								
Neutral	12	107	57	53					
Pathogenic	14	105	234	205					
	Destabilizing effect:								
Neutral	60	22	85	70					
Pathogenic	580	199	305	278					

Benchmark 4 set

	Predicted energetic effect					Non-predicted energetic effect			
	FoldX	**mCSM**	**pyDock**	**pyDockEneRes**		**FoldX**	**mCSM**	**pyDock**	**pyDockEneRes**
Unclassified	107	90	130	125	Unclassified	350	367	327	332
Pathogenic	28	18	14	10	Pathogenic	39	49	53	57
	Stabilizing effect:								
Unclassified	5	60	45	37					
Pathogenic	1	13	2	1					
	Destabilizing effect:								
Unclassified	102	30	85	88					
Pathogenic	27	5	12	9					

4.3.3.4. Application to case studies

We selected six disease-related proteins for further analysis. They have annotated pathogenic mutations that are involved in homo-oligomeric and hetero-oligomeric

interfaces. We wanted to characterize the energetic effects of these pathogenic SAVs on the protein-protein interactions, in order to try to detect possible edgetic variants. For this, we specifically focused on those cases in which at least two of the predictor methods (considering pyDock and pyDockEneRes as a single one) predicted the same stabilizing or destabilizing effects.

ACADM , ETFA and ETFB

This is an interesting case study involving three interacting disease-related proteins that are involved in different types of protein-protein interfaces. We computed the energetical effect of pathogenic SAVs located at interfaces affecting the homo-tetramer of ACADM or the hetero-oligomers of ETFA-ETFB and ACADM-ETFA/ETFB. From the 119 ACADM SAVs annotated as pathogenic (see Appendix 2: Table 8.2.1), we analyzed 16 SAVs located at the homo-tetrameric ACADM interface, and 11 pathogenic SAVs located at the hetero-oligomeric interface between ACADM and the ETFA/ETFB complex (three of the variants, E43K, K395R and T266M are also located at the ACADM homo-oligomeric interface). On the other hand, from 17 ETFA SAVs and 7 ETFB SAVs annotated as pathogenic (see Appendix 2: Table 8.2.1), we analyzed 3 ETFA and 2 ETFB SAVs located

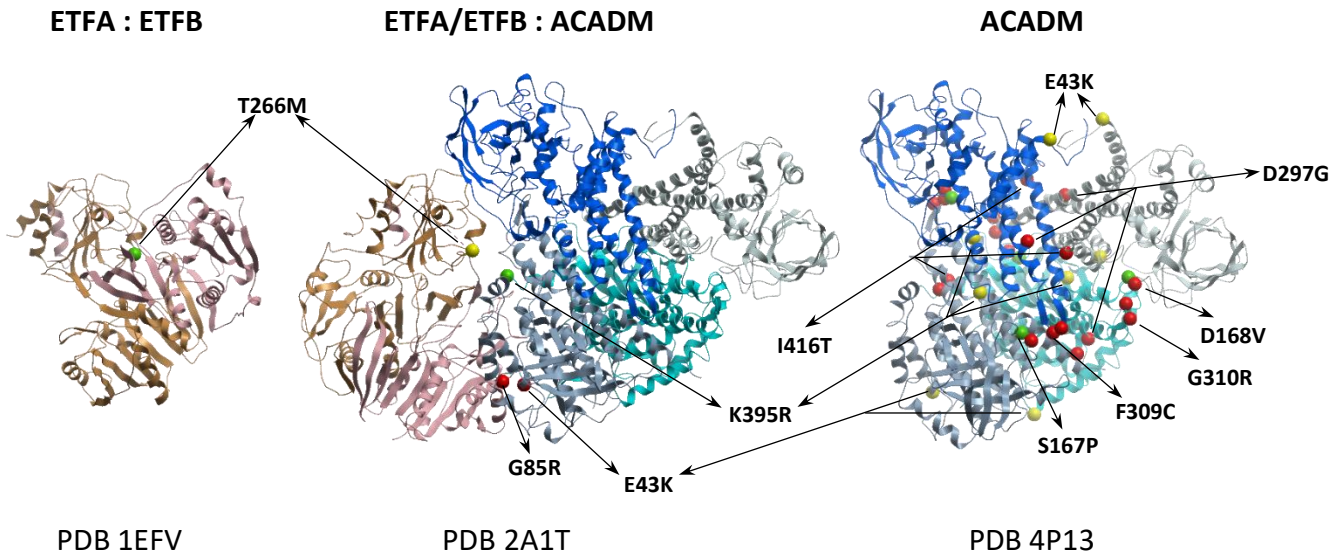


Figure 4.15. Energetically relevant disease-related SAVs at the interfaces of ACADM-ETFA-ETFB hetero-oligomer.
Predicted stabilizing (green spheres), neutral (yellow spheres) and destabilizing (red spheres) variants in at least 2 different computational methods. ETFA (gold) interacts with ETFB (pink). The ETFA/ETFB complex interacts with the homo-tetramer ACADM (represented in 4 different tonalities of blue).

at the hetero-dimeric interface ETFA-ETFB, and 2 ETFA (T266M and R249C are also involved in ETFA/ETFB interaction) and 1 ETFB SAVs located at the hetero-oligomeric of ETFA/ETFB-ACADM homo-tetramer.

Analyzing the energy calculations for each variant on specific protein-protein interactions, we found mutations that showed varying predicted energetic effects on different PPIs. Such is the case of the predicted stabilizing effect of the K395R variant on the hetero-hexamer ACADM-ETFA/ETFB, while its predicted effect on the homo-tetrameric interaction is neutral (Figure 4.15). Concerning mutations affecting specific protein-protein interfaces, five pathogenic variants (A168V, A309C, G310R, D297G, and I416T) were predicted as destabilizing for the interaction between ACADM monomers, while G85R and E43K were predicted as destabilizing for the hetero-hexameric interaction. S167P was predicted as stabilizing for the interaction between ACADM monomers (Figure 4.15). Finally, T226M in ETFA seems to play an interesting role in stabilizing ETFA/ETFB complex and blocking the conformational change necessary to interact with the hetero-tetrameric ACADM, as discussed in more detail in Section 4.3.4 (Figure 4.17). For further information on individual predictions, see Table 8.2.4 of Appendix 2.

HBB protein and its interactions

The next case study is related to HBB homo-dimer, which can interact with HBZ, HBA1, and HP proteins. We computed the energetical effect of pathogenic SAVs located at interfaces affecting the self-assembly or the hetero-oligomer. From the 389 HBB SAVs annotated as pathogenic (see Appendix 2: Table 8.2.1), we analyzed 78 pathogenic SAVs in HBB located at the homo-dimeric interface, besides 78, 80, and 45 pathogenic SAVs in HBB located at the interface with HBZ, HBA and HP. Based on the predicted energetic effects in which at least to methods coincide, only one variant is predicted to stabilize the PPI of the homo-dimeric HBB complex (Table 4.7). For further information of individual predictions, see Table 8.2.4 of Appendix 2. The rim variant H147Y is predicted as stabilizing. Analyzing the predictions involving the hetero-tetramer formed by HBB and HBZ, 41% of variants have been predicted to have effect on the PPI, where four of them were stabilizing (H117Y, H98Y, R105S, R105T) and the remaining ones are destabilizing. 32% of variants have a destabilizing effect on the hetero-tetramer of HBB and HBA1. Regarding the interaction with HP, 24% of variants were predicted as destabilizing (Table 4.7).

Table 4.7 Energetical characterization of pathogenic SAVs affecting HBB PPIs.

	HBB : HBB (A2)[1]		HBB : HBZ (A2:B2)		HBB : HBA1 (A2:B2)		HBB : HP (A:B)	
Variant	**Class[2]**	**Loc[3]**	**Class**	**Loc**	**Class**	**Loc**	**Class**	**Loc**
A116D			**Destabilizing**	CORE	**Destabilizing**	CORE		
A116P			**Destabilizing**	CORE	**Destabilizing**	CORE		
A116V			**Destabilizing**	CORE	**Destabilizing**	CORE		
A129P			**Destabilizing**	CORE	**Destabilizing**	CORE		
A129V			**Destabilizing**	CORE	**Destabilizing**	CORE		
C113R			**Destabilizing**	CORE	Neutral	CORE		
D100A			**Destabilizing**	CORE	**Destabilizing**	CORE	**Destabilizing**	CORE
D100G			**Destabilizing**	CORE	**Destabilizing**	CORE	**Destabilizing**	CORE
D100H			**Destabilizing**	CORE	**Destabilizing**	CORE	Neutral	CORE
D100N			**Destabilizing**	CORE	**Destabilizing**	CORE	Neutral	CORE
D100V			**Destabilizing**	CORE	Neutral	CORE	**Destabilizing**	CORE
D100Y			**Destabilizing**	CORE	**Destabilizing**	CORE	Neutral	CORE
E102A			Neutral	RIM	Neutral	CORE	**Destabilizing**	CORE
E102G			Neutral	RIM	Neutral	CORE	**Destabilizing**	CORE
E102K			Neutral	RIM	Neutral	CORE	**Destabilizing**	CORE
H117P			Neutral	CORE	**Destabilizing**	CORE		
H117Y			**Stabilizing**	CORE	Neutral	CORE		
H98P			**Destabilizing**	RIM	**Destabilizing**	RIM	Neutral	RIM
H98Y			**Stabilizing**	RIM	Neutral	RIM	Neutral	RIM
P101R			**Destabilizing**	CORE	Neutral	CORE		
P125Q			**Destabilizing**	CORE	**Destabilizing**	CORE		
P125R			**Destabilizing**	CORE	**Destabilizing**	CORE		
P37R			**Destabilizing**	RIM	**Destabilizing**	CORE	**Destabilizing**	CORE
P37S			**Destabilizing**	RIM	**Destabilizing**	CORE	**Destabilizing**	CORE
Q128K			**Destabilizing**	CORE	**Destabilizing**	CORE		
Q128P			**Destabilizing**	CORE	**Destabilizing**	CORE		
Q128R			**Destabilizing**	CORE	**Destabilizing**	CORE		
Q132P			**Destabilizing**	CORE	**Destabilizing**	CORE		
R105S			**Stabilizing**	RIM	Neutral	RIM	Neutral	RIM
R105T			**Stabilizing**	RIM	Neutral	RIM	Neutral	RIM
W38C			**Destabilizing**	CORE	**Destabilizing**	CORE	**Destabilizing**	CORE
W38R			**Destabilizing**	CORE	**Destabilizing**	CORE	**Destabilizing**	CORE
W38S			**Destabilizing**	CORE	**Destabilizing**	CORE	**Destabilizing**	CORE
H147Y	**Stabilizing**	RIM	Neutral	RIM	Neutral	RIM	Neutral	RIM
Y146C	Neutral	RIM	**Destabilizing**	CORE	**Destabilizing**	CORE	**Destabilizing**	CORE
Y146D	Neutral	RIM	Neutral	CORE	**Destabilizing**	CORE	**Destabilizing**	CORE
Y146H	Neutral	RIM	**Destabilizing**	CORE	Neutral	CORE	**Destabilizing**	CORE
Y146N	Neutral	RIM	**Destabilizing**	CORE	**Destabilizing**	CORE	**Destabilizing**	CORE

[1] Between parenthesis, the stoichiometry of the complex suggested by the annotated biological unit.

[2] Predicted energetic effect (stabilizing, neutral, or destabilizing) in which at least two methods coincide (considering pyDock and pyDockEneRes as a single method)

[3] The specific location of each variant at the interface core or rim regions. Empty cells mean no variant is located at the indicated protein-protein interface

HBZ and HBA1 share a very similar protein-protein interface, while HP shares only specific regions with them (Figure 4.6 of section 4.1). Table 4.7 shows several examples of edgetic effects of specific variants located in different protein-protein interfaces. Some variants are predicted to have the same destabilizing effect in all involved interfaces: P37R/S, W38C/R/S, A116D/P/V, A125Q/R, Q128K/P/R, A129P/V, Q132P. We should note the predicted destabilizing effect of E102A/G/K mutations at the interface core region

of HBB/HP complex, compared to neutral effect at the core and rim interfaces for interaction with HBA1 and HBZ, respectively. Another interesting variant is C113R at the interface core region with HBA1/HBZ proteins. This variant is predicted to destabilize the interaction with HBZ while is predicted to have no effect for the interaction with HBA. Most of the mutations in tyrosine 146 are predicted to cause destabilizing effects on the interactions with HBZ/HBA1/HP, while there is not apparent effect on the self-assembly. Most mutations on aspartic acid have a destabilizing effect, with the exception of D100H/N, predicted to have a neutral effect on the interaction with HP, and D100V, predicted to have a neutral effect on the interaction with HBA1. Two HBB mutations, H117Y and H98Y, are predicted to have stabilizing effects on the interaction with HBZ, while they are predicted to be neutral on the interaction with HP. On the other hand, the mutation of the same residues to proline was predicted to have a destabilizing effect on the interaction with HBA1, while H117P is predicted to be neutral for the interaction with HBZ, and H98P is predicted to be destabilizing on the interactions with HBA1 and HBZ, but neutral on the interaction with HP. Besides, R105S/T, involved in the interaction with HBZ/HBA1/HP (in all cases located at the interface rim region), has been predicted as stabilizing for the HBB/HBZ interaction. Last, histidine 147 is located at the rim region in all interfaces, with the H147Y mutation predicted as stabilizing for HBB homo-dimer, while there is not predicted any effect on the interactions with the other proteins (Table 4.7).

HADHA and HADHB

The last case study involves HADHB homo-dimer and its interactions. HADHB homo-dimer forms a complex with two monomers of HADHA protein (PDB 6DV2). From 31 HADHA and 35 HADHB SAVs annotated as pathogenic (see Appendix 2: Table 8.2.1), we analyzed 1 pathogenic HADHA SAV located at the protein-protein interface with HADHB, as well as 2 and 6 pathogenic HADHB SAVs located at the interface of homo-dimeric and hetero-oligomeric complexes, respectively. All disease-related SAVs affecting HADHA/HADHB interface (R235W of HADHA and L121P, R229L of HADHB) are predicted to have neutral effect on the interaction. Contrary to HADHB homo-dimer, R174C is predicted to have a stabilizing effect on the PPI, while T133P is predicted to be destabilizing (Figure 4.16). The rest of the mutations involved in the hetero-oligomeric interaction are predicted to have neutral effect (N114D, N114S, N117G, L121P of HADHB).

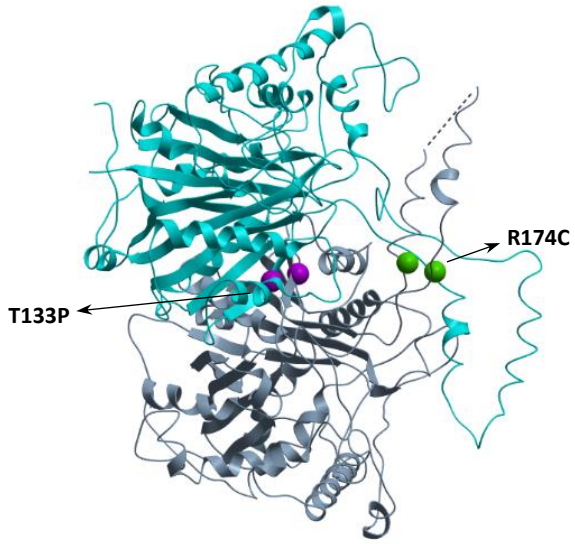


Figure 4.16 Characterization of disease-related SAVs predicted to be energetically relevant for HADHB homo-dimer.
HADHB homo-dimer (PDB 6DV2). The R174C variant, predicted to be stabilizing, is located at the interface rim region (green dots). The T133P variant, predicted to be destabilizing, is located at the interface core region (magenta dots).

4.3.4. Discussion

The low availability of 3D structural information of protein-protein is limiting the number of protein-protein complex structures that we have studied, and as a consequence, the number of SAVs that can be located at an interface. The knowledge of the interacting domains, as well as the oligomeric state of the interacting domains, is also determining the number of variants that can be analyzed. However, to really interpret variants at the phenotypic level, such as the pleiotropic effects in disease genes that are explained by the edgetic effect of mutations on specific interactions, we would need to have structural knowledge of all the interactions for the mutated protein [17, 249]. Despite these structural limitations, we have extended the structural characterization of pathogenic and neutral SAVs from the previous study (section 4.1) and further analyzed another variant set on protein-protein complexes from BM4. In the latter, we unexpectedly found that 80% of pathogenic SAVs at the protein-protein interfaces were located at the rim region. This value is clearly above the expected value by random, considering that in the interfaces of this BM4 there are 51% core residues and 49% rim residues.

This unexpected percentage is probably due to structural limitations, since we are analyzing only a specific number of hetero-oligomeric protein-protein complexes that were compiled in this BM4 set for testing docking programs, therefore not including other

biologically-relevant interactions in which the mutated proteins might be involved. On the contrary, in proteins related to diseases detected in newborn screening that are involved in homo-oligomers and hetero-oligomers, pathogenic SAVs are mostly found at the interface core region, while neutral SAVs are more abundant at the interface rim region, as expected. This is consistent with previous studies that find that SAVs are less energetically tolerant at the interface core region than at the interface rim or at the non-interacting surface [41].

We have applied computational protocols for the energetical characterization of SAVs. Although the distribution for the predicted binding energy values shows a strong overlap between pathological mutations and neutral variants, we observed that the disease-related variants had more extreme values for mCSM and FoldX. This overlap could be due to different reasons. One reason might be that the specific pathogenic SAVs do not have effects on the specific protein-protein interfaces (perhaps they affect to other interfaces for which there is no structural coverage). Another reason might be related to the existing difficulties in predicting binding affinity changes upon mutations, since state-of-the-art methods achieve PCC around 0.34-0.5 (Chapter 4 part 2 section 3.2)[128].

Nevertheless, the results show a small proportion of disease-unrelated (neutral) variants in the newborn screening mutational set that are predicted to be stabilizing or destabilizing in comparison to the pathogenic variants (Figure 4.14). This is not so clear in the mutational data from BM4, in which pathogenic mutations do not seem to have a larger energetic effect on the protein-protein interactions from BM4, perhaps because they might have an effect on other protein-protein interactions different from the ones analyzed here, as above discussed. For mutations with predicted energetic effect on PPI, pyDock/pyDockResEne shows 80% of predicted stabilizing mutations as pathological. On the other hand, mutations predicted as destabilizing by all methods are mostly pathological. On contrary, low percentages observed in the latter set could be caused by that most of these pathogenic mutations might have an effect in other protein-protein interactions that are not included in our analysis (see section 4.3.3.2).

This computational analysis of the energetic effects of SAVs involving protein-protein interacting complexes, besides helping the detection of edgetic variants, can provide substantial information for a better comprehension of molecular mechanisms of diseases involving homo-oligomeric and hetero-oligomeric interfaces. Despite structural limitations for understanding pleiotropic effects, around 30% of mutations have a significant effect on a protein-protein interaction (this percentage is much lower for BM4 ~10%). For those

proteins involved in more than one protein-protein interaction, we detect edgetic effects of SAVs. Few reported studies provide structural and energetical insights of the mechanistic effects of mutations at the molecular level. One study on the pathogenicity of ACADM mutations described the physicochemical changes for a few disease-related SAVs [258]. In the case of E43K, whose impact has been previously discussed in physico-chemical terms [259] our study predicted it as destabilizing in the hetero-oligomeric complex, but neutral in the homo-oligomer. Moderate changes in G310R or I416T have been predicted as destabilizing the homo-oligomeric interaction, while these two variants are not involved in the interaction with ETFA/ETFB. Another pathogenic variant described experimentally is Y64C from CDC42 (1GRN PDB) of BM4, predicted to destabilize the interaction with RHOGAP1. A reported study describes that cysteine substitution affects the interaction with regulatory proteins and effectors, also affecting the catalytic activity of the GTPase and/or its capability to transduce signalling [260]. With our results, we can describe the mechanistic effects at the molecular and energetic levels, providing suitable information helping to interpret the SAVs effects, which can help to improve diagnosis decisions in cases of therapeutic interest.

Three interacting disease-related proteins from distinct oligomeric states are involved in different types of protein-protein interfaces (ACADM-ETFA-ETFB). This is an interesting case to be discussed since the analysis of disease-related SAVs on all possible interactions discloses several mutations with an important impact on the protein-protein interaction. The homo-tetrameric structure of ACADM is predicted to be affected by the destabilizing effect of D168V, D297G, F309P, G310R and I416T, while mutation S167P is predicted to be stabilizing. Besides, the hetero-hexameric structure, formed by the interaction of ACADM and ETFA/ETFB, is predicted to be affected by the destabilizing effect of E43K and G85R. These two variants are predicted to have neutral effect on the tetramer. Thus, we predicted potential edgetic effects of pathogenic SAVs involving ACADM and ETFA/B, which might be involved in medium-chain acyl CoA dehydrogenase deficiency (OMIM: 201450) and glutaric acidaemia type 2 (OMIM: 231680) (Figure 4.15). On the other hand, the ETFA T266M mutation is predicted as stabilizing on the structure of ETFA/ETFB complex, but is predicted to have a neutral effect on the interaction with ACADM. As can be observed in Figure 4.15, the ETFA/ETFB hetero-dimer requires a conformational change to interact with ACADM. Indeed, the

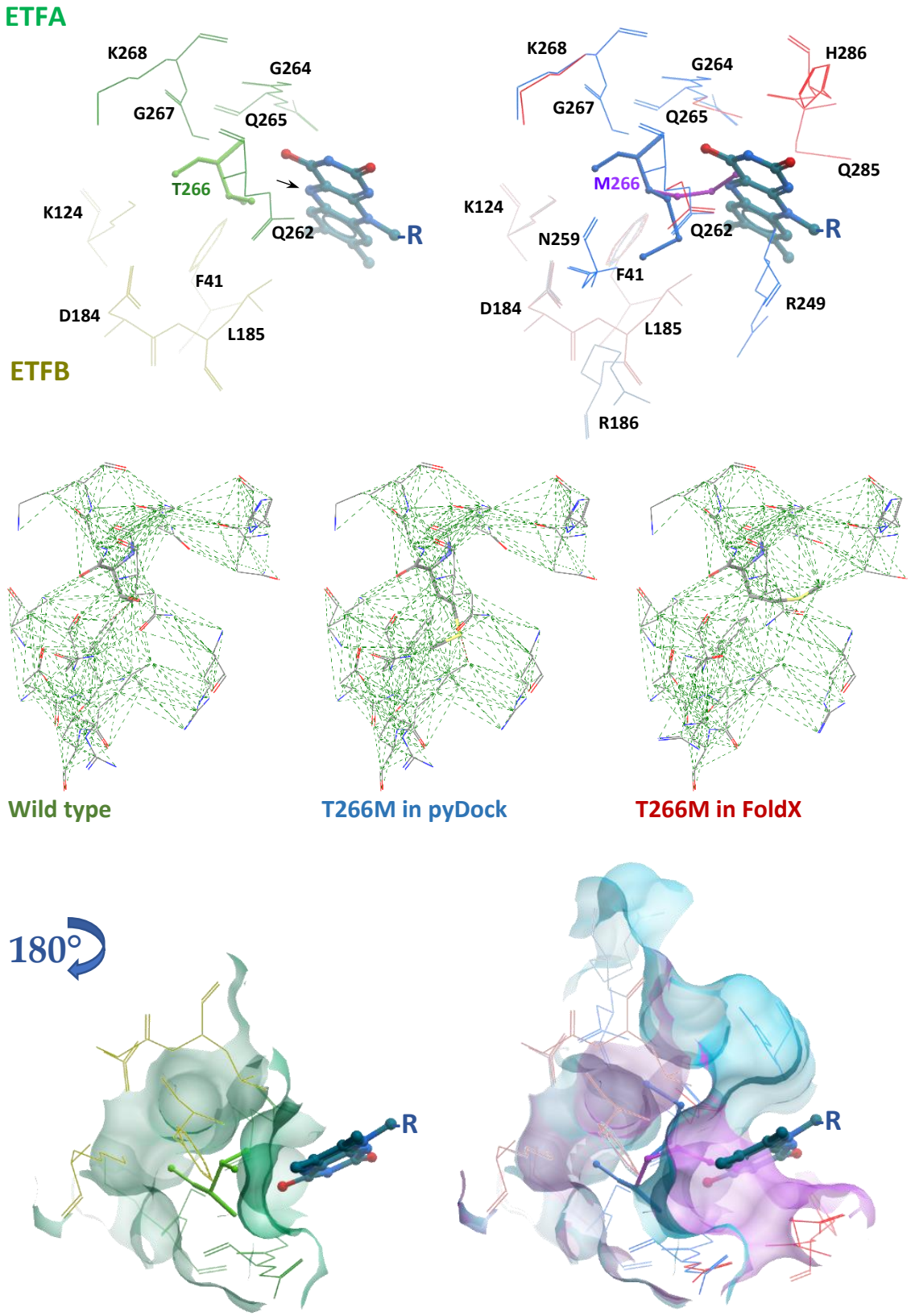


Figure 4.17 Energetic and molecular characterization of ETFA T266M variant
Wild type is represented in green, methionine substitution modeled and scored by FoldX in red, and modeled by SCWRL3 and scored by pyDock in blue. On the top, all residues within 5Å to the mutated residue are shown. ETFB protein is shown in light colors, ETFA in dark colors. Only the flavin group of FAD is shown (row indicates the blocked N5 interacting with T266). In the middle, all contacts involving residue 266 and the residues around it are shown. On the bottom, residues within 5Å to the residue 266 are shown as surface representation. Wild type pocket is occluded in mutated residue.

Cα-Cα distance between ETFB Thr266 and Leu185 ranges from 7 Å in ETFA/ETFA to 27 Å in ACADM/ETFA/ETFB. Then, the T266M variant seems to perturb the interaction with ACADM. We observed a larger number of contacts of this variant with ETFA R249, N259, Q285, H286 residues, and with ETFB F41 and L185 residues (within 5 Å distance) (Figure 4.17). Furthermore, different conformational orientations of the variant revealed a steric hindrance for the interaction with the flavin group of FAD, and the new methionine in T266M variant prevents the interaction of the flavin (Figure 4.17). This is consistent with the low redox activity reported in the literature produced by the methionine in T266M variant blocking the interaction with the N5 of flavin group [261, 262]. This variant is an interesting example showing the interplay between functionality and PPIs.

As above mentioned, the present study has some limitations derived from the low availability of 3D structures of protein-protein complexes, in addition to the difficulties in the description of the conformational states of PPIs. This is especially relevant in BM4 subset, where we limited the specific set of interactions to the ones in the reported benchmark, without exploring other possible protein partners, and without checking alternative oligomerization states. This may be the reason why in this subset there were more SAVs at the interface rim than at the core region. Further characterization of PPIs could be performed by protein-protein docking and modeling procedures, but this step needs a better assessment of the methodologies, oligomerization states, available interface information, etc. for the achievement of accurate protein-protein complex models.

4.3.5. Conclusions

Energetical characterization provides new insights on disease-related variants, for instance to detect potential edgetic effects that could explain the pleiotropic effects by impairing only a subset of the interactions. With our analysis, we detect potential egdetic effects in some variants that could not be explained without exploring the energetic effect on binding.

Despite the limitations of current predicting methods, pathological mutations have higher FoldX and mCSM values compared to neutral variants, and pyDock/pyDockEneRes have the best performance on predicting pathological stabilizing mutations. However, variants from BM4 show no differences between unclassifed and disease-related, which

could be related to the limited set of protein-protein interactions analyzed for these variants, or that some unclassified variants are involved in disease but not annotated.

An accurate consideration of oligomerization states of the interacting molecules and the complexes is essential for a correct energetic description of the mutations and their impact on the interactions, which in turn is key for a successful comprehension of molecular mechanisms of diseases.

This work has contributed to extend the characterization of pathological SAVs affecting protein-protein interactions, beyond structural-only analysis, which can help to identify protein-protein interactions that might be affected by a given pathological situation and therefore, constitute interesting targets for drug discovery purposes.

5. Docking-based identification of small-molecule binding sites at protein-protein interfaces

This work has been submitted to *Comput Struct Biotechnol J*.

5.1. Background

As described in previous chapters, we used computational tools to structurally model protein assemblies, as well as to characterize the effect of pathological SAVs on protein-protein interfaces based on computing the energy changes on binding affinity. Indeed, experimental studies show that many pathological mutations can affect protein-protein interactions, either by disrupting the entire interaction network of the mutated protein, or by specifically affecting some interactions [2, 9, 17]. In this context, PPIs emerge as attractive targets for drug discovery, and the field is shifting the focus of target identification and characterization from individual proteins to interaction networks [139]. Recent examples of protein-protein interfaces as potential therapeutic targets are the mitochondria–endoplasmic reticulum contact sites (MERCs), which are involved in different neurogenerative and metabolic disorders and cancer [263], or the disassembling of the trimeric structure of SARS-CoV-2 spike glycoprotein, which has been proposed as a therapeutic strategy against COVID-19 [264]. Modulation of PPIs with small molecules could thus contribute to new therapeutic developments, and several strategies have been reported for this purpose [140]. A variety of small-molecule inhibitors of PPIs have been identified, most of them by optimization of peptides or natural ligands, fragment-based, high-throughput screening, or rational design [150, 265, 266]. However, the path from a hit to a therapeutic drug is challenging. Indeed, few small-molecule PPI inhibitors have already been approved by FDA or are in clinical trials [2, 267].

Computational identification of small molecules that can modulate PPIs faces important challenges. A major difficulty is the absence of natural pockets in protein-protein interfaces. Contrary to standard ligand design, which usually aims to target a protein cavity (e.g. enzyme active site) that is often well defined and characterized from the structural and even energetic point of view, small-molecule modulators of protein-protein interactions need to target a protein-protein interface that does not usually contain obvious cavities [121, 143]. Indeed, analysis of available structures of PPI inhibitors bound to one of the partner proteins shows that these PPI inhibitor binding sites are less clearly defined than those of enzyme inhibitors [268]. Moreover, in the majority of cases these cavities need to be identified in the 3D structure of the unbound proteins or in that of the protein-protein complex, which is even more challenging, since these pockets may show large conformational rearrangement and thus remain mostly hidden. This imposes a real limitation to the use of structure-based ligand identification tools, such as small-molecule

docking or virtual high-throughput screening (VHTS). To solve this, conformational sampling with computational molecular dynamics (MD) has been reported to identify transient cavities at protein-protein surfaces that are similar to the known PPI inhibitor binding sites [143].

However, among the different transient cavities generated across a given protein-protein interface, it could be difficult to select a suitable cavity for ligand binding, which can be optimally located to compete with a protein interaction [121]. Regarding this, it has been shown that residue composition of protein-protein interfaces is not homogeneous. There are usually a few amino acids (so-called "hot-spots") that contribute to most of the free energy of binding [109]. Targeting such hot-spot residues with a small-molecule could have a significant impact on a protein-protein interaction, which can be exploited for the discovery of PPI modulators [156]. But the identification of hot-spots is not trivial. There is experimental information on the energetic impact of mutations for a limited number of cases [88], but for large-scale applications a variety of bioinformatics approaches have been reported, from statistical analyses to molecular modeling and energetic calculations. Several studies found that hot-spots are enriched in arginine, tyrosine, and tryptophan, whereas leucine, serine, threonine, and valine are less frequent [110, 269]. The number of hot-spots is proportional to the interface size, and they are usually found at the center of the contact interface [121, 270, 271].

Lastly, an important limitation is that the identification of suitable cavities and hot-spots at protein-protein interfaces, as above mentioned, requires the 3D structure of the protein-protein complex and/or the precise location of the interface residues, and this information is not available for the majority of interactions. Indeed, there is available 3D structure for only around 7,500 interactions between human proteins (Interactome3D, https://interactome3d.irbbarcelona.org/, 2020_1 version) [25], a small fraction of the estimated total number of PPIs in human, ranging between 130,000 [44] and 650,000 [45] interactions. In this scenario, computational docking can provide structural models for a protein-protein complex for which there is no available structure, and there are many reported methods using a variety of approaches [65-67, 71, 73, 76-78]. Among the docking methods using energy-based scoring, pyDock [80] has shown excellent predictive results in the most recent CASP-CAPRI and CAPRI assessment experiments [170, 176], and can also be used for the identification of interface residues and hot-spots [95, 96]. Actually, this

is one of the few methods that can identify hot-spots on two specific interacting proteins without requiring the 3D structure of the protein-protein complex.

Here, we have systematically explored the application of these docking-based interface and hot-spot predictions to the selection of transient cavities generated by MD at protein-protein interfaces. The results show that computational hot-spot predictions can help to identify PPI-inhibitor binding sites on protein surfaces in cases in which the structure of the protein-protein complex is not available.

5.2. Materials and Methods

5.2.1. Benchmark set

Benchmark cases were extracted from structural databases of PPIs with known small-molecule modulators: TIMBAL version 1 [164] and 2P2I version 1 [165]. We selected a total of nine non-redundant PPIs (sequence identity < 30%) with available 3D structure for the complex between the small-molecule modulator and one of the interacting proteins, as well as for the protein-protein complex and the unbound proteins (Table 5.1; see Appendix 3: Table 8.3.1 for more details).

Table 5.1 Structural data of PPIs with known modulators

PPI	protein-protein complex PDB	receptor PDB	ligand PDB	PPI-inhibitor PDB
Bcl-xL / Bak	1BXL[c]	1R2D[a]	2YV6	2YXJ
HPV E2 / E1	1TUE	1R6K[a]	(2V9P)[b]	1R6N
IL-2 / IL-2R	1Z92	1M47	(1Z92)[b]	1PY2
HIV Integrase / LEDGF	2B4J	3L3U	1Z9E	3LPU
MDM2 / p53	1YCR[c]	1Z1M	2K8F	4ERF
XIAP BIR3 / Caspase	1NW9	1F9X	1JXQ	1TFT
XIAP BIR3 / Smac	1G73	1F9X	1FEW	2JK7
TNFR1A / TNF-β	1TNR	1EXT	(1TNR)[b]	1FT4
ZipA / FtsZ	1F47[c]	1F46	(2VAW)[b]	1S1J

[a] Missing loops were built with MODELLER 9v10

[b] Structural model built with MODELLER (template PDB ID is indicated in brackets)

[c] Only structure for protein-peptide complex is available

Six of the nine benchmark cases are related to human signalling pathways involved in cell growth and death. Bcl-xL plays a relevant role as a regulator of apoptosis, since its interaction with Bak activates the mitochondrial apoptotic process [272]. Several inhibitors are known for this interaction, many of them with available structure for the protein-inhibitor complex (Table 8.3.1). The best known inhibitor is ABT-737 (813.4 Da), which has been approved by FDA for the treatment of cancer [273, 274]. X-linked inhibitor of apoptosis protein (XIAP) is also an important regulator of apoptosis. XIAP BIR3 domain binds caspase 9 to prevent the formation of its active dimeric form, thus inhibiting its enzymatic apoptotic initiator activity. XIAP activity is inhibited by binding of Smac protein, which promotes apoptosis [275, 276]. Several small molecules are known to target XIAP, which can inhibit specifically the interaction with caspase (e.g. PubChem CID: 5388929; 534.7 Da) or with Smac (e.g. PubChem CID: 139033416; 486.6 Da). MDM2 inhibits p53 mediated cell cycle arrest and apoptosis by binding its transcriptional activation domain, which also promotes the nuclear export of p53 [277]. A known inhibitor is 0R3 molecule (478.4 Da). Another benchmark case involved in apoptosis is the functional homo-trimeric form of TNFR1A (also known as TNFR1; gene TNFRSF1A) bound to the homo-trimeric cytokine tumor necrosis factor-beta TNF-β (also known as TNFB; gene LTA), which also binds to other proteins, such as TNFBR (also known as TNFR2; gene TNFRSF1B) or HVEM (gene TNFRSF14) [278]. A known inhibitor for this interaction is PubChem CID 703. Another case involving cell growth and death is ZipA, an essential cell division protein that binds FtsZ to stabilize it [279]. Several inhibitors are known for this interaction (e.g. WAC, IQZ, WAI, or CL3, with molecular weights ranging from 240.3 to 423.9 Da).

The remaining three benchmark cases are related to viral infection and immune response. HIV integrase binds human lens epithelium-derived growth factor (LEDGF), a transcriptional coactivator involved in neuroepithelial stem cell differentiation and neurogenesis, which facilitates the virus replication and survival [280]. Several inhibitors are known for this interaction (Table 8.3.1). Another case is the E2 from human papillomavirus type 11 (HPV-11), which plays a role in the initiation of viral DNA replication, and interacts with E1 to improve the specificity of E1 DNA binding activity [281]. There is only one inhibitor for this interaction with available complex structure (PDB ID: 1R6N; PubChem CID 5287508), a large molecule (608.5 Da) formed by several aromatic and carbonyl groups, 2 chloride groups, and a thiadiazole group. Another case is

Interleukine-2 (IL-2), which acts as a central regulator of the immune response, by binding to the hetero-trimeric IL-2 receptor (IL-2R) and stabilizing this functional oligomeric state [282]. A known inhibitor for this interaction is FRH (662.56 Da), also quite large and flexible molecule.

5.2.2. Surface cavity detection

We applied Fpocket [144] (http://fpocket.sourceforge.net/) to identify surface cavities on unbound protein structures and MD-based conformational models (see next section). We usually analyzed the top three pockets predicted by Fpocket according to their *Pscore* score.

To evaluate the performance of Fpocket for pocket detection, the predicted pocket residues (based on the list of pocket atoms predicted by Fpocket) were compared to those in the reference pocket of the complex-inhibitor complex structure (i.e. protein residues within 5 Å from the inhibitor), and different evaluation metrics were calculated: precision or positive predicted value (PPV), and sensitivity or coverage (COV) (Eq. 5.1). We usually defined as correct predictions (hits) those ones with PPV $\geq$ 40% and COV $\geq$ 40% with respect to the real position of the residues in the inhibitor pocket.

$$PPV_i = \frac{TP_i}{TP_i + FP_i} \qquad Coverage = \frac{TP_i}{TP_i + FN_i} \qquad (5.1, 5.2)$$

where for each predicted pocket *i*, true positives (*TP*) are the predicted pocket residues that are also found in the reference pocket, false positives (*FP*) are the predicted pocket residues that are not found in the reference pocket, and false negatives (FN) are the real residues in the reference pocket that are not found in the predicted pocket.

5.2.3. Molecular dynamics and transient cavities detection

For each analyzed protein structure, ten nanoseconds (ns) of molecular dynamics (MD) simulations were carried out using AMBER10. The unbound structures were prepared with pyDock *setup* module to use the same files in MD and later in pyDock docking, as described in the next section [80]. In some cases, protein structure from available PDB (Table 5.1) was modified for more realistic conditions (see details in Table 8.3.1). This *setup* step removed all hydrogen atoms. The topology and coordinates of the receptors were obtained

by using Force Field 99 and general AMBER force field (GAFF). The water molecules were added keeping their coordinates as in the x-ray structures. A fast minimization with cartesian restraints was performed to remove severe clashes, followed by minimization with explicit solvation. Each receptor was embedded in a solvated system within a periodic truncated octahedron box, and 150 mM NaCl was added into the system. Lastly, a fast solvent minimization was performed with a restraint mask of waters and ions. Then equilibration was performed at constant volume, using a 12 Å non-bonding cutoff. The equilibration process started by running 120 ps with protein constraints (50 kcal/mol·$Å^2$, from 0 to 300K) using Langevin dynamics (LD). The next 40 ps restraints were reduced from 50 to 25 kcal/mol·$Å^2$, and the next 40 ps from 25 to 10 kcal/mol·$Å^2$ at constant pressure. Finally, restraints were reduced to 5 kcal/mol·$Å^2$ including backbone atoms, followed by 20 ps with backbone restraints (1 kcal/mol·$Å^2$), and 60 ps with no restraints using LD, at 300K. After equilibration, 10 ns MD were performed at 300K at constant pressure, with a collision frequency of 0.2 ps^{-1} excluding bonds involving hydrogen atoms, and 2 ps of relaxation time. Then 1,000 conformational states uniformly selected out of the 10,000 snapshots from MD trajectories were analyzed using Fpocket, selecting the top three pockets per protein conformation according to *Pscore* scoring.

5.2.4. Docking simulations and hot-spot predictions

We applied pyDock docking and scoring method [80] to the unbound structures of the benchmark cases analyzed here. This docking approach consists in a two-step procedure in which docking orientations are generated with the FFT-based approach ZDOCK 2.1[181], and they are later scored by an energy-based scoring function, composed of desolvation, electrostatics and van der Waals energy terms. The results of docking were used to compute Normalized Interface Propensity (NIP) values [96] for every protein residue. Residues with NIP ≥ 0.2 are predicted hot-spots. The interacting proteins used in docking were defined according to the biological assembly of the available protein-protein complexes (see Appendix 3: Table 8.3.1). All cases are hetero-dimeric complexes, except the hetero-tetrameric HIV integrase/LEDGF complex, which had 2A:B2 stoichiometry, with two symmetric hetero-dimeric interfaces, and was thus treated as hetero-dimer. According to this, all unbound proteins were treated as monomers independently on their biological unit annotation. We used ICM browser (http://www.molsoft.com) to visualize structures,

compute atom distances, select atoms within a given distance, align and superimpose structures.

5.2.5. Protein-ligand docking

Protein-ligand was performed by Glide XP [283] and rDock [198], in order to evaluate the suitability of the identified inhibitor pockets for their use in docking. Schrödinger software platform (https://www.schrodinger.com/) was used for the preparation of proteins with *Protein Preparation Wizard,* and that of ligands with *LigPrep*. The OPLS2005 forcefield was used for the preparation of ligands and for Glide docking. The same prepared protein and ligand structures were used for the docking executions with rDock. Receptor grids were generated from the inhibitor pockets (either the predicted ones or taken from the protein-inhibitor complexes). In rDock, the center of mass of the residues in a given pocket was used to set the center of the docking box whose side length was defined as the distance between this center to the farthest atom of the pocket, plus 5 Å (radial distance). The side of the cubic OUTER box in Glide was defined as twice the side length of the docking box in rDock, while the INNER box in Glide was kept at the default value (10 $Å^3$).

5.3. Results

5.3.1. PPI inhibitor pockets are more difficult to identify than general ligand pockets

Fpocket was previously reported to detect 94% and 92% of the pockets within the three best-ranked predictions (i.e. top 3 performance) on the *holo* and *apo* proteins, respectively, in a set of 150 protein-ligand complexes [144]. In order to confirm this performance on protein-ligand complexes, and compare it with that on protein-PPI inhibitor complexes, we applied here Fpocket on the bound proteins (*holo*) of two different datasets: the first one composed of 102 protein-ligand complex structures from DUD-e database (http://dude.docking.org/) [284], and the second one with 264 non-redundant protein-PPI inhibitor complex structures from TIMBAL and 2P2I databases (involving 26 different proteins). For comparison purposes, we initially used the same evaluation metrics as in above mentioned study [144], where the Pocket Picker criterion (PPC) defined a correctly predicted pocket as that one in which its center of mass lies within 4 Å from at least one

atom of the ligand. Using this criterion, a correctly predicted pocket (*hit*) was found within the three best-ranked predictions in 89% of the DUD-e cases, and in 69% of the 2P2I/TIMBAL cases (Figure 5.1-A). The performance on DUD-e cases is similar to the results described in the above mentioned study. When analyzing only the best-scoring predicted pocket, the success rate is 70% for DUD-e and 43% for 2P2I/TIMBAL. The top 1 performance thus obtained here for DUD-e is worse than the one obtained in the above mentioned study for the larger set of 150 cases (83%) [144].

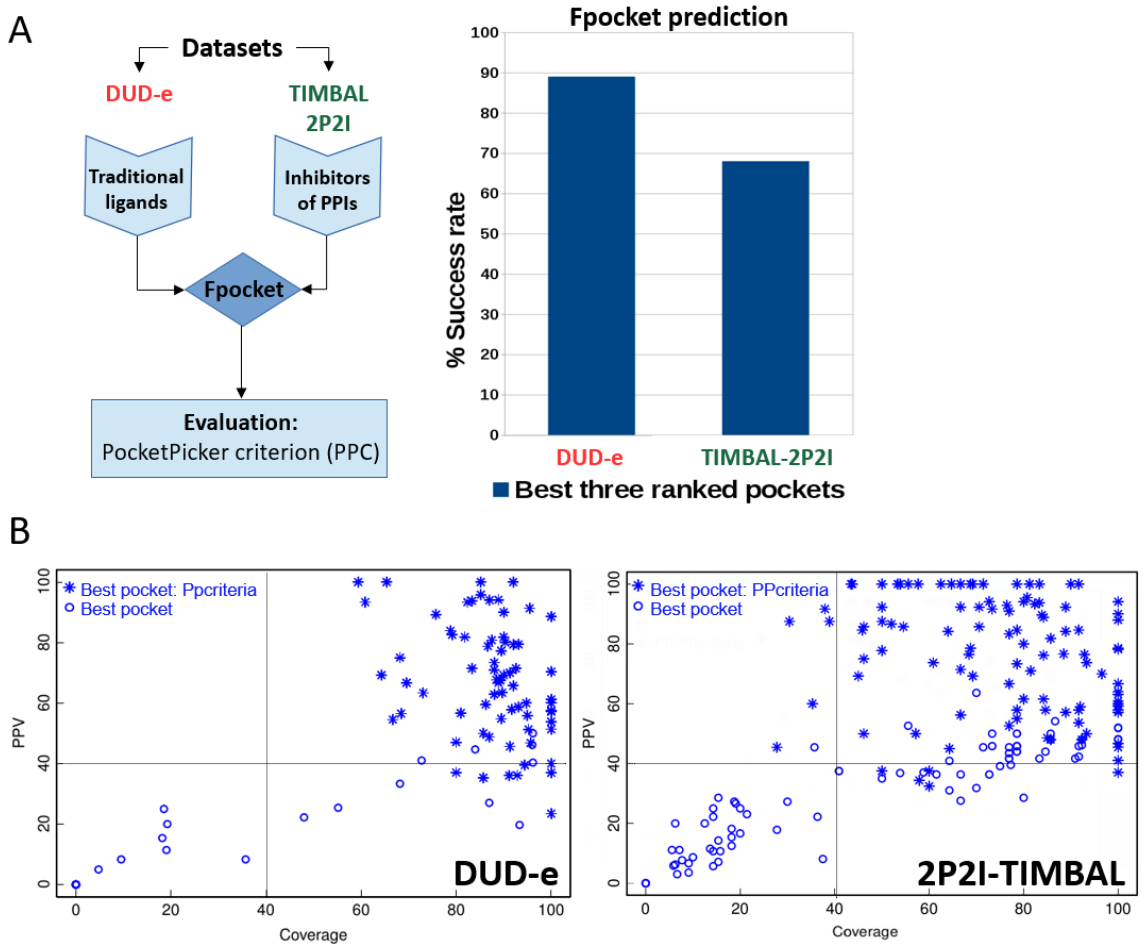


Figure 5.1 Prediction of protein-ligand and protein-PPI inhibitor pockets by Fpocket
(A) Left: scheme of the test performed here; right: top 3 performance of Fpocket on predicting known protein-ligand (DUD-e database) and protein-PPI inhibitor (TIMBAL/2P2I databases) pockets. (B) Positive predicted value (PPV) and coverage (COV) of best-scoring predicted pockets on DUD-e (left) and TIMBAL/2P2I (right) databases. The plot shows the best-scoring predicted pocket for each case, with those considered a hit (by PPC) represented as * symbol.

Since we were concerned that the PPC metrics to define correct pockets might not be too restrictive, we used other evaluation metrics like the PPV and coverage of the predicted pocket residues with respect to the reference ones. As can be seen in Figure 5.1-B, the large majority of pockets considered good by PPC have PPV and coverage over 40% (both in protein-ligands and in protein-PPI inhibitors). Protein-ligand pockets detected by PPC have better coverage with respect to the reference pockets. Based on the above findings, when using this criterion to define a correctly predicted pocket (PPV and coverage $\geq$ 40%), the top 1 success for DUD-e and 2P2I/TIMBAL is 68% and 53%, respectively, while the top 3 success is 82% and 74%, respectively. This follows the same trends as the PPC criterion, and confirms that the PPI inhibitor sites are more difficult to identify than the ligand ones, even when using the protein-ligand and -inhibitor complex structures for test purposes. Previous studies discussed about the importance of pocket size and accessibility [140, 141]. In this context, PPI inhibitor binding sites tend to be large and flat, making it difficult to identify them with methods initially developed to detect deep cavities. Indeed, the predicted ligand cavities have on average higher *Pscore* scores than the predicted inhibitor cavities.

To explore the possibility that perhaps in the unsuccessful cases Fpocket might have found the correct PPI inhibitor pockets but with poorer score (i.e. ranking > 3), we computed the predictive performance when considering all the predicted pockets for each case, independently on their ranking. In this case, Fpocket can identify a correct pocket (i.e. PPV and coverage $\geq$ 40% with respect to the reference pocket) in 86% of the DUD-e cases, and in 77% of the 2P2I/TIMBAL cases, which is similar to the above shown top 3 performance, thus indicating that no additional correct pockets are found beyond the top 3 predicted pockets in the unsuccessful cases. To further evaluate potential problems in the scoring of PPI inhibitor cavities, we have restricted the predicted pockets to only those at the protein-protein interface (predictions were made on the protein coordinates from the protein-ligand or -inhibitor complexes, and no information from the protein-protein complexes was used until this last test). When selecting only the predicted pockets at the protein-protein interface, the top 1 predictive performance for DUD-e and 2P2I/TIMBAL cases improve up to 79% and 67%, respectively, and the top 3 predictive performance up to 86% and 77%, respectively, obtaining values more similar to those when considering all predicted pockets independently of the ranking. This shows that knowledge of the protein-

protein interface might be important to complement Fpocket scoring and improve the predictions.

In addition to the above discussed difficulties in detecting PPI inhibitor cavities, there are also problems in the assessment of predictions. Perhaps the definition for a successful predicted pocket used in traditional ligand sites is not optimal for PPI inhibitor cases. Figure 8.3.1 shows an example of a predicted pocket that englobes the real PPI inhibitor site. Therefore, it has a good coverage, but its PPV is slightly below the cutoff (40%) used to define a correct prediction (PDB ID: 3VNG). In the same figure, there is another case in which PPC indicates unsuccessful prediction, but it is assessed as correct according to PPV and coverage criteria (PDB ID: 3U5L).

5.3.2. Identifying PPI inhibitor pockets on unbound proteins

After the above described general test on protein-ligand and -inhibitor complex structures, we aimed to perform a more realistic benchmark test by applying Fpocket to identify surface cavities on the unbound proteins (*apo*) of the PPIs of our benchmark set (Table 5.1) and compare them with the reference pocket in the protein-PPI inhibitor complex. In most of the benchmark cases, there were several available structures with the protein bound to different inhibitors (see Appendix 3: Table 8.3.1). For the initial evaluation of the predictions, we selected the structure of the complex with the inhibitor with the best IC50. As an example, Figure 5.1-A shows the best-*Pscore* pocket predicted by Fpocket in the unbound IL-2 structure, which is not located near the PPI inhibitor with the best IC50, not even at the known protein-protein interface. The predicted pockets for all the cases are shown in Figure 8.3.2 of Appendix 3. Only in one case (11% of the benchmark set; see Table 5.2) we can find a correctly predicted pocket (i.e. PPV and coverage $\geq$ 40%) (see Appendix 3: Figure 8.3.3-A). For comparison purposes, if we apply Fpocket on the proteins bound to these PPI inhibitors (*holo*), a correctly predicted pocket is found in 33% of the cases (see Appendix 3: Figure 8.3.3-B; Table 5.2), slightly better than when using the unbound proteins, but still a poor performance that is comparable to that obtained on TIMBAL/2P2I databases described in the previous section (53%). This confirms that PPI inhibitor pockets are difficult to identify on the *holo* proteins, and even more difficult to predict on the unbound proteins. Nevertheless, to disregard the possibility that the inhibitor with the best IC50 might not be the most appropriate to be predicted, we compared the predicted pockets on the unbound proteins with respect to all the other known inhibitors for

each system, and the predictive results do not improve (see Appendix 3: Figure 8.3.3-C; Table 5.2). Finally, we have also explored the pocket prediction on the proteins taken from the protein-protein complexes, to check whether they adopt some conformations that could facilitate pocket prediction, but while some cases can be now correctly predicted, the overall predictive rate (33%) is the same as when using the proteins bound to the inhibitors (see Appendix 3: Figure 8.3.3-D; Table 5.2). We should note that here, despite using the protein-protein complex structure for the predictions, we have not restricted the predicted pockets to be located only at the interface region, since our goal was to evaluate the impact of the different conformational states on the predictions.

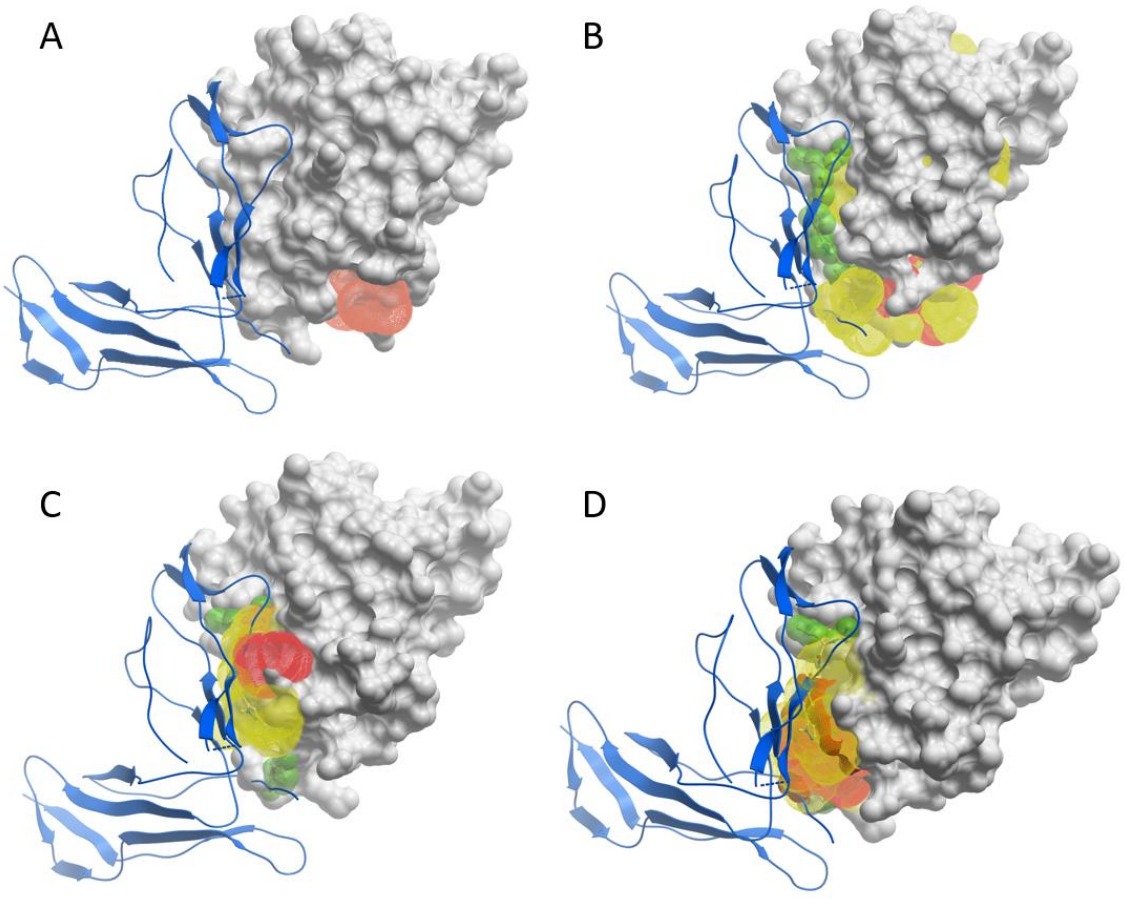


Figure 5.2 Predicted pockets in IL-2 protein using different conformational states.
IL-2 protein is shown in grey surface, with IL-2R parnter protein in blue ribbon, and the inhibitor with the best IC50 in green. (A) Best-scoring predicted pocket (in orange) on the unbound IL-2. (B) 10 best-scoring pockets predicted on MD conformers generated from unbound IL-2 (best-scoring pocket in orange, the others in yellow). (C) 10 best-scoring pockets from MD, restricted to the known protein-protein interface (best-scoring pocket in orange, the others in yellow). (D) 10 best-scoring pockets from MD, containing ≥ 3 predicted hot-spots (best-scoring pocket in orange, the others in yellow).

Table 5.2 Success rates for the best-ranked predicted PPI inhibitor pocket by FPocket on different selected protein structures and filtering conditions.

Protocol	Reference on inhibitor with best IC50	Reference on all inhibitors
Unbound protein	11%	11%
Protein bound to protein	33%	33%
Protein bound to inhibitor	33%	33%
MD conformers	11%	11%
MD conformers + interface	56%	67%
MD conformers + hot-spots	44%	56%

Overall, these results confirm that x-ray protein structures (either in their unbound states or from protein-protein complexes) do not have surface cavities that can be easy to identify and exploited as binding sites for protein-protein inhibitors. All this suggests that we need to consider conformational variability in the protein structures in order to identify suitable pockets for PPI inhibitors.

5.3.3. MD simulations can generate transient PPI inhibitor pockets

Molecular dynamics (MD) simulations were previously reported to identify transient pockets in unbound protein surfaces known to be involved in protein-protein interactions, which were not present in the crystal structure of the unbound proteins [143]. We wanted to apply this strategy to the proteins of our benchmark set in order to generate transient pockets that could be suitable for inhibitor binding. With this purpose, we generated 10 ns MD trajectories from the unbound protein structures, from which 1,000 conformers were generated (see Methods). Then we used Fpocket to identify surface cavities in all these conformers. The best-*Pscore* pocket was selected for each conformer as usual, but this yielded near 1,000 predicted pockets per protein (for a few conformers, no pocket was predicted). The *Pscore* ranking indicates the capacity of the pocket to bind a small-molecule [144], but it does not include the druggability of the pocket. Thus, we filtered the total number of best-*Pscore* pockets from all conformers to keep only the 100 pockets with the best druggability score as defined by Fpocket. Finally from these, we selected the pocket with the best *Pscore*. As an example, Figure 5.1-B shows the 10 best-scoring pockets predicted from the MD-based conformers generated from the unbound IL-2. The

best-scoring pocket (in orange) is not located at the PPI inhibitor site, but some of the top 10 pockets are nearby. The performance of pocket prediction on MD-based conformers for all the benchmark cases is shown in Figure 8.3.3-E of Appendix 3, in comparison with all known PPI inhibitors. Correctly predicted pockets are found only in 11% of the cases (Table 5.2), yielding the same performance as with the unbound protein structures. Despite being able to generate suitable pockets with MD in some of the cases (see for instance Figure 5.1-B), they are not identified with the default Fpocket scoring tools.

The problem is that the transient cavities are found across the entire protein surface, while we are mostly interested in those cavities at the protein-protein interfaces (apart from possible allosteric effects, one would expect that a small-molecule modulator would affect a PPI only if it is bound at the interface). Therefore, we analyzed whether knowing the location of the protein-protein interface could help to improve the predictions. This information is known in the cases of our benchmark, so we filtered the MD-based cavities to select only those in which ≥ 40% of the pocket residues are located at the protein-protein interface (defined as those residues within 5 Å inter-atomic distance from the partner protein). The impact of including the information of the protein-protein interface can be visualized in our IL-2 example, in which now the best-scoring pocket is correctly predicted (Figure 5.1-C). The benefit of including the protein-protein interface information is clear for many of the benchmark cases: under these assumptions, the best-scoring pocket would be correctly located at a known inhibitor binding site in 67% of the cases (see Appendix 3: Figure 8.3.3-F; Table 5.2). This indicates that, although molecular dynamics can generate transient binding pockets that could be suitable for binding PPI inhibitors, such pockets are still difficult to identify unless we know the protein-protein interface location. The main difficulty here is that the 3D structure of the complex is available only for a small fraction of all possible protein-protein interactions (see Introduction), which limits its applicability in real cases. In this context, we will analyze in the next sections whether the computational prediction of protein interfaces and structural modeling of protein-protein interactions can help to identify the correct pockets among the many ones predicted from MD conformers in cases in which the structure of the protein-protein complex is not available.

5.3.4. Computational docking can identify interface hot-spot residues

As we discussed above, in realistic situations in which the location of the protein-protein interface is not known, it would be helpful to rely on computational tools for the prediction of the interface residues. Among the several reported methods, we have explored here the use of pyDock, which can identify interface and hot-spot residues on unbound proteins based on docking calculations (see Methods). First, we tested the capabilities of this method for the prediction of interface hot-spots in our benchmark set. For this, docking orientations were generated from the unbound proteins with pyDock, and the 100 lowest-energy orientations were used to calculate *NIP* values for all residues in each interacting protein, as previously described [96]. Figure 5.3 shows the docking-based *NIP* values for the unbound proteins in the benchmark set. The predicted hot-spot residues are defined as those with $NIP \geq 0.2$. The number of predicted hot-spots in the unbound proteins of our benchmark set ranges from 0 (ZipA) to 21 (HPV E2) (Table 5.3). As can be seen in Figure 5.3, in many of the cases the predicted hot-spot residues are located at the interface with the partner protein. We have quantified in Table 5.3 the number of predicted hot-spot residues that are located at the known protein-protein interfaces (i.e. within 10 Å from any atom of the partner protein). In general, 55% of the predicted hot-spots are located at the protein-protein interface. Moreover, in six out of the nine benchmark cases, more than half of the predicted hot-spot residues are located at the protein-protein interface.

Table 5.3 Docking-based prediction of hot-spot residues compared to real interfaces.

PPI target	Predicted hot-spots (HS)	Predicted HS located at protein-protein interfaces[1]
Bcl-xL	9	7
HPV E2	21	16
IL-2	4	4
Integrase	16	0
MDM2	7	7
XIAP BIR3/caspase	18	10
XIAP BIR3/smac	19	11
TNFR1A	14	4
ZipA	0	0

[1] Predicted hot-spot residues located within 10 Å from any atom of the partner protein in the protein-protein complex structure.

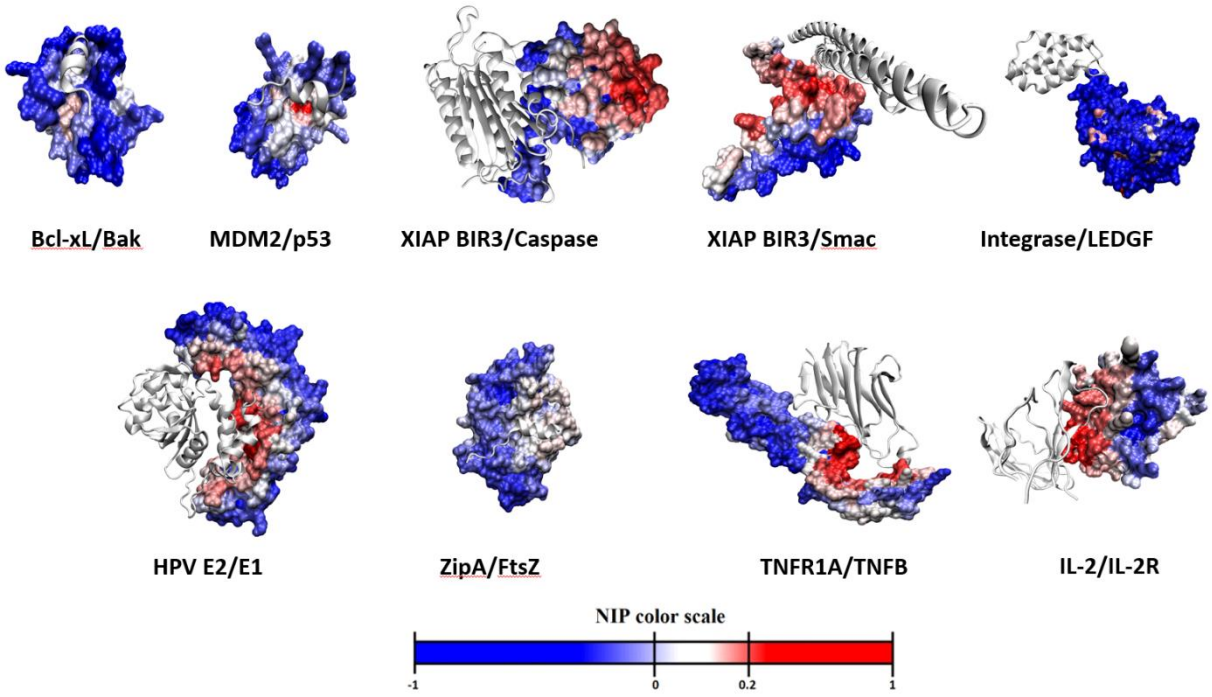


Figure 5.3 Docking-based hot-spot predictions.
Protein residues are colored by NIP value, resulting from the docking calculations on the unbound proteins. For comparison, the partner protein is shown in white ribbon.

For some of the predicted hot-spot residues, there is experimental evidence in the literature. For instance, Bcl-xL F146 residue is predicted here as a hot-spot, which is confirmed in previous experimental studies [285-287]. Several hot-spots have been experimentally identified for E2 (Y19, Q24, E39, Y99 and E100) [288], and we successfully predict two of them (Y19 and E100). In IL-2, we successfully predict three hot-spots (R38, F42 and L72) among the experimentally identified ones (K35, R38, F42, K43, E62 and L72) [133]. In some cases, we successfully predict the hot-spots in the partner protein, but they are not used here for the purpose of identifying PPI inhibitor binding sites.

Regarding the type of predicted residues, we find here, in agreement with previous observations, that *NIP* values are more likely to predict as hot-spots the aromatic phenylalanine, tyrosine and tryptophan residues, the polar lysine, arginine, glutamic acid and threonine residues, and the non-polar leucine residue. *NIP* values are less likely to predict as hot-spots the non-polar methionine, glycine, valine or alanine residues, and the polar glutamine residue.

5.3.5. Predicted hot-spot residues are critical to identify PPI inhibitor pockets

We found in previous sections that the use of information on the protein-protein interface is essential to reduce the number of candidate pockets from MD conformers and thus identify the correct pockets with higher precision. Despite this information is not available in the majority of protein-protein complexes, we have studied here whether using the predicted hot-spots from docking simulations could help to improve the identification of the predicted pockets. In our example protein IL-2, when the pockets predicted on MD conformers from the unbound state are filtered to select only those that contain at least three predicted hot-spot residues, the 10 best-scoring pockets are located at the protein-protein interface, and the best-scoring one is correctly predicted as close to the PPI inhibitor binding site (see Appendix 3: Figure 8.3.3-D). These results are much better than those obtained directly from the MD conformers (see Appendix 3: Figure 8.3.3-B), and similar to the ones obtained when using the location of the known protein-protein interface (see Appendix 3: Figure 8.3.3-C). This strategy of selecting only MD-based predicted pockets containing ≥ 3 predicted hot-spots improves the general predictive performance for the benchmark set, obtaining correctly predicted pockets in 56% of the cases (Figure 5.4; Table 5.2). This is comparable with the performance obtained when using the information of the known protein-protein interface (67%), which suggests that docking calculations with *pyDock* can be a helpful tool to locate PPI inhibitor binding sites in the absence of structural information on the protein-protein interfaces.

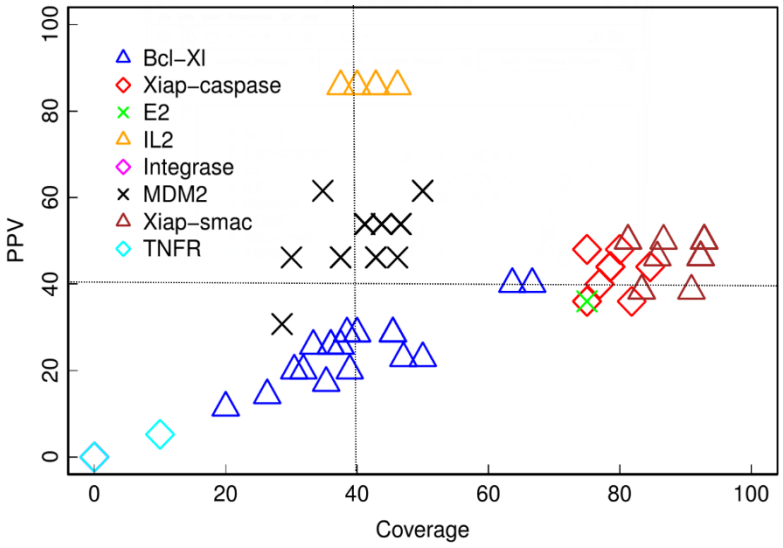


Figure 5.4 Assessment of the identification of PPI inhibitor pockets by integrating MD simulations and docking-based hot-spot predictions. PPV and coverage of the best-scoring predicted pocket on the MD-based conformers, which contain at least three predicted hot-spots.

5.3.6. Assessing the use of the predicted pockets in protein-ligand docking

As above discussed, we have devised a protocol integrating MD simulations and docking-based hot-spot predictions to identify transient cavities on protein-protein interfaces suitable for binding of small molecules. We already showed that this approach is able to predict the location of known PPI inhibitor binding sites in 56% of the cases of the benchmark set. Now, we should evaluate whether these predicted pockets are suitable for binding of PPI inhibitors, and can be used in protein-ligand docking simulations.

As an initial test, we aimed to reproduce the known protein-PPI inhibitor binding modes by protein-ligand docking, starting from the structure of the protein bound (*holo*) to the inhibitor with the best IC50. For the sake of simplicity, the PPI inhibitor conformation was also kept as in the protein-PPI inhibitor complex structure. We used Glide [283] and rDock [198] protein-ligand docking programs, with different parameters. More specifically, we tested several grid size values in order to extend the original grids in up to 10 Å, increased the number of docking poses up to 1000, and tried the OPLS 2005 force field used in Glide in addition to the newest OPLS3. Under these ideal conditions, Glide with OPLS 2005 forcefield found correct docking orientations (i.e. ligand RMSD < 2 Å) within the 5 best-scoring docking models in 7 of the 9 cases, while rDock found correct models in only 3 cases (sampling during docking was extended to up to 1000 poses). According to these results, we decided to assess the use of Glide with the predicted pockets, in realistic conditions.

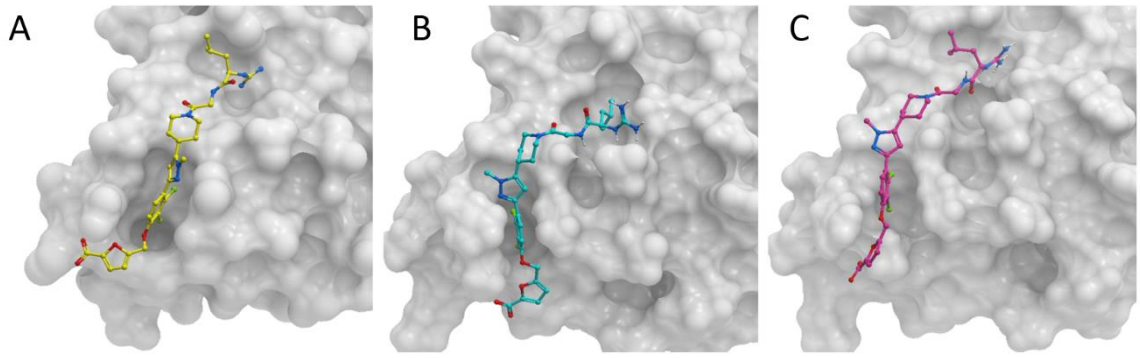


Figure 5.5 Docking of inhibitor on the predicted cavities of IL-2.
IL-2 is shown in grey surface, and the different binding modes of the inhibitor are shown as ball and sticks. (A) IL-2 bound to FRH inhibitor (PDB ID: 1PY2). (B) GLIDE best-scoring docking model, using the 1st ranked pocket from MD and hot-spot prediction (from a MD snapshot at 2.241 ns). (C) The closest docking model to the reference in terms of ligand RMSD, obtained with the 9th ranked pocket from MD and hot-spot prediction (from a MD snapshot at 2.261ns).

With this purpose, we applied Glide to the unbound proteins, using the pockets predicted by MD and hot-spot predictions, a more realistic situation. In this case, the performance was much worse. We can get reasonable models (ligand RMSD ≤ 4.0 Å) within the 20 best-scoring docking poses in two of the benchmark cases, but for this we had to use the 10 best-scoring pockets. In the case of IL-2, the rank 1 docking model obtained when using the best-scoring pocket from MD and hot-spot prediction (Figure 5.5-B) has a conformation that is slightly different (ligand RMSD 5.2 Å) to that of the PPI inhibitor in the complex structure (Figure 5.5-A). When using the pocket ranked 9th from MD and hot-spot prediction, the rank 1 docking model is closer to the reference (ligand RMSD 3.5 Å) (Figure 5.5-C). But in general, it seems that the transient pockets generated by MD and selected by the predicted hot-spots, even if they are well located at the PPI inhibitor sites, their conformation is not optimal for ligand docking.

5.4. Discussion

5.4.1. Other possible criteria to select the pockets

In our protocol, we used the hot-spot predictions by *pyDock* to select the pockets predicted by Fpocket during the MD simulation. This helped to identify the correct pockets. In Figure 8.3.4 of Appendix 3, we show the PPV values for the best-scoring pockets in the 1,000 MD conformers as a function of Fpocket score (*Pscore*). Since *Pscore* alone is not able to discriminate the correct pockets in some cases, the addition of filtering criterion based on the predicted hot-spots help to improve the predicted pockets.

We wanted to further explore other possible criteria to select the MD-based pockets. For instance, perhaps the size of the pocket in terms of number of residues could help to identify the cavities more suitable for inhibitor binding. In Figure 8.3.5 of Appendix 3, we show the PPV of the MD-based predicted pockets for every case, as a function of the number of residues of each pocket. While in some cases (e.g. XIAP interacting with caspase, or the same protein interacting with smac) the largest pockets are good in terms of coverage, they have PPV < 40%. So in general, the size of the pocket does not help to identify a successful pocket. This is consistent with the fact that PPI inhibitor pockets are in average smaller (15.0 ± 6.2 residues; average and SD computed on pockets in TIMBAL/2P2I databases) than traditional ligand binding sites (23.7 ± 6.4 residues; average and SD computed from pockets in DUD-e set).

In our protocol, we finally sorted the best-scoring predicted pockets from MD by druggability score, selecting only the top 100 pockets according to that score. We also explored the inclusion of druggability score as a filter, with different cutoffs. In Figure 8.3.6 of Appendix 3 we show the effect of using a druggability cutoff of > 0.7%, which is almost not affecting the predictions.

5.4.2. Lessons from unsuccessful predictions

Unsuccessful cases might be related to conformation and/or oligomerization issues. For instance, the TNFR1A/TNFB complex, besides the difficulty of its large protein-protein interface, the main challenge is that the functional homo-trimeric form of TNFB is interacting with the homo-trimeric form of TNFR1A in addition to other proteins [278]. Since we docked TNFR1A and TNFB as monomers, and did not consider the rest of interacting proteins, the docking calculations might not be fully accurate (indeed, most of the predicted hot-spots for this case are not located at the studied protein-protein interface, perhaps because they are part of the interface with the other functional interactions). Moreover, the known PPI inhibitors bound to TNFR1A are only partially located at the interface with TNB. We could speculate that the inhibitor may also affect the interaction with the other partner proteins, which indirectly would disrupt TNFR1A/TNFB interaction.

Another unsuccessful case is Bcl-xL. This is another example in which different oligomerization states and conformational arrangements can critically affect the predictions. The structure of this protein in complex with Bak has been determined by NMR (PDB ID: 1BXL), and the unbound structure by x-ray crystallography (PDB ID: 1R2D). In the latter, there is a long missing loop after the N-terminal α-helix, which we modeled based on the NMR structure. In a newer x-ray structure for the Bcl-xL/Bak complex (PDB ID: 5FMK), this N-terminal α-helix has been assigned to a different chain (Figure 5.6-A). The biological assembly assigned by the Protein Data Bank (PDB) is a hetero-tetramer, formed by two copies of Bcl-xL/Bak complex. Since the location of Bak does not seem to affect Bcl-xL homo-dimer interface, this suggests that unbound Bcl-xL could be a homo-dimer (this structure contains a few more Bcl-xL C-terminal residues that might explain the observed different oligomerization). Interestingly, the structure of the homologue Bcl-2 (PDB ID: 5VAX) has the same missing loop and the same N-terminal *α*-helix as unbound Bcl-xL (PDB ID: 1R2D), and the biological assembly for Bcl-2/peptide interaction is hetero-dimeric. Despite not participating directly in the interaction, this flexible N-terminal

α-helix seems to have an essential role in the oligomerization of Bcl-xL. In addition, Bcl-xL shows a large conformational rearrangement when the hinge/turn region between α-helices 5 and 6 is extended, which induces a different dimer orientation (PDB ID: 4PPI) (Figure 5.6-B). Similar dimerization is found by domain swapping, which seems to be important to avoid interaction with p53 [286, 289]. The pocket predicted by our method is far from the known inhibitor of Bcl-xL/Bak interaction (Figure 5.6-C). This could be due to problems derived from the above discussed conformational flexibility of this protein. Indeed, in the unbound protein structure, there could be potential clashes between the inhibitor and the C-term α-helix, which is not present in the protein-PPI inhibitor complex structure (PDB ID: 2YXJ). Another reason is that our approach might be detecting another cavity of biological relevance. Actually, in the structure of Bcl-xL in complex with p53 (PDB ID: 2MEJ), p53 residues 176-193 are found in the predicted pocket region (Figure 5.6-D).

In the case of XIAP, although our approach correctly located the PPI inhibitor binding site, this predicted pocket was not sufficiently open to efficiently bind the small-molecule inhibitor. Perhaps in some cases, MD sampling would need to be run for longer. In other cases such as ZipA, TNFR1A or integrase, a major limitation was that the docking calculations provided incorrect hot-spots. In the case of ZipA/FtsZ interaction, the complex structure contained only residues 367-383 of FtsZ, but the entire FtsZ was modeled and used in docking, which perhaps caused problems in the predictions. Finally, the hetero-tetrameric HIV integrase/LEDGF complex was actually treated as hetero-dimer, since its stoichiometry was 2A:2B, with two equal hetero-dimeric interfaces. However, the inhibitor pocket was formed by the two homo-dimer integrase molecules, and the use of only one integrase monomer might have affected the pocket and hot-spot predictions. This shows once again the importance of considering the appropriate oligomerization state of the interacting proteins.

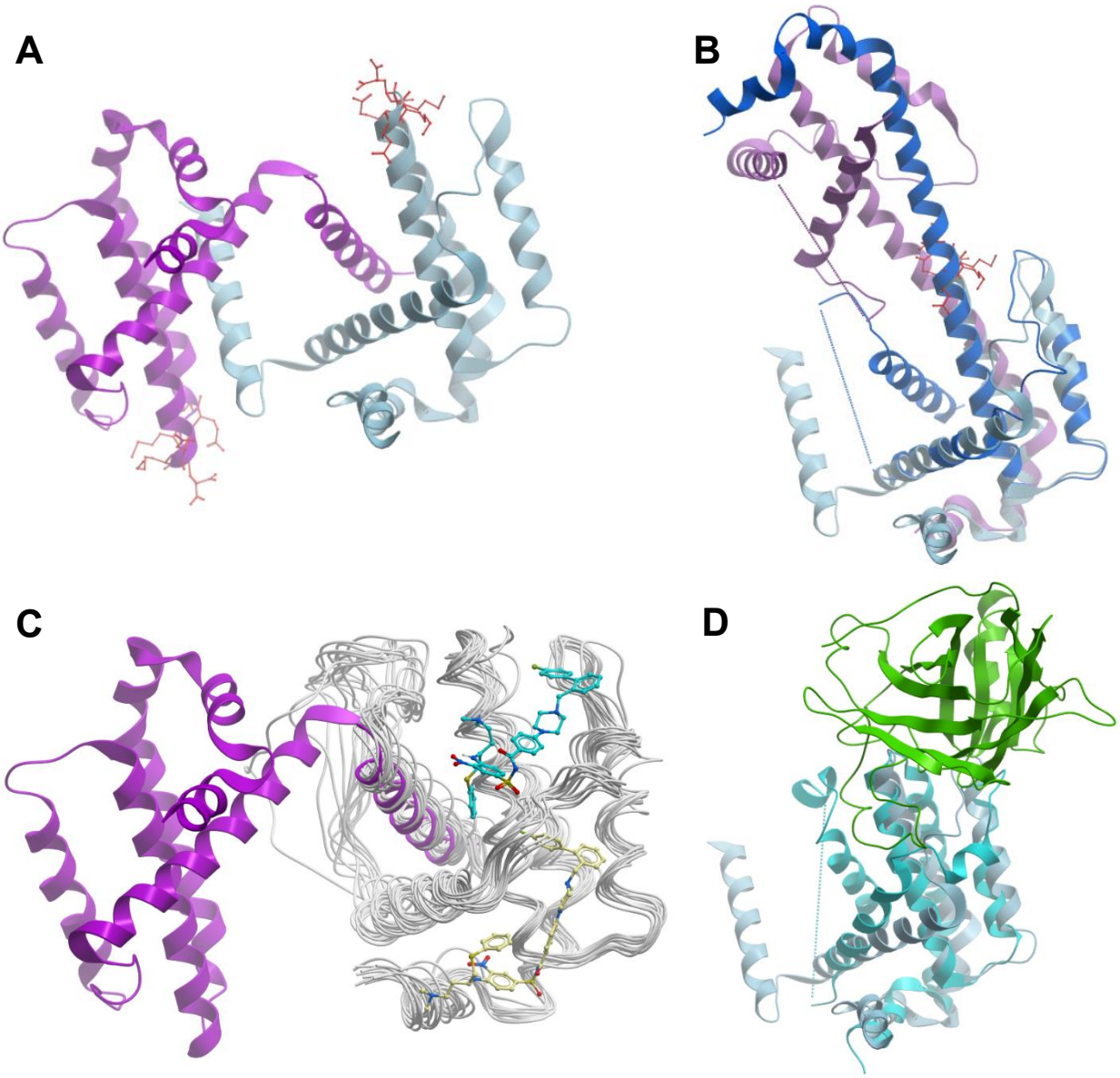


Figure 5.6 Structural analysis of Bcl-xL conformational and oligomeriation states.
(A). Bcl-xL is a homo-dimer in the x-ray structure in complex with Bak (PDB 5FMK). In red, hinge region (residues 156-161) in the hairpin turn between 5th and 6th α-helices. (B). Extended hinge/turn region between α5 and α6 in Bcl-xL induces different dimer orientation (PDB 4PPI). In light blue is represented the conformation shown in (A). (C) Different Bcl-xL conformers from MD sampling are represented in white. We can observe the swapping of the N-terminal α-helix with respect to the conformation shown in (A). The best docking orientation found by our approach for this PPI inhibitor is represented in blue sticks. The real location of the PPI Inhibitor is represented in yellow sticks (PDB 2YXJ). (D) Bcl-xL (cyan) in complex with p53 (green) (PDB 2MEJ). In light blue is represented the Bcl-xL conformation shown in (A).

5.5. Conclusions

In this work we have shown how the integration of structure-based cavity detection, molecular dynamics, and docking-based prediction of hot-spots, can help to identify PPI inhibitor binding sites on the surface of unbound proteins. Two critical aspects here are: i) the use of MD-generated conformers, which produces thousands of transient cavities across the protein surface, and ii) the use of protein-protein docking methods to predict hot-spots, which can help to locate the interface pockets among all the generated by MD. This has the advantage that can be applied in the absence of structural information on the protein-protein complex, a realistic situation in the majority of cases.

One of the limitations is the small size of the benchmark set. There are few cases for which the structure is available for the unbound proteins as well as for the protein-protein complex and the protein-PPI inhibitor complex, which are important for benchmarking new computational approaches. The other problem is related to the existence of large conformational rearrangements, which are not described by standard MD, and limited knowledge of the possible oligomerization states in some of the proteins. Both are essential for accurate predictions.

Despite the difficulties, we propose here a protocol that can improve the detection of surface cavities for the identification of small-molecule modulators of protein-protein interactions. More work needs to be done in the conformational description of proteins and their interactions, as well as in the optimal use of these predicted pockets in ligand docking protocols, but this study can be helpful towards the goal of targeting PPIs of therapeutic interest.

6. General discussion

In this thesis, we have tried to improve the understanding of protein-protein interactions (PPIs) and the impact of single amino acid variants (SAVs) on them. We also explored the dynamics of protein-protein interfaces to help drug discovery targeting new therapeutic targets. To that end, varied *in silico* techniques have been applied and integrated to face the different obstacles in the field.

In human, around 1-5% of PPIs have available 3D structure [25, 44, 45]. Despite the advances in the experimental determination of protein structures, such as Cryo-EM and NMR, solving the structure of proteins and PPIs, there is a need to cover this enormous gap through *in silico* techniques. Thus, many efforts have been focused on modeling proteins and protein-protein interactions. As we have mentioned in the Introduction, different computational techniques have been developed and continuously upgraded. Regarding the modeling of PPIs, computational docking methods are being continuously evaluated at the CAPRI blind assessment [85] and recently, also at the CASP rounds focused on multi-molecular assemblies [86]. In this regard, last CAPRI rounds and the two joint CASP-CAPRI rounds imposed new challenges for our docking and scoring approaches, which were integrated into a broader modeling scheme, including docking, template-based modeling, flexible refinement, and experimental restraints. For most of protein structure prediction targets the oligomerization state is highly relevant. In a broader sense, our performance on these validation experiments was successful concerning predictors and scorers (see Figure 3.3). For predictors, our approach did not stand out of the best performances in the 7th CAPRI edition, but it was the second best approach in CASP (first in CAPRI). For scorers, our approach had the best performance in both experiments.

For the predictors experiment, the most successful docking approach in the recent 7th CAPRI edition was that of the Andreani and Guerois group, which went beyond their traditional rigid-body and InterEvScore approach, and they applied different strategies for the inclusion of evolutionary constraints, such as template-based modeling with RosettaCM-based protocol [79, 214]. In addition, the most successful approach as predictors in CASP13-CAPRI was that of Venclovas group. They basically used template-based models when reliable templates were found, and free docking with HEX [71] otherwise, with the use of VoroMQA [202] for the evaluation and selection of the final models. In addition to these successful methods, our approach was one of the most efficient approaches based on a combination of template-based and *ab initio* docking followed by pyDock scoring [170], besides the integration of available experimental restraints and

symmetry data [4](Figure 3.1). PyDock was used to evaluate all the proposed models, and in case of reliable templates, consistency between energy-based scoring and template-based data was sought. Comparing our performance with other methodologies was more evident on the "difficult" cases for which no clear template was available. Taking into account unsuccessful predictions in 7th CAPRI edition, we learned the critical importance to choosing the optimal docking approach for each case, depending on the template availability.

For the scorers experiment, pyDock got the best performance when considering top 10 predictions, which shows its capabilities to evaluate complex models derived from combined approaches (template-based, *ab initio*, refinement) [176](Figure 3.3). Somehow unexpectedly, Venclovas group were less efficient in the scorers experiment (rank 7th), which might indicate that this function seems mostly optimized for their own pipeline for template-based and docking generation, while its application to models generated by other sources represents a challenge to be solved. Another successful approach was the use of CONSRANK [203, 204] for the ranking of docking models based on the most frequent inter-residue contacts in the ensemble of decoys with a clustering procedure [205].

All approaches converge to an integrative computational approach that aim to efficiently use experimental structural data and additional information from a variety of sources for the structural modeling of complexes [215]. For example, the targets T150 and T151 were the same complex as T149, a challenging multi-domain dimer, for which SAXS and chemical cross-linking data were provided, respectively. Interestingly, the inclusion of restraints from SAXS data improved the models submitted by pyDock to the original target (which had few successful groups), and the cross-linking data further improved pyDock submissions [171].

Thus, the success of pyDock predicting protein-protein complexes in CAPRI and CASP supports the computational prediction of hot-spot residues. The main advantages of this method are that it does not require the structure of a complex and that hot-spot predictions are specific for two given interacting proteins (as opposed to general interface prediction on one unbound protein as above described). Through pyDock we can characterize protein-protein interfaces and compute the residues that contribute the most to the binding free energy of the protein-protein complex (above mentioned hot-spots).

Another important aspect analyzed in the thesis is related to a better understanding of the impact of genetic variants at the protein level, through the structural characterization of SAVs and their interactions. As mentioned above, a major limitation is the low availability of 3D structures for protein-protein complexes. Consequently, a large fraction of SAVs cannot be precisely located at protein interfaces. For this reason, using docking models to estimate whether SAVs can be involved in PPIs may be auspicious. We have collected structural information for the proteins and protein-protein complexes (models from Interactome3D), the oligomeric state, the structural coverage, and the identification of interacting proteins (Table 8.2.2). That compiled data can be valuable information in order to run docking simulations in the most realistic conditions, as well as for accurate calculations of binding affinity changes upon mutations.

Mutations in the same gene can affect different phenotypic traits (pleiotropy). In this context, the number of interactions affected by specific mutation key to understand pleiotropic effects in disease genes. Recent studies show that SAVs located at distinct protein-protein interfaces of the same protein are prone to produce different disease phenotypes [17, 39, 249]. Moreover, it has been demonstrated that one-third of the SAVs produce an 'edgetic' effect by impairing only a subset of the interactions [17]. We assessed this premise with the structural analysis of the case example hemoglobin subunit beta (HBB), which showed that the same SAV could affect the interaction with different partner proteins if their interface patches are the same, and different SAVs could perturb different partner proteins if these have distinct interface patches (Figure 4.6). Analyzing the structural distribution of SAVs from the newborn screening program, these are more likely to be located at the interface rather than at the non-interacting surface. More precisely, we found that they are more probable to occur at the interface core region rather than at the rim, in agreement with previous studies [40, 41, 234, 243]. On the contrary, neutral SAVs occur significantly more often in the interface rim region as well as in the non-interacting surface, as compared to the interface core region. Furthermore, the predicted hot-spot residues are more likely to be located at the interface core. This is in line with previous studies [41, 245, 250], which revealed that hot-spot residues are not equally distributed among interface regions, but they tend to be clustered within the interface core. Thus, this core region is critical for the stabilization of PPIs; this is reflected in the fact that core residues show a higher level of conservation and coevolution among homologous proteins as compared to those in the rim [246, 251]. This energetical relevance of the core region

also explains why SAVs are not as likely to be tolerated there as in the interface rim or in the non-interacting surface [41]. We found that arginine, tryptophan and tyrosine are over-represented among disease-causing mutated residues. This is consistent with previous studies reporting that the most frequent hot-spot residues are tryptophan (21%), arginine (13.3%) and tyrosine (12.3%) [110, 252, 253].

The structural characterization of SAVs on protein-protein complexes it is not often sufficient for understanding the edgetic effect on PPIs. Therefore, we have analyzed the energetic effect of interface mutations by applying diverse *in silico* methods to model the mutations and compute the change in binding affinity, using techniques such as energy-based, machine learning, and empirical forcefields. These predictive methods have been validated on experimental data from the SKEMPIv2.0 database [88]. 39.67% of the SAVs in SKEMPI were destabilizing, while only 5.57% were stabilizing (Figure 4.7). In total, at least 45% of the analyzed mutations have an energetic impact on protein-protein interactions. We have used experimental cutoffs that are used in other studies (as stabilizing if $\Delta\Delta G_{bind} \leq$ -1.0 kcal/mol, or as destabilizing if $\Delta\Delta G_{bind} \geq$ 1.0 kcal/mol) [96]. The Pearson correlation coefficients are in general low, ranging from r = 0.31 (pyDockEneRes) to r = 0.34 (FoldX). In a recent study, using a set of mutations to alanine in binary complexes from SKEMPIv2.0 not included in SKEMPIv1.0, the prediction of changes in binding affinity by the different methods showed slightly better correlations: r = 0.44 (FoldX); r = 0.34 (mCSM); r = 0.48 (pyDockEneRes, based only on electrostatics + van der Waals) [128]. In general, the distributions for the different mutation types are very similar in all methods, indicating poor separation between classes. Nonetheless, some of the destabilizing mutations have higher scoring values for all predictors, and a few of the destabilizing mutations have smaller values for some of the predictors, especially for pyDock, pyDockEneRes and mCSM.

Based on these observations, we next explored whether we could use these predictors as a classifier to identify the destabilizing or stabilizing mutations. For destabilizing mutations, the cutoff value that seems to provide the best balance between PPV and sensitivity overall explored is 2.0 a.u. (arbitrary units) for all predictors (Figure 4.10). Positive Predicted Values (PPV) are quite similar, ranging from 58% (pyDockEneRes; sensitivity 56%) to 70% (mCSM; sensitivity 29%) being pyDock the most balanced method with PPV of 62% and sensitivity of 54%. Compared to other studies, PPV/sensitivity values are slightly lower ranging from 38% PPV (BeAtMuSiC; sensitivity

27%) to 41% PPV (UEP; sensitivity 70%), besides the predictive methods already used in the thesis that also have lower PPV/sensitivity values [256]. The difference between the rates obtained from the compared one is the form to assign the positive/negative predictions. We tested different experimental cutoffs for assign stabilizing/destabilizing mutations, while the compared study is based if mutations increase/decrease the experimental binding affinity. Regarding the predicted energy changes for the stabilizing variants defined from SKEMPIv2.0 data, the PPV and sensitivity values are, in general, much worse for all predictors (Figure 4.10). Based on that, the best predicted energy cutoff for FoldX and pyDock/pyDockEneRes seems to be -2.0 a.u. again, with this cutoff, pyDock and pyDockEneRes show the best balance between PPV and sensitivity (PPV 14.5%, sensitivity 35.8% for pyDock; PPV 14.3%, sensitivity 43.4% for pyDockEneRes), better than FoldX (PPV 14.3%, sensitivity 3.4%). The best cutoff for mCSM seems to be 0 a.u., showing PPV 18.6%, and sensitivity 20.7% (Figure 4.10, see Appendix 2: Figure 8.2.1 and Figure 8.2.2).

After this analysis, we classified the mutations according to the selected cutoff values for each predictor. Despite significant overlapping, a large part of the most destabilizing mutations (as experimentally defined) are correctly predicted as destabilizing (Figure 4.11). Moreover, many of the most stabilizing mutations are correctly predicted as stabilizing. This is reflected in the large sensitivity shown for most of the methods in Figure 4.10. Overall, despite the large overlapping between the predicted classes, there is an important number of predictions correctly classified. The selected cutoffs are not able to clearly separate mutations in an accurate way but helped to improve the classification.

We have extended the structural characterization of pathogenic and neutral SAVs from the previous study and further analyzed another variant set on protein-protein complexes from BM4 (based on specific protein-protein interactions). In the latter, we unexpectedly found that 80% of pathogenic SAVs at the protein-protein interfaces were located at the rim region. This value is clearly above the expected value by random, considering that in the interfaces of this BM4 there are 51% core residues and 49% rim residues. This unexpected percentage is probably due to structural limitations since we are analyzing only a specific number of hetero-oligomeric protein-protein complexes that were compiled in this BM4 set for testing docking programs, therefore not including other biologically-relevant interactions in which the mutated proteins might be involved. On the contrary, in proteins related to diseases detected in newborn screening that are involved in

homo-oligomers and hetero-oligomers, pathogenic SAVs are mostly found at the interface core region while neutral SAVs are more abundant at the interface rim region, as expected. This is consistent with previous studies that find that SAVs are less energetically tolerant at the interface core region than at the interface rim or at the non-interacting surface [41].

We have applied the validated computational protocols for the energetical characterization of SAVs. Although the distribution for the predicted binding energy values show a strong overlap between pathological mutations and neutral variants, we observed that the disease-related variants had more extreme values for mCSM and FoldX. This overlap could be due to different reasons. One reason might be that the specific pathogenic SAVs do not have effects on the specific protein-protein interfaces (perhaps they affect to other interfaces for which there is no structural coverage). Another reason might be related to the existing difficulties in predicting binding affinity changes upon mutations, since state-of-the-art methods achieve PCC around 0.34-0.5 (Chapter 4 part 2 section 3.2)[128]. Nevertheless, the results show a small proportion of disease-unrelated (neutral) variants in the newborn screening mutational set that are predicted to be stabilizing or destabilizing in comparison to the pathogenic variants (Figure 4.14). This is not so clear in the mutational data from BM4, in which pathogenic mutations do not seem to have larger energetic effect on the protein-protein interactions from BM4, perhaps because they might have an effect on other protein-protein interactions different from the ones analyzed here, as above discussed. For mutations with predicted energetic effect on PPI, pyDock/pyDockResEne shows 80% of predicted stabilizing mutations as pathological. On the other hand, mutations predicted as destabilizing by all methods are mostly pathological. On contrary, low percentages observed in the BM4 set could be caused by that most of these pathogenic mutations might have an effect in other protein-protein interactions that are not included in our analysis (see section 4.3.3.2).

This computational analysis of the energetic effects of SAVs involving protein-protein interacting complexes, besides helping the detection of edgetic variants, can provide important information for a better comprehension of molecular mechanisms of diseases involving homo-oligomeric and hetero-oligomeric interfaces. Despite structural limitations for understanding pleiotropic effects, around 30% of mutations have a significant effect on protein-protein interaction (this percentage is much lower for BM4 ~10%). For those proteins involved in more than one protein-protein interaction, we detect potential edgetic effects of SAVs. With our results, we can describe the mechanistic effects at the molecular

and energetic levels, providing relevant information helping to interpret the SAVs effects, which can contribute to improve diagnosis decisions in cases of therapeutic interest (see section 4.3).

The present study has some limitations derived from the low availability of 3D structures of protein-protein complexes, in addition to the difficulties in the description of the conformational states of PPIs. This is especially relevant in BM4 subset, where we limited the specific set of interactions to the ones in the reported benchmark, without exploring other possible protein partners, and without checking alternative oligomerization states. This may be the reason why in this subset, there were more SAVs at the interface rim than at the core region. Further characterization of PPIs could be performed by protein-protein docking and modeling procedures, but this step needs a better assessment of the methodologies, oligomerization states, available interface information, etc. for the achievement of accurate protein-protein complex models. Despite these limitations, we have performed a structural and energetic characterization of SAVs on PPIs, thus identifying those mutations that are more prone to have an energetic impact on PPI (potentially involved in disease). But further work is needed to get a robust classifier using decision algorithms through experimental energetic mutational data. From a very preliminary classifier of the energetic impact of SAVS, we provide novel information on disease-related variants. This is extremely useful in order to characterize SAVs in future studies, interpret them at the molecular level, improve the accuracy of pathogenicity predictors on new mutations, and help to advance toward precision medicine by helping clinicians to provide adequate diagnosis and treatments, for example targeting PPI.

From the validation of *ab initio* protein-protein docking techniques. pyDock provides accurate predictions of hot-spot residues of specific PPIs. Specifically, NIP values ≥ 0.2 achieves PPV of 68% (sensitivity 43%). We used the hot-spot predictions by *pyDock* to select suitable cavities for targeting PPIs. Despite the limitations of the small size of the benchmark. We have shown how the integration of structure-based cavity detection, molecular dynamics, and docking-based prediction of hot-spots, can help to identify PPI inhibitor binding sites on the surface of unbound proteins. This has the advantage that can be applied in the absence of structural information on the protein-protein complex, a realistic situation in the majority of cases.

PPI inhibitor pockets are in average smaller (15.0 ± 6.2 residues; average and SD computed on pockets in TIMBAL/2P2I databases) than traditional ligand binding sites

(23.7 ± 6.4 residues; average and SD computed from pockets in DUD-e set). Fpocket helps to detect and characterize pockets, but in some cases, the scoring function of Fpocket (*Pscore)* alone is not able to discriminate the correct pockets for targeting PPIs. The application of this protocol contributes to better identification of pockets for targeting PPIs. Through molecular dynamics we are able to explore transient openings from x-ray structures. Using *ab initio* docking we are able to predict protein-protein interfaces in blind conditions, and identify specific cavities surrounded by residues that contribute the most to the binding affinity (hot-spots).

Lessons from unsuccessful predictions are mostly related to oligomerization states and/or conformational arrangements. In some cases, we do not select the correct biological unit getting wrong docking and hot-spots predictions and wrong surface exploration (i.e. Integrase and TNFR1A). XIAP need more time during the exploration of transient cavities. Bcl-xL, this is another example in which different oligomerization states and conformational arrangements can critically affect the predictions.

7. General conclusions

- The scoring function from pyDock was particularly successful for the multimeric targets of the last CASP13 edition and, for the varied protein complexes of the lasts CAPRI rounds. As predictors, we should note the good performance of *ab initio* docking in predicting difficult targets in which template-based was not helpful, as well as the improvement in the quality of the submitted models from the integration of template-based and energy-based docking.

- We deal with the present limitations facing the structural characterization of SAVs involving PPIs. Structural availability of proteins and protein-protein complexes, as well as an accurate consideration of oligomerization states of the interacting molecules and the complexes are essential for a correct energetic description of the mutations and their impact on the interactions, which in turn is key for a successful comprehension of molecular mechanisms of diseases.

- Pathogenic SAVs are more likely to be located at the interface rather than at the non-interacting surface. More precisely, we found that they are more probable to occur at the interface core region rather than at the rim. On the contrary, neutral SAVs occur significantly more often in the interface rim as well as in the non-interacting surface, as compared with the interface core region.

- The hot-spot residues are more likely to be located at the interface core, then, SAVs located at core region have energetic relevance in the interaction. We found that arginine, tryptophan and tyrosine are over-represented among disease-causing mutated residues.

- The analysis of experimental binding affinity changes of interface mutations shows a larger proportion of destabilizing mutations in comparison to the stabilizing ones. Both stabilizing and destabilizing mutations are more enriched in the interface core region compared to the interface rim region, while neutral SAVs are indistinctly distributed on the interface.

- After evaluating the prediction of binding energy changes in mutations in comparison with their experimental values, all of the studied predictive methods provide a reasonable prediction of destabilizing effects of the mutations, but more limited prediction of stabilizing mutations.

- Energetical characterization provides new insights on disease-related variants, for instance to detect potential edgetic effects that could explain the pleiotropic effects by impairing only a subset of the interactions. With our analysis, we detect potential egdetic effects in some variants that could not be explained without exploring the energetic effect on binding.

- Despite the limitations of current predicting methods, pathological mutations have higher FoldX and mCSM values compared to neutral variants, and pyDock/pyDockEneRes have the best performance on predicting pathological stabilizing mutations. However, variants from BM4 show no differences between unclassifed and disease-related, which could be related to the limited set of protein-protein interactions analyzed for these variants, or that some unclassified variants are involved in disease but not annotated.

- This work has contributed to extend the characterization of pathological SAVs affecting protein-protein interactions, beyond structural-only analysis, which can help to identify protein-protein interactions that might be affected by a given pathological situation and therefore, constitute interesting targets for drug discovery purposes.

- From the reduced size of the PPI inhibitor benchmark, we have shown how the integration of structure-based cavity detection, molecular dynamics, and docking-based prediction of hot-spots, can help to identify PPI inhibitor binding sites.

- More work needs to be done in the conformational description of proteins and their interactions, as well as in the optimal use of these predicted pockets in ligand docking protocols, but this study can be helpful towards the goal of targeting PPIs of therapeutic interest.

8. Appendices

8.1. Appendix 1. Supplementary material for Chapter 3

Table 8.1.1 Structural availability of the interacting molecules and additional information for the preparation of the submitted models as servers and predictors

Target[a]	RECEPTOR		LIGAND		COMPLEX		
	Unbound Structure (int RMSD)[b]	Template (SI, RMSD)[c]	Unbound Structure (int RMSD)[b]	Template (SI, RMSD)[c]	Template (SI, RMSD)[c]	Docking restraints[d]	Target structure[e]
Protein-protein							
T122	5MXA (3.4 Å)	-	-	1I1R (25%, 10.4 Å)	-	Literature	5MZV
T123	5LMW (N/A)	5BOZ (76%, N/A)	-	multi-template I-TASSER (N/A, N/A)	-	-	-
T124	5FWO (2.2 Å)	4GRW:H (78%, 3.0 Å)	-	multi-template I-TASSER (N/A, N/A)	Dimer: 3MTR (17%, 21.9 Å)	-	6EY6
T125 *hetero*	4QKH (1.4 Å)	-	-	3T3A (47%, 10.6 Å)	-	Literature	5MGT
T125 *homo*	4QKH (1.4 Å)	-	4QKH (1.3 Å)	-	4QKH (100%, 0.7 Å)	-	5MGT
T131	4WHD (1.4 Å)	-	-	5LP2 (96%, 0.9 Å)	-	Literature	6GBG
T132	4WHD (1.0 Å)	-	-	5LP2 (57%, 2.9 Å)	-	Literature	6GBH
T133	-	3U43:B (87%, 0.8 Å)	-	3U43:A (83%, 1.5 Å)	-	-	6ERE
T136	-	5FKZ:E (45%,N/A), multi-template I-TASSER (N/A)	-	5FKZ:E (45%, N/A), multi-template I-TASSER (N/A, N/A)	2VYC (40%, N/A), 5FKZ (45%, N/A)	5FKZ interface	6Q6I
Protein-peptide							
T121	1LR0 (N/A)	-	-	multi-template I-TASSER (N/A,N/A), 2HQS (42%, N/A), 2W8B (38%, N/A)	-	1TOL interface	-
T134	1F3C (2.0 Å)	-	-	1F95 (33%, 1.2 Å), 1F96 (9%, 2.4 Å), 3P8M (36%, 2.3 Å)	-	1F95 interface	6GZL
T135	1F3C (2.0 Å)	-	-	1F95 (33%, 1.2 Å)	-	1F95 interface	6GZL
Protein-saccharide							
T126	-	5F7V (30%, N/A)	2C7F,2X8S, 3QEF, CID 53356682 (N/A)	-	-	-	6RKH
T127	-	5F7V (30%, N/A)	CID 74539968 (N/A)	-	-	-	6RKX
T128	-	5F7V (30%, N/A)	5HOF, CID 74539969 (N/A)	-	-	-	6RL2
T129	-	5F7V (30%, N/A)	1GYE, CID 74539970 (N/A)	-	-	-	6RL1
T130	3D5Z (0.7 Å)	-	5HOF (2.6 Å), CID 74539969 (2.3 Å)	-	-	-	6F1G

Table 8.1.2 Overall 7th CAPRI performance ranking for the protein-protein targets. Extracted from [176].

			top 1		top 5		top 10	
Rank	Group name	#targets	rank	performance	rank	performance	rank	performance
1	Kozakov/Vajda	11	2	5/1***/3**	1	6/1***/5**	1	6/1***/5**
2	Venclovas	8	1	5/2***/3**	2	5/2***/3**	2	5/2***/3**
3	Seok	11	6	4/3**	3	5/1***/4**	3	5/1***/4**
3	Pierce	9	3	4/2***/1**	3	5/2***/2**	3	5/2***/2**
5	Andreani/Guerois	11	6	4/3**	5	5/1***/3**	5	5/1***/3**
6	Zou	11	6	3/1***/2**	6	4/1***/3**	6	4/1***/3**
6	Zacharias	11	14	4/1***	6	5/1***/2**	6	5/1***/2**
6	Kihara	11	6	4/1***/1**	6	5/1***/2**	6	5/1***/2**
6	Gray	11	3	5/1***/2**	6	5/1***/2**	6	5/1***/2**
10	Shen	11	14	3/1***/1**	10	4/1***/2**	6	5/1***/2**
10	Moal	8	24	2**	10	4/1***/2**	11	4/1***/2**
10	MDOCKPP	11	5	4/1***/2**	10	4/1***/2**	11	4/1***/2**
10	HADDOCK	11	6	3/1***/2**	10	4/1***/2**	11	4/1***/2**
14	Grudinin	11	6	3/1***/2**	14	3/1***/2**	16	3/1***/2**
14	Fernandez-Recio	11	6	3/1***/2**	14	3/1***/2**	16	3/1***/2**
14	CLUSPRO	11	20	3/2**	14	4/3**	16	4/3**
14	Bonvin	11	6	3/1***/2**	14	3/1***/2**	16	3/1***/2**
18	Weng	11	14	3/1***/1**	18	3/1***/1**	20	3/1***/1**
18	SWARMDOCK	11	14	3/1***/1**	18	3/1***/1**	20	3/1***/1**
18	LZERD	11	14	3**	18	3**	20	3**
18	Chang	11	20	2/1***/1**	18	3/1***/1**	11	4/1***/2**
18	Bates	11	14	3/1***/1**	18	3/1***/1**	11	4/1***/2**
23	Vakser	11	24	2**	23	2/1***/1**	24	2/1***/1**
23	Takeda-Shitaka	8	20	2/1***/1**	23	2/1***/1**	24	2/1***/1**
23	Huang	9	20	2/1***/1**	23	2/1***/1**	20	3/1***/1**
26	Wolfson	7	24	2/1***	26	2/1***	27	2/1***
26	PYDOCKWEB	11	28	1***	26	2/1***	24	3/1***
26	HDOCK	8	30	1**	26	2**	27	2**
26	GALAXYPPDOCK	8	24	2**	26	2**	27	2**
30	Laine	3	28	1***	30	1***	30	1***
31	Ritchie	5	30	1**	31	1**	32	1**
31	Iwadate	1	30	1**	31	1**	32	1**
31	INTERPRED	1	30	1**	31	1**	32	1**
31	Del Carpio	11	30	1**	31	1**	32	1**
31	Carbone	6	36	0	31	1**	30	2/1**
31	Brini	1	30	1**	31	1**	32	1**
37	ZDOCK	1	36	0	37	0	37	0
37	Wang	0	36	0	37	0	37	0
37	Wallner	4	36	0	37	0	37	0
37	UUcourse	0	36	0	37	0	37	0
37	Tuffery	0	36	0	37	0	37	0
37	Schueler-Furman	0	36	0	37	0	37	0
37	Schneidman	1	36	0	37	0	37	0
37	Sanner	0	36	0	37	0	37	0
37	Niv	0	36	0	37	0	37	0
37	Negi	4	36	0	37	0	37	0
37	GRAMM-X	3	36	0	37	0	37	0
37	Gong	0	36	0	37	0	37	0
37	Czaplewski	1	36	0	37	0	37	0
37	Carazo	1	36	0	37	0	37	0

Table 8.1.3 Overall 7th CAPRI performance ranking for the protein-peptide targets. Extracted from [176].

			top 1		top 5		top 10	
Rank	Group name	#targets	rank	performance	rank	performance	rank	performance
1	Zacharias	3	1	2/1***/1**	1	2***	2	2***
1	Schueler-Furman	3	1	2/1***/1**	1	2***	1	3/2***/1**
1	Andreani/Guerois	3	1	2/1***/1**	1	2***	2	2***
4	Venclovas	3	11	1	4	2/1***/1**	2	2***
4	Seok	3	7	2/1**	4	3/2**	5	3/2**
4	Moal	3	1	2/1***/1**	4	2/1***/1**	5	2/1***/1**
4	Huang	3	1	2/1***/1**	4	2/1***/1**	5	2/1***/1**
8	HDOCK	2	6	2/1***	8	2/1***	8	2/1***
9	Zou	2	11	1	9	2	12	2
9	Shen	3	8	1**	9	1**	12	1**
9	Kozakov/Vajda	2	11	1	9	1**	8	2**
9	GALAXYPPDOCK	3	21	0	9	1**	12	1**
9	Fernandez-Recio	3	11	1	9	2	12	2
9	CLUSPRO	2	11	1	9	1**	12	1**
9	Brini	2	8	1**	9	1**	12	1**
9	Bonvin	3	8	1**	9	1**	8	2**
17	UUcourse	1	11	1	17	1	12	1**
17	SWARMDOCK	3	21	0	17	1	12	2
17	PYDOCKWEB	2	21	0	17	1	23	1
17	Kihara	3	11	1	17	1	12	1**
17	HADDOCK	3	21	0	17	1	12	2
17	Del Carpio	2	11	1	17	1	23	1
17	Czaplewski	2	11	1	17	1	12	2
17	Bates	3	11	1	17	1	11	3
25	Wang	1	21	0	25	0	25	0
25	Wallner	1	21	0	25	0	25	0
25	Vakser	2	21	0	25	0	25	0
25	Tuffery	1	21	0	25	0	25	0
25	Takeda-Shitaka	1	21	0	25	0	25	0
25	Sanner	1	21	0	25	0	25	0
25	Ritchie	1	21	0	25	0	25	0
25	Niv	1	21	0	25	0	25	0
25	Negi	1	21	0	25	0	25	0
25	MDOCKPP	2	21	0	25	0	25	0
25	LZERD	3	21	0	25	0	25	0
25	Grudinin	3	21	0	25	0	25	0
25	Gong	1	21	0	25	0	25	0
25	Chang	3	21	0	25	0	25	0

Table 8.1.4 Overall 7th CAPRI performance ranking for the protein-oligosaccharide targets. Extracted from [176].

			top 1		top 5		top 10	
Rank	Group name	#targets	rank	performance	rank	performance	rank	performance
1	Andreani/Guerois	5	4	4/1**	1	5/1***/2**	1	5/1***/3**
2	Seok	5	1	5/1***/1**	2	5/1***/1**	2	5/1***/2**
2	LZERD	5	4	5	2	5/1***/1**	2	5/1***/2**
2	Chang	5	2	4/1***	2	5/1***/1**	4	5/1***/1**
5	Kozakov/Vajda	5	9	3/1**	5	5/2**	8	5/2**
5	Huang	5	9	4	5	5/2**	8	5/2**
5	HDOCK	5	24	1	5	4/3**	4	5/3**
5	CLUSPRO	5	24	1	5	5/2**	8	5/2**
9	Zou	5	9	4	9	5/1**	14	5/1**
9	Zacharias	5	4	5	9	5/1**	14	5/1**
9	Venclovas	5	4	3/1***	9	4/1***	14	4/1***
9	Takeda-Shitaka	5	9	4	9	5/1**	14	5/1**
9	MDOCKPP	5	14	3	9	5/1**	14	5/1**
9	Kihara	5	9	4	9	4/2**	4	5/3**
9	Grudinin	5	2	4/1***	9	4/1***	8	5/1***
9	Gray	5	24	1	9	4/2**	14	4/2**
9	Fernandez-Recio	5	19	2	9	5/1**	8	5/2**
9	Bonvin	5	14	2/1**	9	3/1***/1**	8	4/1***/1**
19	Shen	5	19	2	19	5	14	5/1**
19	HADDOCK	5	14	2/1**	19	3/1***	4	5/1***/1**
19	Carbone	4	4	3/1***	19	3/1***	22	3/1***
22	Vakser	4	19	2	22	4	24	4
22	Moal	5	30	0	22	4	22	5
22	GALAXYPPDOCK	5	14	3	22	3/1**	14	4/1***
25	Pierce	2	14	2/1**	25	2/1**	25	2/1**
25	Bates	5	19	2	25	3	25	3
27	SWARMDOCK	5	24	1	27	2	28	2
27	Negi	4	19	2	27	2	28	2
29	Ritchie	1	24	1	29	1	28	1**
29	Del Carpio	5	24	1	29	1	25	3

Table 8.1.5 Overall CAPRI-CASP13 performance ranking for the "Easy" targets [170]

	Predictor group							
Rank	*Human*	#[a]	rank	Top-1	rank	Top-5	rank	Top-10
1	Venclovas	9	6	7/3***/3**	1	9/5***/4**	1	9/5***/4**
2	Fernandez-Recio	9	1	9/3***/6**	2	9/4***/5**	2	9/4***/5**
3	Kihara	9	2	8/3***/4**	3	9/4***/4**	3	9/4***/4**
	Seok	9	9	8/1***/5**	3	9/3***/6**	3	9/3***/6**
5	Vakser	9	2	9/1***/7**	5	9/2***/7**	3	9/3***/6**
	Bates	9	4	8/2***/5**	5	8/4***/4**	6	8/4***/4**
7	Zou	9	4	9/8**	7	9/2***/6**	6	9/3***/5**
	Chang	9	9	8/7**	7	8/3***/5**	8	8/3***/5**
9	Kozakov	9	6	8**	9	9**	9	9**
	Huang	8	9	7/3***/2**	9	8/4***/2**	9	8/4***/2**
	Weng	9	12	6/3***/2**	9	7/4***/3**	9	7/4***/3**
12	Eisenstein	8	6	7/3***/3**	12	7/4***/2**	12	7/4***/2**
13	Shen	9	9	7/2***/4**	13	7/2***/4**	13	7/2***/4**
	Pierce	9	14	5/1***/3**	13	6/3***/3**	13	6/3***/3**
15	Elofsson	9	16	5/3**	14	7/2***/3**	14	7/2***/3**
16	Czaplewski	9	17	4/3**	15	6/1***/4**	15	6/2***/3**
17	Moal	8	15	5/1***/2**	17	5/2***/2**	16	5/2***/3**
	Grudinin	9	17	3/1***/2**	17	5/1***/4**	18	5/1***/4**
19	Carbone	9	20	2/1***/1**	19	4/2***/1**	19	4/2***/1**
20	Schneidman	7	17	3/2***	20	3/2***	19	4/2***/1**
21	Hou	3	21	2**	21	2/1***/1**	21	2***
22	Ritchie	2	22	1**	22	1**	22	1**
23	Liwo	6	23	0	23	0	23	1
	Crivelli	6	23	0	23	0	23	1
	EMBO2017 course	0	23	0	23	0	23	0
	DelCarpio	7	23	0	23	0	23	0
Rank	*Server*	#[a]	rank	Top-1	rank	Top-5	rank	Top-10
1	HDOCK	9	2	6/4***/2**	1	9/5***/4**	1	9/5***/4**
2	SWARMDOCK	9	1	8/2***/5**	2	8/4***/4**	2	8/4***/4**
3	HADDOCK	9	2	8/3***/2**	3	9/3***/3**	3	9/3***/3**
	CLUSPRO	9	2	8**	3	9**	3	9**
5	LZERD	9	5	6/3***/3**	5	7/3***/4**	5	7/3***/4**
6	MDOCKPP	9	6	8/1***/4**	6	8/1***/5**	6	8/2***/4**
7	GALAXYPPDOCK	6	7	4**	7	5/2***/2**	7	5/2***/2**
8	HAWKDOCK	2	8	1**	8	2/1**	8	2/1***
Rank	*Scorers*[b]	#[a]	rank	Top-1	rank	Top-5	rank	Top-10
1	Fernandez-Recio	9	12	5/1***/4**	1	9/4***/5**	1	9/4***/5**
2	Oliva	9	1	9/3***/6**	2	9/3***/6**	1	9/4***/5**
3	Huang	9	4	7/2***/4**	3	9/3***/5**	3	9/3***/5**
4	Zou	9	4	8/7**	4	9/1***/8**	3	9/2***/7**
	Chang	9	4	8/7**	4	9/1***/8**	3	9/2***/7**
6	Venclovas	9	2	8/1***/6**	6	8/2***/6**	3	9/2***/7**
	HDOCK	9	2	8**	6	9**	3	9/2***/7**
	MDOCKPP	9	7	7/1***/5**	6	9/1***/7**	3	9/2***/7**
9	Kihara	9	9	7/6**	9	8/2***/5**	3	8/4***/4**
	LZERD	9	7	8/6**	9	8/2***/5**	10	8/4***/3**
11	Bates	9	13	6/4**	11	8/1***/6**	13	8/1***/7**
12	Bonvin	9	11	7/5**	12	8/7**	11	8/2***/6**
	Carbone	9	9	6/1***/5**	12	7/2***/4**	13	8/2***/5**
14	Weng	9	14	5/4**	14	7/1***/5**	11	9/1***/7**
15	HAWKDOCK	6	14	4/1***/3**	15	5/2***/3**	18	5/2***/3**
16	Grudinin	8	16	4**	16	5/1***/4**	15	6/2***/4**
17	QASDOM	8	17	3**	17	5**	15	7**
18	Seok	8	18	2**	18	5/4**	17	7/1***/4**

[a] Target participation, out of 9 (for both predictors and scorers)
[b] Human and Server together

Table 8.1.6 Overall CAPRI-CASP13 performance ranking for the " Difficult" targets [170].

	Predictor group							
Rank	*Human*	#[a]	rank	Top-1	rank	Top-5	rank	Top-10
1	Venclovas	11	3	2/1***/1**	1	4/1***/2**	1	4/1***/2**
2	Shen	11	1	4/1***/1**	2	4/1***/1**	2	4/1***/1**
3	Weng	11	2	3/1***/1**	3	3/1***/1**	2	4/1***/1**
	Fernandez-Recio	10	7	2/1**	3	3/1***/1**	2	4/1***/1**
	Kozakov	10	7	2/1**	3	4/2**	5	4/2**
6	Huang	11	5	2**	6	3/2**	5	4/2**
	Seok	11	3	2/1***/1**	6	2/1***/1**	7	2/1***/1**
8	Kihara	11	5	2**	8	2**	9	2**
	Zou	11	13	2	8	4	9	4
10	Eisenstein	4	7	2/1**	10	2/1**	7	3/2**
	Pierce	11	7	1***	10	1***	9	2/1***
	Chang	11	7	2/1**	10	2/1**	12	2/1**
	Bates	9	7	1***	10	1***	12	1***
	Vakser	11	13	2	10	2/1**	12	2/1**
15	Grudinin	11	18	0	15	2	12	2/1**
16	Carbone	11	18	0	16	1	16	1**
	Moal	9	15	1	16	1	17	1
	Elofsson	11	15	1	16	1	17	1
	Czaplewski	8	15	1	16	1	17	1
20	Hou	8	18	0	20	0	17	1
	Schneidman	5	18	0	20	0	21	0
	Ritchie	2	18	0	20	0	21	0
	Liwo	5	18	0	20	0	21	0
	EMBO2017 course	1	18	0	20	0	21	0
	Del Carpio	6	18	0	20	0	21	0
	Crivelli	6	18	0	20	0	21	0
Rank	*Server*	#[a]	rank	Top-1	rank	Top-5	rank	Top-10
1	GALAXYPPDOCK	11	1	2/1***	1	2/1***	1	3/1***
	LZERD	11	1	2**	1	2**	3	2**
	CLUSPRO	11	4	2	1	3/1**	3	3/1**
4	SWARMDOCK	11	3	1***	4	1***	5	1***
	MDOCKPP	11	4	2	4	3	5	3
6	HDOCK	11	4	1**	6	1**	1	3/2**
7	HAWKDOCK	5	7	0	7	0	7	0
	HADDOCK	10	7	0	7	0	7	0
Rank	*Scorers*[b]	#[a]	rank	Top-1	rank	Top-5	rank	Top-10
1	MDOCKPP	10	11	1**	1	4/1***/1**	1	4/1***/1**
2	Seok	9	1	2/1***/1**	2	3/1***/1**	5	3/1***/1**
	Oliva	10	1	2/1***/1**	2	3/1***/1**	5	3/1***/1**
	Kihara	10	1	2/1***/1**	2	3/1***/1**	1	4/1***/1**
	Fernandez-Recio	10	1	2/1***/1**	2	3/1***/1**	5	3/1***/1**
	Bates	10	6	2**	2	3/1***/1**	5	3/1***/1**
	Zou	10	10	2/1**	2	3/1***/1**	1	4/1***/1**
	HDOCK	10	11	1**	2	3/1***/1**	5	3/1***/1**
	Bonvin	9	16	1	2	4/1***	5	4/1***
10	Chang	10	1	2/1***/1**	10	2/1***/1**	5	3/1***/1**
	Venclovas	10	6	2**	10	3/2**	1	4/3**
	LZERD	10	6	2**	10	2/1***/1**	5	3/1***/1**
	Carbone	10	6	2**	10	2/1***/1**	5	3/1***/1**
14	Weng	10	11	1**	14	2/1**	15	3/2**
	Huang	10	11	1**	14	2/1**	5	3/1***/1**
16	Grudinin	10	11	1**	16	1**	16	2/1**
17	QASDOM	5	17	0	17	0	18	0
	HAWKDOCK	7	17	0	17	0	17	1**

[a] Target participation, out of 11 (for predictors) or 10 (for scorers)
[b] Human and Server together

8.2. Appendix 2. Supplementary material for Chapter 4

Table 8.2.1 Proteins associated to diseases detected in newborn screening, with known pathological and neutral mutations.

Gene	UniProt	Disease [1]	Neutral mutations	Disease mutations
ABCD4	O14678	Methylmalonic aciduria and homocystinuria (cblJ type)	80	2
ACADM	P11310	Medium chain acyl CoA dehydrogenase deficiency	43	119
ACADS	P16219	Small chain Acyl-CoA-dehydrogenase deficiency Ethylmalonic aciduria	51	66
ACADSB	P45954	2-Methylbutyryl-CoA dehydrogenase deficiency	34	10
ACADVL	P49748	Very long chain acyl-CoA dehydrogenase deficiency Carnitine palmitoyltransferase 2 deficiency	114	153
ACAT1	P24752	Acetoacetyl-CoA-thiolase deficiency Alpha-methylacetoacetic aciduria 3-ketothiolase deficiency	22	41
ACSF3	Q4G176	Malonic & methylmalonic aciduria Methylmalonic acidaemia	47	15
ARG1	P05089	Argininemia / Arginase I deficiency	46	26
ASL	P04424	Argininosuccinic aciduria Argininosuccinate lyase deficiency	120	111
ASS1	P00966	Citrullinaemia	152	84
BCKDHA	P12694	Maple syrup urine disease	43	64
BCKDHB	P21953	Maple syrup urine disease	47	64
BTD	P43251	Biotinidase deficiency	65	180
CBS	P35520	Homocystinuria Myelomeningocele	39	130
CFTR	P13569	Cystic fibrosis Pancreatitis chronic Hypertrypsinaemia, neonatal Asthma Chronic pulmonary disease Congenital absence of vas deferens Primary sclerosing colangitis Bronchiectasis	225	963
CPT1A	P50416	Carnitine palmitoyltransferase 1 deficiency	89	34
CPT1B	Q92523	Carnitine palmitoyltransferase 1 deficiency Autism spectrum disorder Obesity	83	2
CPT1C	Q8TCG5	Carnitine palmitoyltransferase 1 deficiency	56	1
CPT2	P23786	Carnitine palmitoyltransferase 2 deficiency Hypercholesterolaemia Muscular dystrophy, limb-girdle	74	73
CYP11B1	P15538	Congenital Adrenal hyperplasia	64	88
CYP17A1	P05093	Congenital Adrenal hyperplasia Pseudohermaphroditism	29	74
CYP21A2	P08686	Congenital Adrenal hyperplasia Premature pubarche	81	170
DBT	P11182	Maple syrup urine disease	38	39
ETFA	P13804	Glutaric acidaemia type 2	116	17
ETFB	P38117	Glutaric acidaemia type 2	83	7
ETFDH	Q16134	Glutaric acidaemia type 2 Acyl-CoA dehydrogenation deficiency Coenzyme Q10 deficiency Lipid storage myopathy Subacute myopathy	128	130
FAH	P16930	Tyrosinaemia type 1	39	57
FCGR2A	P12318	Cystic fibrosis Lupus nephritis	14	3
GALT	P07902	Galactosaemia	33	249
GCDH	Q92947	Glutaric acidaemia type I	78	162
HADHA	P40939	Long-chain 3-hydroxyacyl-CoA dehydrogenase deficiency Mitochondrial trifunctional protein deficiency Rhabdomyolysis	116	31
HADHB	P55084	Mitochondrial trifunctional protein deficiency Charcot-Marie-Tooth disease Hypoparathyroidism, infantile-onset Peripheral polyneuropathy	70	35
HBB	P68871	Thalassaemia beta Sickle cell anaemia Erythrocytosis Dyspnea, progressive	31	389
HCFC1	P51610	Methylmalonic acidemia with homocystinuria Dysmorphic features Hypospadias Intellectual disability (nonsyndromic/X-linked) Mental retardation (X-linked)	215	14

HMGCL	P35914	HMG-CoA lyase deficiency 3-hydroxy-3-methylglutaric aciduria	27	30
HPD	P32754	Tyrosinaemia Hawkinsinuria	34	8
HSD3B2	P26439	Congenital adrenal hyperplasia Hypospadias / Idiopathic hypospadias Pseudohermaphroditism	48	46
IVD	P26440	Isovaleric acidaemia Complex I deficiency	73	50
LMBRD1	Q9NUN5	Methylmalonic aciduria and homocystinuria (cblF type)	202	0
MCCC1	Q96RQ3	3-methylcrotonyl-CoA carboxylase deficiency	63	60
MCCC2	Q9HCC0	3-methylcrotonyl-CoA carboxylase deficiency Complex I deficiency	36	85
MLYCD	O95822	Malonyl-CoA decarboxylase deficiency	51	17
MMAA	Q8IVH4	Methylmalonic acidaemia/aciduria (cblA type)	49	31
MMAB	Q96EY8	Methylmalonic acidemia (cblB type)	8	21
MMACHC	Q9Y4U1	Methylmalonic acidemia (cblC type) Homocystinuria (cblC type)	38	43
MMADHC	Q9H3L0	Methylmalonic aciduria (cblD type) Homocystinuria (cblD type)	93	8
MTHFR	P42898	Homocystinuria Critical congenital Heart Disease Neural tube defect Preeclampsia Inborn errors of metabolism Cleft lip and/or palate	51	94
MUT	P22033	Methylmalonic acidemia	224	224
PAH	P00439	Classic phenylketonuria Benign hyperphenylalaninaemia	86	582
PAX8	Q06710	Primary congenital hypothyroidism	136	27
PCCA	P05165	Propionic acidaemia	101	51
PCCB	P05166	Propionic acidaemia	217	63
SLC22A5	O76082	Carnitine deficiency	70	79
SLC25A13	Q9UJS0	Citrullinemia, type 2	355	64
SLC25A20	O43772	Carnitine-acylcarnitine translocase deficiency	87	19
TAT	P17735	Tyrosinaemia type 2	50	15
TGFB1	P01137	Cystic fibrosis Osteoporosis / Ostosclerosis Camurati-Engelmann Cleft-lip	65	16
TSHR	P16473	Primary congenital hypothyroidism	73	107

[1] Disease detected in newborn screening (first line) and other associated diseases.

Table 8.2.2 Structural data on analyzed proteins and their interacting partners: location of mutations in protein-protein interfaces.

		Individual				Complex	Neutral mutations [7]		Disease mutations [7]	
Gene [1]	**UniProt [1]**	**PDB [2]**	**Glob. Cov. [3]**	**Domain Cov. [4]**	**Biological Unit [5]**	**PDB [6]**	**CORE**	**RIM**	**CORE**	**RIM**
ABCD4	**O14678**	(2hyd_A)	96	98/100	Homo 2-mer	-	-	-	-	-
HSPB2	Q16082	(3j07_A)	95	100	Homo 24-mer	-	-	-	-	-
PEA15	Q15121	1n3k_A	100	100	Monomer	-			-	-
XRCC6	P12956	1jeq_A	90	100/78/97/100	Hetero 2-mer	-	-	-	-	-
SARAF	Q96BY9	-	-	-	-	-	-	-	-	-
DLEU1	O43261	-	-	-	-	-	-	-	-	-
All PPIs [8]							-	-	-	-
ACADM	**P11310**	4p13_A	92	100/100/100/100	Homo 4-mer	4p13	1	3	9	11
USP50	Q70EL3	(2hd5_A)	88	100/100	Monomer	-	-	-	-	-
ETFB	P38117	1efv_B	99	100	Hetero 2-mer	2a1t	0	0	3	2
ETFA	P13804	1efv_A	94	100/100	Hetero 2-mer	2a1t	0	0	1	4
USP20	Q9Y2K6	5z4i_A	11	0/0/0/100	Monomer	-	-	-	-	-
All PPIs [8]							***0***	***3***	***10***	***12***
ACADS	**P16219**	2vig_E	92	100/100/100/100	Homo 4-mer	2vig	2	0	8	3
IRF7	Q92985	2o61_A	65	100/59	Homo 2-mer	-	-	-	-	-
All PPIs [8]							***2***	***0***	***8***	***1***
ACADSB	**P45954**	2jif_A	88	100/100/100/100	Homo 4-mer	2jif	1	2	1	1
USP19	O94966	1wh0_A	10	100/0/0/0/15	Monomer	-	-	-	-	-
All PPIs [8]							***1***	***1***	***1***	***1***
ACADVL	**P49748**	2uxw_A	87	100/100/100/100	Homo 2-mer	2uxw	5	8	13	7
SOCS3	O14543	(2c9w_A)	85	100	Monomer	-	-	-	-	-
All PPIs [8]							*5*	***8***	***13***	*7*
ACAT1	**P24752**	2ib8_D	92	100/100/100	Homo 4-mer	2ib8	0	2	2	2
PEX7	O00628	(3w15_A)	93	93	Hetero 3-mer	-	-	-	-	-
EIF1B	O60739	(2if1_A)	100	100	Monomer	-	-	-	-	-
MPG	P29372	1f6o_A	71	97	Monomer	-	-	-	-	-
HLA-C	P30504	1im9_E	75	0/100/100	Hetero 5-mer	-	-	-	-	-
GH1	P01241	1hwh_A	85	88	Hetero 2-mer	-	-	-	-	-
TNIK	Q9UKE5	5d7a_A	21	96/96/0	Monomer	-	-	-	-	-
HLA-B	P30480	4u1m_A	76	0/100/100	Hetero 2-mer	-	-	-	-	-
MCC	P23508	-	-	-	-	-	-	-	-	-
All PPIs [8]							***0***	*2*	***1***	*2*
ACSF3	**Q4G176**	(3nyq_A)	89	100/93	Monomer	N/A				
MATN4	O95460	(1ck4_A) (1min_A)	30	0/0/100/100/100/100/0/0/0/0	Monomer	-	-	-	-	-
TRIM27	P14373	(4cg4_A) (2cvs_A)	71	0/100/100/0/100/0	Homo 2-mer	-	-	-	-	-
PPP1R13B	Q96KQ4	(1ycs_B)	18	100/100/100/100/100/100	Monomer	-	-	-	-	-
KRT40	Q6A162	-	-	-	-	-	-	-	-	-
ARG1	**P05089**	3gmz_A	98	100	Homo 3-mer	2aeb	0	2	2	0
FLOT1	O75955	(1win_A)	32	77/0	Monomer	-	-	-	-	-
USP53	Q70EK8	-	-	-	-	-	-	-	-	-
All PPIs [8]							***0***	*2*	*2*	***0***

ASL	**P04424**	1k62_B	99	100/100	Homo 4-mer	1k62	6	8	26	12
HMOX1	P09601	(1wov_A)	81	100	Homo 2-mer	-	-	-	-	-
TRIM3	O75382	(1jm7_A)	10	0/0/0/0/0/100/0/0/100/100/100/100/100/0	Homo 2-mer	-	-	-	-	-
WDYHV1	Q96HA8	4w79_A	98	100	Monomer	-	-	-	-	-
MCMBP	Q9BTE3	4kg9_B	1	1	Monomer	-	-	-	-	-
USP53	Q70EK8	-	-	-	-	-	-	-	-	-
All PPIs [$]							*6*	*7*	*26*	*5*
ASS1	**P00966**	2nz2_A	98	99	Homo 4-mer	2nz2	11	12	13	6
ARAF	P10398	(3lb7_A) (1faq_A)	45	100/100/100/100/0	Homo 2-mer Monomer	-	-	-	-	-
PTPN1	P18031	(1lar_A)	82	100	Monomer	-	-	-	-	-
TERF2	Q15554	3bu8_B	38	0/100/0	Homo 4-mer	-	-	-	-	-
USP4	Q13107	2y6e_B (3ivu_A)	35	100/52/81/100	Homo 6-mer Homo 2-mer	-	-	-	-	-
TERF1	P54274	3bqo_A 1ity_A	46	100/97	Homo 2-mer Monomer	-	-	-	-	-
MLH1	P40692	(3na3_A) 3rbn_A	44	100/100/100/94	Homo 2-mer	-	-	-	-	-
ASCC2	Q9H1I8	2di0_A	9	100	Monomer	-	-	-	-	-
JAK2	O60674	4gmy_A 4z32_A 4fvr_A	39	100/100/100/100/100	Monomer	-	-	-	-	-
CDR2	Q01850	-	-	-	-	-	-	-	-	-
AKTIP	Q9H8T0	-	-	-	-	-	-	-	-	-
All PPIs [$]							*11*	*12*	*13*	*3*
BCKDHA	**P12694**	1olx_A	88	98	Hetero 4-mer	2bfd	1	1	3	3
BCKDHB	P21953	2bev_B	86	100/100	Hetero 4-mer	2bfd	1	1	6	3
BCKDK	O14874	(1gkz_A)	83	100/100	Monomer	-	-	-	-	-
All PPIs [$]							*2*	*2*	*9*	*3*
BCKDHB	**P21953**	2bev_B	86	100/100	Hetero 4-mer	2bfd	0	0	6	2
BCKDHA	P12694	1olx_A	88	98	Hetero 4-mer	2bfd	0	0	9	2
All PPIs [$]							*0*	*0*	*14*	*1*
BTD	**P43251**	(4cyf_A)	90	100	Monomer	N/A				
CBS	**P35520**	4l3v_A	90	100/100	Homo 2-mer	4l28	1	5	6	5
PRKAR1A	P10644	(4din_B)	96	100/100/100	Monomer	-	-	-	-	-
PIAS1	O75925	(4mvt_A)	43	100/100	Monomer	-	-	-	-	-
EHHADH	Q08426	(3zwc_A)	100	100/100/100/100/100	Monomer	-	-	-	-	-
UBASH3A	P57075	(3d4i_A) (1gri_A)	42	100/100	Homo 2-mer	-	-	-	-	-
FXR2	P51116	(4qw2_A) (2qnd_A)	29	100/0/4/100	Homo 2-mer	-	-	-	-	-
CBX4	O00257	2k28_A	11	0/100	Monomer	-	-	-	-	-
PRKAG1	P54619	5ezv_E	91	100/100/100	Hetero 3-mer	-	-	-	-	-
ZDHHC17	Q8IUH5	3eu9_A	34	100/0/100/100/100/100/100/100/100/100	Homo 6-mer	-	-	-	-	-
WDYHV1	Q96HA8	4w79_A	98	100	Monomer	-	-	-	-	-
PSMA1	P25786	4r3o_F	90	100/100	Hetero 28-mer	-	-	-	-	-
PIAS3	Q9Y6X2	4mvt_A	41	100/82	Monomer	-	-	-	-	-
UGP2	Q16851	3r2w_D	92	95	Homo 8-mer	-	-	-	-	-
VTA1	Q9NP79	2lxm_A	53	100	Hetero 2-mer	-	-	-	-	-
RANBP9	Q96S59	5jiu_B	29	100/0/0	Homo 2-mer	-	-	-	-	-
COL4A3BP	Q9Y5P4	2e3p_A 4hhv_A	38	100/100/100	Homo 2-mer	-	-	-	-	-
PIN1	Q13526	1nmv_A	100	100/100/100	Monomer	-	-	-	-	-
HID1	Q8IV36	-	-	-	-	-	-	-	-	-
All PPIs [$]							*1*	*5*	*6*	*4*

CFTR	**P13569**	5uak_A	77	100/91/6/100	Homo 5-mer Monomer	1xmi	4	5	26	30
LPAR2	Q9HBW0	(4z35_A)	87	100/100	Monomer	-	-	-	-	-
CANX	P27824	(1jhn_A)	67	0/100	Monomer	-	-	-	-	-
DERL1	Q9BUN8	5glf_B	5	0	Homo 4-mer	-	-	-	-	-
BCAP31	P51572	4jzl_C	25	24	Homo 2-mer	-	-	-	-	-
SLC9A3R1	O14745	2krg_A 2m0u_A	60	100/100/100/ 100/100	Monomer	-	-	-	-	-
FLNA	P21333	(4b7l_A) 1wlh_A) 2j3s_A 2k3t_A 4m9p_A 2k7p_A 3rgh_A 3cnk_A	13	100/3/100/ 100/100/0/0/0/ 0/100/0/100/ 100/100/100/ 100/0/100/100 /0/100/95/100/ 100	Monomer Homo 2-mer	-	-	-	-	-
PPP2R1A	P30153	1b3u_B	100	100/100	Monomer	-	-	-	-	-
YWHAZ	P63104	4zdr_A	96	99	Homo 2-mer	-	-	-	-	-
GOPC	Q9HD26	4q6h_A	19	100	Monomer	-	-	-	-	-
PRKAG1	P54619	5ezv_E	91	100/100/100	Hetero 3-mer	-	-	-	-	-
GORASP2	Q9H8Y8	4edj_A	45	100/100	Monomer	-	-	-	-	-
PDZD3	Q86UT5	2v90_A	17	0/0/100/0/0	Homo 6-mer	-	-	-	-	-
YWHAG	P61981	4j6s_B	97	100	Monomer	-	-	-	-	-
CFTR	P13569	5uak_A	77	100/91/6/100	Monomer Homo 5-mer	-	-	-	-	-
KCNN4	O15554	-	-	-	-	-	-	-	-	-
KRT8	P05787	-	-	-	-	-	-	-	-	-
ANO1	Q5XXA6	-	-	-	-	-	-	-	-	-
SLC26A8	Q96RN1	-	-	-	-	-	-	-	-	-
RNF5	Q99942	-	-	-	-	-	-	-	-	-
DERL1	Q9BUN8	-	-	-	-	-	-	-	-	-
All PPIs [$]							***1***	***3***	***21***	***16***
CPT1A	**P50416**	(2fy5_A)	77	100/0	Monomer			N/A		
CLIC1	O00299	3uvh_A	96	92/100/100	Homo 2-mer	-	-	-	-	-
CACYBP	Q9HB71	1x5m_A 2jtt_C (1ysm_A)	56	69/100/55	Monomer	-	-	-	-	-
KBTBD7	Q8WVZ9	-	-	-	-	-	-	-	-	-
CPT1B	**Q92523**	(1nm8_A)	77	0/100	Monomer			N/A		
TEX44	Q53QW1	-	-	-	-	-	-	-	-	-
CPT1C	**Q8TCG5**	2m76_A	6	100/0	Monomer			N/A		
CPT2	**P23786**	(2deb_A)	95	100	Monomer			N/A		
CYP11B1	**P15538**	(4dvq_A)	93	100	Homo 6-mer	-	-	-	-	-
CYP17A1	**P05093**	5irv_C	93	100	Monomer			N/A		
CDH1	P12830	(3q2v_A) (1op4_A)	61	100/0/100/100 /100/100/100	Monomer	-	-	-	-	-
TIGD7	Q6NT04	(1hlv_A)	24	100/88/0	Monomer	-	-	-	-	-
APC	P25054	3nmz_A 1deb_B 1m5i_A 1th1_C 1emu_B 2rqu_B	14	0/0/0/0/0/0/0/ 0/0/100/0/100/ 2/0/0/50/76/93	Monomer	-	-	-	-	-
PTPRJ	Q12913	2nz6_A	22	93	Monomer	-	-	-	-	-
IGF1R	P08069	1igr_A 1p4o_B	34	100/100/100/ 100/100	Monomer	-	-	-	-	-
CDO1	Q16878	2ic1_A	92	97	Monomer	-	-	-	-	-
CHEK2	O96017	3i6u_A	72	91/100/91/91	Homo 2-mer	-	-	-	-	-
HSP90AB1	P08238	5fwk_B	87	100/100/87	Monomer Hetero 4-mer	-	-	-	-	-
BARX1	Q9HBU1	2dmt_A	31	100	Monomer	-	-	-	-	-
STK11	Q15831	2wtk_C	66	99/100/99	Monomer Hetero 3-mer	-	-	-	-	-
ERBB2	P04626	3n85_A	48	99/100/99/92/	Hetero 9-mer	-	-	-	-	-

		3pp0_B		100/97	Monomer					
ATP5E	P56381	(2v7q_I)	92	96	Homo 3-mer	-	-	-	-	-
RPAIN	Q86UA6	-	-	-	-	-	-	-	-	-
CYP21A2	**P08686**	4y8w_C	89	96	Monomer			N/A		
DBT	**P11182**	(2ii3_A) 2coo_A 1k8m_A	48	100/100/100	Homo 24-mer Monomer	-	-	-	-	-
DLD	P09622	1zmd_A	93	100/100/100/ 100/100/100/ 100	Homo 2-mer	-	-	-	-	-
USP16	Q9Y5T5	2i50_A	15	100/0/0	Monomer	-	-	-	-	-
ADI1	Q9BV57	4qgn_A	99	100/100	Monomer	-	-	-	-	-
ETFA	**P13804**	1efv_A	94	100/100	Hetero 2-mer			N/A		
ETFB	P38117	1efv_B	99	100	Hetero 2-mer	1efv	5	6	2	1
ACADM	P11310	4p13_A	92	100/100100/ 100	Homo 4-mer	2a1t	0	0	0	2
HSPB2	Q16082	(3j07_A)	94	100	Homo 24-mer	-	-	-	-	-
IKBKE	Q14164	(4im0_A)	90	100/100	Homo 2-mer	-	-	-	-	-
FTSJ1	Q9UET6	(1ej0_A)	55	100	Monomer	-	-	-	-	-
EIF1B	O60739	(2ifl_A)	100	100	Monomer	-	-	-	-	-
ELF3	P78545	3jtg_A 2e8p_A	25	100/100	Monomer	-	-	-	-	-
TRAF6	Q9Y4K3	3hcs_A	30	12/100/100/ 100/100/100/ 100/	Homo 2-mer	-	-	-	-	-
CDC42	P60953	5c2j_B	100	100/100/100	Hetero 2-mer	-	-	-	-	-
PSEN1	P49768	5fn5_B	48	54	Monomer	-	-	-	-	-
MPG	P29372	1f6o_A	71	97	Monomer	-	-	-	-	-
PRKAB1	Q9Y478	5ezv_D	69	100/100	Hetero 3-mer	-	-	-	-	-
MCC	P23508	-	-	-	-	-	-	-	-	-
All PPIs [$]							*5*	*6*	*2*	*1*
ETFB	**P38117**	1efv_B	99	100	Hetero 2-mer			N/A		
ETFA	P13804	1efv_A	94	100/100	Hetero 2-mer	1efv	12	10	1	0
ACADM	P11310	4p13_A	92	100/100/100/ 100	Homo 4-mer	2a1t	0	2	0	1
IKBKE	Q14164	(4im0_A)	90	100/100	Homo 2-mer	-	-	-	-	-
All PPIs [$]							*11*	*11*	*1*	*3*
ETFDH	**Q16134**	(2gmh_A)	94	100/100/100	Homo 2-mer	-	-	-	-	-
ZNF581	Q9P0T4	(1ubd_C)	53	100/100	Monomer	-	-	-	-	-
MYH7B	A7E2Y1	(4p7h_B)	40	0/100/100	Monomer	-	-	-	-	-
KRTAP13-2	Q52LG2	-	-	-	-	-	-	-	-	-
OTX1	P32242	(2dms_A)	21	100/0	Monomer	-	-	-	-	-
KRTAR11-1	Q8IUC1	-	-	-	-	-	-	-	-	-
HSD17B10	Q99714	1u7t_B	98	100/100/100	Homo 4-mer	-	-	-	-	-
AIRE	O43918	1xwh_A 2lri_C	12	100/12/0	Monomer	-	-	-	-	-
TRIM69	Q86WT6	(2iwg_B)	36	0/0/100/100/0/ 0	Monomer	-	-	-	-	-
FAH	**P16930**	(1hyo_A)	99	100/100	Homo 2-mer	-	-	-	-	-
EGFR	P00533	3qwq_A 3w32_A	51	100/100/100/ 100/100/100	Monomer Homo 2-mer	-	-	-	-	-
TCF4	P15884	2kwf_B	3	0	Monomer	-	-	-	-	-
KRTAP5-9	P26371	-	-	-	-	-	-	-	-	-
KRTAP10-8	P60410	-	-	-	-	-	-	-	-	-
ADAMTSL4	Q6UY14	-	-	-	-	-	-	-	-	-
SERTAD1	Q9UHV2	-	-	-	-	-	-	-	-	-
FCGR2A	**P12318**	1h9v_A	54	100/100/100/ 100	Homo 2-mer Monomer	1fcg	0	0	1	0
UBQLN1	Q9UMX0	(1j8c_A) 2iv5_A	18	100/100/98	Monomer	-	-	-	-	-
APCS	P02743	4avv_A	92	100	Homo 5-mer	3d5o	0	0	0	0
IGHG1	P01857	1hzh_H	100	100/100/100/ 100	Homo 2-mer	3ry6	0	0	0	1

CRP	P02741	3l2y_A	92	100/100	Homo 5-mer	-	-	-	-	-
All PPIs [$]							***0***	***0***	***1***	***0***
GALT	**P07902**	5in3_B	87	91/99	Homo 2-mer	5in3	1	2	18	9
GRPEL1	Q9HAV7	(4ani_A)	76	98	Homo 4-mer	-	-	-	-	-
TNK2	Q07912	4hzs_D 1cf4_B	31	95/62/95/100/0	Monomer	-	-	-	-	-
TRIP13	Q15645	5vqa_A	85	88/84/86	Monomer	-	-	-	-	-
NQO1	P15559	1gg5_A	100	100	Homo 2-mer	-	-	-	-	-
KRTAP6-2	Q3LI66	-	-	-	-	-	-	-	-	-
All PPIs [$]							***0***	***2***	***17***	***9***
GCDH	**Q92947**	1siq_A	89	100/100/100	Homo 4-mer	1siq	9	11	19	7
NOS3	P29474	(1tll_A) 5uoc_A	55	100/100/99/100	Homo 2-mer	-	-	-	-	-
RNF32	Q9H0A6	(1iym_A)	12	100/96/100/100/96/100/100/0/100/0	Monomer	-	-	-	-	-
APOE	P02649	2l7b_A	95	100	Monomer	-	-	-	-	-
CDC37	Q16543	5fwp_E 2n5x_A	68	99/88/100	Monomer	-	-	-	-	-
PSEN1	P49768	5fn5_B	48	54	Monomer Homo 2-mer	-	-	-	-	-
GRB2	P62993	1gri_A	97	88/100/97/100/100/100	Homo 2-mer	-	-	-	-	-
A2M	P01023	4acq_C	95	96/100/100/100/100/100	Homo 4-mer	-	-	-	-	-
EXOC6	Q8TAG9	(2a2f_X)	40	90	Monomer	-	-	-	-	-
ECSIT	Q9BQ95	-	-	-	-	-	-	-	-	-
All PPIs [$]							***9***	***9***	***18***	***4***
HADHA	**P40939**	(2wtb_A)	94	100/100/100/100	Monomer	N/A				
IKBKE	Q14164	(4im0_A)	90	100/100	Homo 2-mer	-	-	-	-	-
TRAF6	Q9Y4K3	3hcs_A	30	12/100/100/100/100/100/100/	Homo 2-mer	-	-	-	-	-
PRKAB1	Q9Y478	5ezv_D	69	100/100	Hetero 3-mer	-	-	-	-	-
GABARAP	O95166	1kot_A	100	100/100	Monomer	-	-	-	-	-
BCAR3	O75815	3t6a_C	37	0/100	Monomer	-	-	-	-	-
GABARAPL1	Q9H0R8	5lxi_B	100	100/100	Homo 2-mer	-	-	-	-	-
HSD17B10	Q99714	1u7t_B	98	100/100/100	Homo 4-mer	-	-	-	-	-
EPB41	P11171	1gg3_C	32	100/100/0/96/0/0	Monomer	-	-	-	-	-
HLA-B	P30480	4u1m_A	76	0/100/100	Monomer	-	-	-	-	-
MAP1LC3B	Q9GZQ8	2zjd_C	98	100/100	Monomer	-	-	-	-	-
GABARAPL2	P60520	4co7_B	100	100/100	Monomer	-	-	-	-	-
USP49	Q70CQ1	-	-	-	-	-	-	-	-	-
HADHB	**P55084**	(2iik_A)	89	100/100/100	Homo 2-mer	-	-	-	-	-
HSPB2	Q16082	(3j07_A)	94	100	Homo 24-mer	-	-	-	-	-
SMN1	Q16637	-	-	-	-	-	-	-	-	-
ANXA7	P20073	(1dm5_A)	64	100/100/100	Homo 6-mer	-	-	-	-	-
TNNI3K	Q59H18	(1n11_A) 4yfi_C	42	93/100/100/0/100/100/100/100/100/100/100/100/100/100	Monomer Homo 2-mer	-	-	-	-	-
DUX1	O43812	(1fjl_A)	36	0/100	Homo 2-mer	-	-	-	-	-
TK1	P04183	1w4r_A	74	100	Homo 4-mer	-	-	-	-	-
RCC1	P18754	1a12_C	95	100/100/100/100/100/100/100/100	Homo 3-mer	-	-	-	-	-
CDKN1A	P38936	4rjf_B	12	0	Hetero 6-mer	-	-	-	-	-
GABARAPL1	Q9H0R8	5lxi_B	100	100/100	Homo 2-mer	-	-	-	-	-
HSD17B10	Q99714	1u7t_B	98	100/100/100	Homo 4-mer	-	-	-	-	-
MAP1LC3C	Q9BXW4	2ncn_A	86	100	Monomer	-	-	-	-	-

GRB7	Q14451	(3hk0_A)	47	100/100	Homo 2-mer	-	-	-	-	-
CDKN1A	P38936	-	-	-	-	-	-	-	-	-
HBB	**P68871**	1dxt_D	100	100	Homo 2-mer	1bbb	0	1	1	3
HP	P00738	(1gpz_A)	91	100	Monomer	4wjg	1	2	6	6
HBZ	P02008	3w4u_C	99	100	Homo 2-mer	3w4u	2	2	15	9
HBA1	P69905	1bz1_C	100	100	Homo 2-mer	2w72	2	3	17	12
SELENOT	P62341	-	-	-	-	-	-	-	-	-
All PPIs [$]							***1***	***3***	***17***	***8***
HCFC1	**P51610**	4go6_D	8	0	Homo 4-mer	4go6	5	4	1	0
GABPB1	Q06547	(1n11_A)	85	100/100/100/ 100/100/100/ 100/100/100	Monomer	-	-	-	-	-
CLP1	Q92989	(4ohx_A)	97	100/100/100	Monomer	-	-	-	-	-
MAGI1	Q96QZ7	2kpl_A 2q9v_A 2r4h_C 2yse_A 2kpl_A 2bpu_A (1ex7_A)	9	100/97/100/0/ 100/3/7/100/ 100/100	Monomer	-	-	-	-	-
ASF1B	Q9NVP2	(2hue_A)	80	99	Monomer	-	-	-	-	-
E2F1	Q01094	2aze_B	23	100	Monomer	-	-	-	-	-
E2F3	O00716	(1cf7_A)	14	97/0	Monomer	-	-	-	-	-
E2F4	Q16254	5tuu_B 1cf7_A	26	98/100	Monomer	-	-	-	-	-
OGT	O15294	4xif_B 1w3b_A	67	100/100/100/ 100/100/100/ 100/100/100/ 100/100/100/ 100/100/100/ 100/100/100/ 100/78/100/ 100/100/100/ 100/100/100	Monomer Homo 2mer	-	-	-	-	-
FOXO3	O43524	2k86_A 2lqi_B	15	46/30/100	Monomer	-	-	-	-	-
THAP1	Q9NVV9	2jtg_A	41	100	Monomer	-	-	-	-	-
SIRT1	Q96EB6	5btr_A	49	100	Monomer	-	-	-	-	-
SIN3A	Q96ST3	2n2h_B 2l9s_B 2rms_B	10	0/46/100/100/ 0	Monomer	-	-	-	-	-
FHL3	Q13643	2ehe_A 2cuq_A	29	100/24/100/7	Monomer	-	-	-	-	-
THAP11	Q96EK4	2lau_A	26	100	Monomer	-	-	-	-	-
ZBTB17	Q13105	4u2m_D 2lvr_A 2n26_A 2n25_A 2m0f_A 2m0d_A 2lvu_A	30	100/52/56/56/ 52/54/4/100/ 100	Monomer	-	-	-	-	-
FHL2	Q14192	2d8z_A 2miu_A 1x4l_A 1x4k_A	25	100/100/100/ 100	Monomer	-	-	-	-	-
RNF2	Q99496	4r8p_L 3ixs_A	76	100/100/100/ 100/100	Monomer Homo 2-mer	-	-	-	-	-
TET2	Q6N021	5d9y_A	21	39	Monomer	-	-	-	-	-
NFE2L1	Q14494	-	-	-	-	-	-	-	-	-
SP1	P08047	-	-	-	-	-	-	-	-	-
PDCD2	Q16342	-	-	-	-	-	-	-	-	-
USP53	Q70EK8	-	-	-	-	-	-	-	-	-
BAP1	Q92560	-	-	-	-	-	-	-	-	-
THAP7	Q9BT49	-	-	-	-	-	-	-	-	-

CREBZF	Q9NS37	-	-	-	-	-	-	-	-	-
CREB3	O43889	-	-	-	-	-	-	-	-	-
TET3	O43151	-	-	-	-	-	-	-	-	-
All PPIs [$]							*5*	*2*	*1*	*0*
HMGCL	**P35914**	2cw6_A	91	100	Homo 2-mer	2cw6	0	1	0	0
RNF126	Q9BV68	(1x4j_A) 2n9p_A	21	100/100/100/ 100/100/100/ 100/100/100	Monomer	-	-	-	-	-
ADAMTS10	Q9H324	(1lsl_A)	10	0	Monomer	-	-	-	-	-
DNAJA1	P31689	2lo1_A (1nlt_A)	58	100/100/100	Monomer	-	-	-	-	-
HSD17B10	Q99714	1u7t_A	98	100/100/100	Homo 4-mer	-	-	-	-	-
HES1	Q14469	2mh3_A	25	0/100	Homo 2-mer	-	-	-	-	-
GTF2B	Q00403	5iyb_M	98	100/100/100/ 100	Monomer Hetero 21-mer	-	-	-	-	-
ARL6IP1	Q15041	-	-	-	-	-	-	-	-	-
MS4A7	Q9GZW8	-	-	-	-	-	-	-	-	-
All PPIs [$]							*0*	*1*	*0*	*0*
HPD	**P32754**	3isq_A	96	100/100/100	Homo 2-mer	3isq	1	5	1	0
CDKN1A	P38936	4rjf_B	12	0	Monomer Hetero 6-mer	-	-	-	-	-
IKBKG	Q9Y6K9	4bwn_A 3brt_D	20	98/92	Homo 2-mer	-	-	-	-	-
All PPIs							*1*	*5*	*1*	*0*
HSD3B2	**P26439**	(3wj7_A)	95	100	Homo 3-mer	-	-	-	-	-
IVD	**P26440**	1ivh_A	91	100/100/100/ 100	Homo 4-mer	1ivh	1	2	5	3
ACTN3	Q08043	(4d1e_A)	95	100/100/100/ 100/100/100/ 100/100	Homo 2-mer	-	-	-	-	-
GPSM3	Q9Y4H4	-	-	-	-	-	-	-	-	-
All PPIs [$]							*1*	*2*	*3*	*2*
LMBRD1	**Q9NUN5**	-	-	-	-	-	-	-	-	-
MCCC1	**Q96RQ3**	(3u9s_A)	92	100/100/100/100	Hetero 12-mer	N/A				
MCCC2	**Q9HCC0**	(3u9r_B)	95	100	Homo 6-mer	-	-	-	-	-
HSPB2	Q16082	(3j07_A)	94	100	Homo 24-mer	-	-	-	-	-
MLYCD	**O95822**	4f0x_A	92	100/100	Homo 2-mer Homo 4-mer	2ygw, 4f0x	1	3	2	4
UBALD1	Q8TB05	(2dzl_A)	33	100	Monomer	-	-	-	-	-
RBM5	P52756	2lkz_A	12	0/0/100/0	Monomer	-	-	-	-	-
All PPIs [$]							*1*	*2*	*2*	*2*
MMAA	**Q8IVH4**	2www_D	75	89	Homo 2-mer	2www	1	2	0	2
All PPIs [$]							*1*	*2*	*0*	*2*
MMAB	**Q96EY8**	2idx_C	72	98	Homo 3-mer	2idx	0	1	3	1
SPERT	Q8NA61	-	-	-	-	-	-	-	-	-
All PPIs [$]							*0*	*1*	*2*	*1*
MMACHC	**Q9Y4U1**	3sc0_A	84	100	Monomer	N/A				
MTR	Q99707	4ccz_A (3bul_A)	48	100/100/100/ 92/100	Monomer	-	-	-	-	-
MMADHC	Q9H3L0	5cv0_A	57	62	Homo 2-mer	-	-	-	-	-
CREB5	Q02930	-	-	-	-	-	-	-	-	-
MMADHC	**Q9H3L0**	5cv0_A	57	62	Homo 2-mer	5cv0	3	8	0	0
MMACHC	Q9Y4U1	3sc0_A	84	100	Monomer	-	-	-	-	-
All PPIs [$]							*3*	*8*	*0*	*0*
MTHFR	**P42898**	(1v93_A)	45	100	Monomer	N/A				
LSM8	O95777	(4c92_A)	96	100	Monomer Hetero 7-mer	-	-	-	-	-

MUT	**P22033**	3bic_	95	99	Homo 2-mer	2xiq	6	8	11	3
PCBD1	P61457	1dcp_A	95	100	Homo 4-mer	-	-	-	-	-
All PPIs [$]							***6***	***8***	***10***	***2***
PAH	**P00439**	(5den_A)	95	100	Homo 2-mer	1j8u, 5fii	0	3	13	12
QDPR	P09417	1hdr_A	97	100	Monomer	-	-	-	-	-
PCBD1	P61457	1dcp_A	95	100	Homo 4-mer	-	-	-	-	-
All PPIs [$]							***0***	***2***	***13***	***11***
PAX8	**Q06710**	2k27_A	35	100/0	Monomer	N/A				
UNK	Q9C0B0	5elh_A (5elk_A)	17	100/100	Homo 2-mer	-	-	-	-	-
ANXA7	P20073	(1dm5_A)	64	100/100/100/ 100	Homo 6-mer	-	-	-	-	-
CXCL9	Q07325	(1qnk_A)	56	100	Homo 2-mer	-	-	-	-	-
GCM2	O75603	(1odh_A)	31	100	Monomer	-	-	-	-	-
HOXC8	P31273	(2r5y_A)	31	100	Monomer	-	-	-	-	-
AES	Q08117	(4om3_A)	59	87	Homo 4-mer	-	-	-	-	-
HOXC9	P31274	2msy_A	26	0/100	Monomer	-	-	-	-	-
SAE1	Q9UBE0	1y8r_A	91	93	Monomer, Hetero 3-mer	-	-	-	-	-
SMAD4	Q13485	1dd1_B 3qsv_A	45	100/96	Homo 2-mer, Homo 3-mer	-	-	-	-	-
HGS	O14964	4avx_A 3fli_H	28	100/100/100	Monomer	-	-	-	-	-
CHUK	O15111	5ebz_A	88	100/100/0	Homo 6-mer	-	-	-	-	-
NCL	P19338	2krr_A 2fc9_A 2fc8_A	24	100/100/100	Monomer	-	-	-	-	-
PKM	P14618	3gr4_A	98	100/100	Homo 4-mer	-	-	-	-	-
EP300	Q09472	4bhw_A 1l3e_B 3io2_A (1tot_A)	23	82/100/100/0/ 0/100/100/100	Homo 2-mer	-	-	-	-	-
POU2AF1	Q16633	1cqt_J	9	9.2	Monomer	-	-	-	-	-
RB1	P06400	4elj_A 2aze_C	64	100/98/28/93	Monomer	-	-	-	-	-
RPLP2	P05387	4beh_B	100	100	Monomer, Hetero 2-mer	-	-	-	-	-
CDK3	Q00526	-	-	-	-	-	-	-	-	-
CLSTN1	O94985	-	-	-	-	-	-	-	-	-
SS18L1	O75177	-	-	-	-	-	-	-	-	-
LONRF1	Q17RB8	-	-	-	-	-	-	-	-	-
RSBN1	Q5VWQ0	-	-	-	-	-	-	-	-	-
TMA16	Q96EY4	-	-	-	-	-	-	-	-	-
SERINC1	Q9NRX5	-	-	-	-	-	-	-	-	-
PCCA	**P05165**	(3n6r_A)	91	100/100/100/ 100/100/100/ 100	Monomer, Hetero 12-mer	N/A				
ERRFI1	Q9UJM3	4zjv_C	10	47	Monomer	-	-	-	-	-
PCCB	P05166	(3n6r_B)	93	100	Homo 6-mer	-	-	-	-	-
PARP2	Q9UGN5	-	-	-	-	-	-	-	-	-
MCC	P23508	-	-	-	-	-	-	-	-	-
PCCB	**P05166**	(3n6r_B)	93	100	Homo 6-mer	-	-	-	-	-
PCCA	P05165	(3n6r_A)	91	100/100/100/ 100/100/100/ 100	Monomer, Hetero 12-mer	-	-	-	-	-
ACTN3	Q08043	(4d1e_A)	95	100/100/100/ 100/100/100/ 100/100	Homo 2-mer	-	-	-	-	-
CSNK2B	P67870	4dgl_B	93	100	Monomer	-	-	-	-	-
SLC22A5	**O76082**	-	-	-	-	-	-	-	-	-
SLC9A3R2	Q15599	-	-	-	-	-	-	-	-	-
PDZD3	Q86UT5	-	-	-	-	-	-	-	-	-
SLC9A3R1	O14745	-	-	-	-	-	-	-	-	-

MTUS2	Q5JR59	-	-	-	-	-	-	-	-	-
NOTCH2NL	Q7Z3S9	-	-	-	-	-	-	-	-	-
KRT27	Q7Z3Y8	-	-	-	-	-	-	-	-	-
SLC25A13	**Q9UJS0**	4p5w_A	46	0/0/0	Homo 2-mer	4p5w	1	9	4	3
USP50	Q70EL3	(2hd5_A)	88	100/100	Monomer	-	-	-	-	-
All PPIs [$]							***1***	***9***	***4***	***3***
SLC25A20	**O43772**	(1okc_A)	95	100/100/94	Monomer			N/A		
TAT	**P17735**	3dyd_A	85	0/100	Homo 2-mer	3dyd	2	1	0	1
UBE3A	Q05086	1c4z_A 2kr1_A	40	100/98	Monomer	-	-	-	-	-
GLUL	P15104	2ojw_C	97	100/100	Homo 10-mer	-	-	-	-	-
JPT2	Q9H910	(4hpo_L)	97	100	Monomer	-	-	-	-	-
All PPIs [$]							***2***	***1***	***0***	***1***
TGFB1	**P01137**	5ffo_H	83	95/94	Monomer, Homo 2-mer	5ffo	1	3	3	1
CDKN2A	Q8N726	(2g7l_A)	98	100	Homo 2-mer	-	-	-	-	-
KLK9	Q9UKQ9	(2psx_A)	91	100	Monomer	-	-	-	-	-
FSTL1	Q12841	(1lr7_A)	22	100/100/100	Monomer	-	-	-	-	-
CCL5	P13501	(1el0_A)	80	100	Monomer	-	-	-	-	-
BMPR1A	P36894	(3mdy_A) 2k3g_A	62	100/100/100/ 100	Monomer	-	-	-	-	-
MMP9	P14780	(1ck7_A)	96	100/100/100/ 100/100/100/ 100/100	Monomer	-	-	-	-	-
LCE3B	Q5TA77	(3tjq_A)	93	99	Monomer	-	-	-	-	-
BUB1	O43683	(3e7e_A) 2lah_A	32	100/100/100/ 100	Monomer	-	-	-	-	-
DIRAS3	O95661	(2p67_A)	95	100/100/100	Monomer	-	-	-	-	-
AXIN2	Q9Y2T1	(1dk8_A) (1wsp_A)	17	0/100/100/11	Homo 2-mer Homo 3-mer	-	-	-	-	-
TGFBR3	Q03167	(3qw9_A)	20	51	Homo 2-mer	-	-	-	-	-
THBS1	P07996	(1yo8_A) 5foe_B 2es3_A 3r6b_A	53	100/100/98/10 0/100/100/100 /100/100/100/ 100/96	Monomer	-	-	-	-	-
THRSP	Q92748	(3omt_A)	94	95	Homo 2-mer	-	-	-	-	-
MEOX2	P50222	(1puf_A)	19	100	Monomer	-	-	-	-	-
LAMB1	P07942	(4aqs_A)	26	100/100/72/0/ 0/0/0/0/0/0/0/ 0/100	Monomer	-	-	-	-	-
PRSS50	Q9UI38	(1z8g_A)	86	100/100	Monomer	-	-	-	-	-
HOXC6	P09630	(1ftz_A)	28	100	Monomer	-	-	-	-	-
MTA3	Q9BTC8	(4bkx_A)	28	0/0/0/100	Monomer	-	-	-	-	-
HOXA1	P49639	(1b72_A)	27	100	Monomer	-	-	-	-	-
LCE3D	Q9BYE3	(1m8n_A)	99	100	Monomer	-	-	-	-	-
NSD3	Q9BZ95	(4yz8_A) 4rxj_A	15	0/100/100	Homo 2-mer	-	-	-	-	-
CST6	Q15828	(2ch9_A)	85	100	Homo 2-mer	-	-	-	-	-
ERBB2	P04626	3n85_A 3pp0_B	48	99/100/99/92/ 100/97	Monomer	-	-	-	-	-
TGFBR1	P36897	3faa_E 2l5s_A	66	100/100/100/ 100	Monomer	3kfd	0	0	0	0
KLK7	P49862	2qxj_A	88	100	Monomer	-	-	-	-	-
AGR3	Q8TD06	3ph9_B	81	100	Monomer	-	-	-	-	-
WFIKKN1	Q96NZ8	2ddi_A	13	0/0/0/0/0/ 100/0/0/0	Monomer	-	-	-	-	-
LAMTOR5	O43504	3msh_A	99	100	Homo 2-mer	-	-	-	-	-
ITGAV	P06756	3ije_A	89	93/100/100	Monomer, Hetero 2-mer	5ffo	1	3	0	0
FNTA	P49354	2h6i_A	83	100/100/100/ 100/100	Monomer, Hetero 2-mer	-	-	-	-	-
LPXN	O60711	1x3h_A (1v6g_A)	21	5/100/100/100	Monomer	-	-	-	-	-

		1m3v_A)								
TLR2	O60603	2z7x_A 1fyx_A	70	98/98/100/100 /100/100/100	Monomer, Hetero 2-mer	-	-	-	-	-
APC	P25054	3nmz_A 1emu_B 2rqu_B 1deb_B 1m5i_A	14	0/0/0/0/0/0/0/ 0/0/100/0/100/ 0/0/0/50/96/93	Monomer, Hetero 4-mer	-	-	-	-	-
PDLIM2	Q96JY6	3pdv_A	25	100/0/0	Monomer	-	-	-	-	-
DAXX	Q9UER7	2kzs_A	13	98	Monomer	-	-	-	-	-
SNAI1	O95863	3w5k_B	42	92/100	Monomer, Hetero 2-mer	-	-	-	-	-
FBXW7	Q969H0	2ovq_B	63	100/100/100/ 100/100/100	Monomer, Hetero 2-mer	-	-	-	-	-
DLC1	Q96QB1	3kuq_A 2gky_A	13	100/100/0	Monomer	-	-	-	-	-
FKBP1A	P62942	3mdy_B	100	100	Monomer, Hetero 2-mer	-	-	-	-	-
AURKA	O14965	2j4z_B	67	100/100	Monomer	-	-	-	-	-
ST14	Q9Y5Y6	4jyt_A	28	100/0/0/0/0/0/ 0/0	Monomer	-	-	-	-	-
EGFR	P00533	3qwq_A 3w32_A	51	100/100/100/ 100/100/100	Monomer	-	-	-	-	-
TGFBR2	P37173	5e8y_A 1plo_A	53	99/99/100	Monomer	3kfd	0	0	0	0
BCL10	O95999	2mb9_A	45	100	Monomer	-	-	-	-	-
APP	P05067	3nyl_A 3ktm_C 1aap_A 2roz_A 2lp1_A	25	79/55/97/100/ 100/100	Monomer	-	-	-	-	-
ENG	P17813	5hzv_A	86	99	Monomer	-	-	-	-	-
MMP2	P08253	1gxd_A	94	100/100/100/ 100/100/100/ 100/100	Monomer, Hetero 2-mer	-	-	-	-	-
ITGB6	P18564	5ffo_B	87	100/100/100/0	Monomer	5ffo	0	1	0	0
LCE1C	Q5T751	-	-	-	-	-	-	-	-	-
WFIKKN2	Q8TEU8	-	-	-	-	-	-	-	-	-
ANGPTL4	Q9BY76	-	-	-	-	-	-	-	-	-
KLK5	Q9Y337	-	-	-	-	-	-	-	-	-
CYSRT1	A8MQ03	-	-	-	-	-	-	-	-	-
FCHO1	O14526	-	-	-	-	-	-	-	-	-
MCC	P23508	-	-	-	-	-	-	-	-	-
KRTAP10-8	P60410	-	-	-	-	-	-	-	-	-
KRTAP1-1	Q07627	-	-	-	-	-	-	-	-	-
LRRC32	Q14392	-	-	-	-	-	-	-	-	-
DIP2A	Q14689	-	-	-	-	-	-	-	-	-
PDGFRL	Q15198	-	-	-	-	-	-	-	-	-
GREB1	Q4ZG55	-	-	-	-	-	-	-	-	-
LCE1D	Q5T752	-	-	-	-	-	-	-	-	-
LCE1A	Q5T7P2	-	-	-	-	-	-	-	-	-
LCE3A	Q5TA76	-	-	-	-	-	-	-	-	-
LCE2C	Q5TA81	-	-	-	-	-	-	-	-	-
ADAMTSL4	Q6UY14	-	-	-	-	-	-	-	-	-
BLID	Q8IZY5	-	-	-	-	-	-	-	-	-
CATSPER1	Q8NEC5	-	-	-	-	-	-	-	-	-
SCGB3A1	Q96QR1	-	-	-	-	-	-	-	-	-
KRTAP4-11	Q9BYQ6	-	-	-	-	-	-	-	-	-
BCAS3	Q9H6U6	-	-	-	-	-	-	-	-	-
PSMC3IP	Q9P2W1	-	-	-	-	-	-	-	-	-
MLH3	Q9UHC1	-	-	-	-	-	-	-	-	-
LCE2B	O14633	-	-	-	-	-	-	-	-	-
All PPIs							*2*	*7*	*3*	*1*
TSHR	**P16473**	2xwt_C	31	0/100/100	Monomer	N/A				

SCRIB	Q14160	4wyt_A 1x5q_A 1wha_A	12	100/100/100 /100/0/0/0/ 100	Monomer	-	-	-	-	-
RABAC1	Q9UI14	-	-	-	-	-	-	-	-	-

[1] Gene name and UniProt code for the proteins associated to diseases detected in newborn screeening (first line, in bold) and for their interacting partners as provided by Interactome3D database (rest of lines).

[2] PDB codes and chain IDs corresponding to the available 3D structure of the individual proteins analyzed here and to their interacting partners (in parenthesis, the template of the homology model provided by Interactome3D, if no experimental structure is available).

[3] Global structural coverage: percentage of the sequence of each individual protein that has available 3D structure in the corresponding PDB files.

[4] Domain structural coverage: structural coverage (%) of each of the individual domains annotated for each individual protein.

[5] Oligomeric state for each individual protein as indicated in the biological unit in the PDB.

[6] PDB code of the structure of the complex formed by the individual proteins analyzed here with each interacting partner, as provided by Interactome3D. The first line corresponding to each individual protein analyzed here refers to the self-interactions (generated by the oligomeric state indicated in the biological unit).

[7] Mutated residues (involved in neutral or pathogenic mutations) that are located in each interface region as indicated by the complex structure. In the case of homo oligomers, a mutated residue could be counted in different regions if so appears in the different subunits.

[8] Global summary of the location of mutations in all the interactions for a given protein. Each mutated residue is counted only at one region.

Figure 8.2.1 Validation of predicted binding affinity changes upon mutation in SKEMPIv2.0. Destabilizing variants.

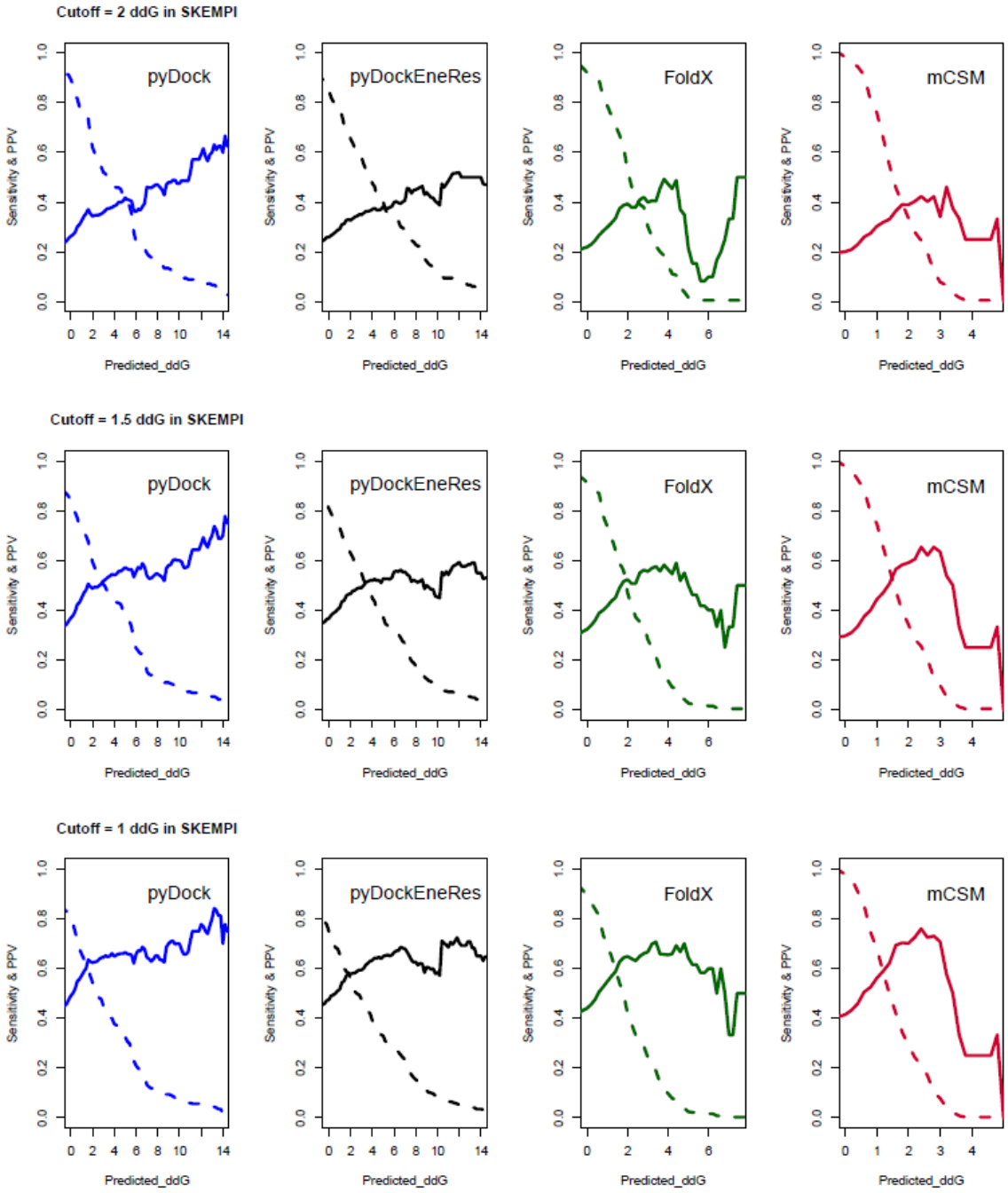


Figure 8.2.2 Validation of predicted binding affinity changes upon mutation in SKEMPIv2.0. Stabilizing variants.

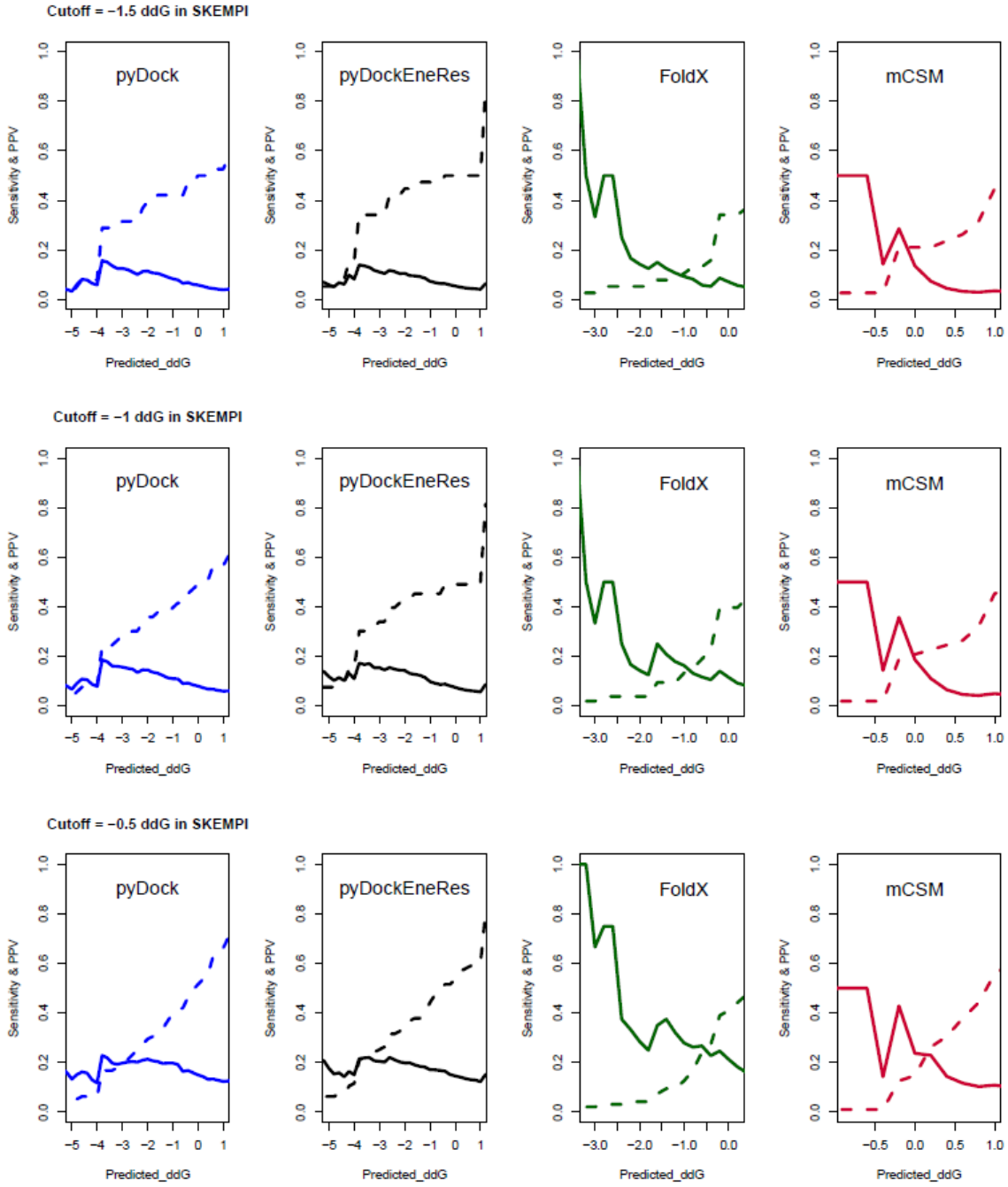


Table 8.2.3 Proteins from protein-protein docking benchmark version 4.0 with known pathological and unclassified mutations at the interface.

PDB	UniProts[1]	Biological Unit[2]	Unclassified mutations	Disease mutations
1B6C	P36897-P62942	Hetero-2-mer	11	1
1BKD	P01112-Q07889	Hetero-8-mer	5	12
1BUH	P24941-P61024	Hetero-2-mer	3	0
1DE4	Q30201-P61769-P02786	Hetero-6-mer	21	2
1E4K	P01857-O75015	Hetero-3-mer	0	0
1E96	P63000-P19878	Hetero-2-mer	12	2
1EER	P01588-P19235	Hetero-3-mer	21	1
1FQ1	Q16667-P24941	Hetero-2-mer	9	1
1GHQ	P01024-P20023	Hetero-3-mer	6	0
1GRN	Q07960-P60953	Hetero-2-mer	18	1
1GXD	P08253-P16035	Hetero-2-mer	16	0
1H9D	Q01196-Q13951	Hetero-2-mer	2	0
1HCF	P34130-Q16620	Hetero-4-mer	62	0
1HE8	P48736-P23175	Hetero-2-mer	0	0
1HIV	P06396-P68135	Hetero-2-mer	13	0
1I2M	P62826-P18754	Hetero-2-mer	1	0
1I4D	P53365-P63000	Hetero-3-mer	10	0
1IBR	Q14974-P62826	Hetero-2-mer	0	0
1IQD[a]	P00451	Hetero-3-mer	8	0
1RLB[b]	P02766	Hetero-6-mer	9	27
1IRA	P18510-P14778	Hetero-2-mer	13	0
1JMO	P00734-P00734-P05546	Hetero-3-mer	0	1
1JPS[a]	P13726	Hetero-3-mer	14	0
1JWH	P68400-P67870	Hetero-4-mer	3	2
1K74	P19793-P37231-Q15788	Hetero-4-mer	6	0
1KTZ	P10600-P37173	Hetero-4-mer	2	0
1M10	P04275-P07359	Hetero-2-mer	31	5
1MQ8	P05362-P20701	Hetero-2-mer	11	0
1NW9	P98170-P55211	Hetero-2-mer	12	0
1OC0	P05121-P04004	Hetero-2-mer	0	0
1PVH	P40189-P15018	Hetero-2-mer	14	0
1RV6	P17948-P49763	Hetero-4-mer	13	0
1S1Q	Q99816-P0CG48	Hetero-2-mer	0	0
1SYX	P83876-O95400	Hetero-2-mer	13	0
1WQ1	P01112-P20936	Hetero-2-mer	4	12
1XD3	P15374-P0CG48	Hetero-2-mer	0	0
1XQS	Q9NZL4-P0DMV8	Hetero-2-mer	0	0
1Z0K	P20338-Q9H1K0	Hetero-2-mer	0	0
2AYO	P54578-P0CG48	Hetero-2-mer	0	0
2C0L	P50542-P22307	Hetero-2-mer	2	0
2CFH	O43617-Q86SZ2	Hetero-2-mer	0	0
2FJU	P63000-Q00722	Hetero-2-mer	9	0
2HLE	P54760-P52799	Hetero-2-mer	5	0
2I9B	Q03405-P00749	Hetero-2-mer	21	0
2J0T	P03956-P01033	Hetero-2-mer	10	0
2NZ8	P63000-O75962	Hetero-2-mer	0	0
2OT3	Q9UJ41-Q9UL25	Hetero-2-mer	0	0
1AKJ	P61769-P01732-P04439	Hetero-5-mer	29	0
1KAC	P36711-P78310	Hetero-6-mer	8	0
1AK4	P62937-P12497	Hetero-2-mer	0	0
1EFN	P06241-P03406	Hetero-4-mer	4	0
1ML0	O41925-P13500	Hetero-4-mer	0	0
2AJF	Q9BYF1-P59594	Hetero-2-mer	6	0
2B4J	P12497-O75475	Hetero-4-mer	0	0

[1] UniProt codes of proteins involved for each protein complex. [a] The PDB structure has a human antibody interacting. [b] PDB 1RLB has treated individually as self-assembly (homo-4-mer).

[2] Biological Unit assigned of each PDB complex.

Table 8.2.4 Predicted binding affinity changes upon mutations in proteins related to diseases detected by newborn screening.

Gene	Uniprot-Partner_SAV	Location	Type	FoldX	mCSM	pyDockEneRes	pyDock	Biological Unit
ACADM	P11310_IA396V	CORE	neutral	0.8	1.1	0.5	0.3	Homo-4-mer
ACADM	P11310_IB396V	CORE	neutral	1.3	1.2	0.5	0.3	Homo-4-mer
ACADM	P11310_IC396V	CORE	neutral	1.1	1.1	0.4	0.3	Homo-4-mer
ACADM	P11310_ID396V	CORE	neutral	0.8	1.2	0.5	0.3	Homo-4-mer
ACADS	P16219_LE406V	CORE	neutral	1.6	0.7	1.9	1.3	Homo-4-mer
ACADS	P16219_LF406V	CORE	neutral	2	0.9	1.2	1.2	Homo-4-mer
ACADS	P16219_LG406V	CORE	neutral	1.9	0.9	1.1	1	Homo-4-mer
ACADS	P16219_LH406V	CORE	neutral	2.1	0.8	1.4	1.4	Homo-4-mer
ACADS	P16219_VE291A	CORE	neutral	3.3	2	1.6	0.7	Homo-4-mer
ACADS	P16219_VF291A	CORE	neutral	2.8	1.8	0.7	0.7	Homo-4-mer
ACADS	P16219_VG291A	CORE	neutral	2.6	1.8	1.6	0.6	Homo-4-mer
ACADS	P16219_VH291A	CORE	neutral	2.8	1.9	0.9	0.7	Homo-4-mer
ACADSB	P45954_LA421I	CORE	neutral	2.7	2.1	0.1	1.3	Homo-4-mer
ACADSB	P45954_LB421I	CORE	neutral	2.9	2.1	-0.2	1.1	Homo-4-mer
ACADSB	P45954_LC421I	CORE	neutral	2.6	2.3	-0.2	1	Homo-4-mer
ACADSB	P45954_LD421I	CORE	neutral	2.8	2.1	0.1	1.3	Homo-4-mer
ACADVL	P49748_FA563A	CORE	neutral	3.8	1.7	6.8	6.1	Homo-2-mer
ACADVL	P49748_HA598Y	CORE	neutral	-3	0.9	-6.2	-5.7	Homo-2-mer
ACADVL	P49748_IA457V	CORE	neutral	1.2	1.5	1.9	0.4	Homo-2-mer
ACADVL	P49748_QA562E	CORE	neutral	-1.6	1	0.5	0.8	Homo-2-mer
ACADVL	P49748_QA566S	CORE	neutral	1.3	0.3	0.1	0.1	Homo-2-mer
ASL	P04424_AA308S	CORE	neutral	1.6	1	0.3	0.3	Homo-4-mer
ASL	P04424_AB308S	CORE	neutral	1.6	1	0.3	0.3	Homo-4-mer
ASL	P04424_AC308S	CORE	neutral	0.7	1	0.5	0.5	Homo-4-mer
ASL	P04424_AD308S	CORE	neutral	0.7	1	0.5	0.5	Homo-4-mer
ASL	P04424_EA189Q	CORE	neutral	0	2.3	0.7	1.1	Homo-4-mer
ASL	P04424_EB189Q	CORE	neutral	-0.3	2.3	0.7	1.1	Homo-4-mer
ASL	P04424_EC189Q	CORE	neutral	-0.2	2.3	1.4	1.7	Homo-4-mer
ASL	P04424_ED189Q	CORE	neutral	0	2.3	1.4	1.7	Homo-4-mer
ASL	P04424_GA309A	CORE	neutral	-0.1	1.6	-0.1	0	Homo-4-mer
ASL	P04424_GB309A	CORE	neutral	-0.1	1.6	-0.1	0	Homo-4-mer
ASL	P04424_GC309A	CORE	neutral	-0.1	1.5	-0.2	-0.1	Homo-4-mer
ASL	P04424_GD309A	CORE	neutral	-0.1	1.5	-0.2	-0.1	Homo-4-mer
ASL	P04424_IA225V	CORE	neutral	1	0.5	0.8	0.5	Homo-4-mer
ASL	P04424_IB225V	CORE	neutral	1.1	0.5	0.8	0.5	Homo-4-mer
ASL	P04424_IC225V	CORE	neutral	0.5	0.6	1.4	0.4	Homo-4-mer
ASL	P04424_ID225V	CORE	neutral	0.7	0.6	1.4	0.4	Homo-4-mer
ASL	P04424_MA284R	CORE	neutral	0.4	0.9	-2.7	-2.7	Homo-4-mer
ASL	P04424_MB284R	CORE	neutral	0.4	0.9	-2.7	-2.7	Homo-4-mer
ASL	P04424_MC284R	CORE	neutral	0.1	1.2	-0.1	-0.4	Homo-4-mer
ASL	P04424_MD284R	CORE	neutral	0.4	1.2	-0.1	-0.4	Homo-4-mer
ASL	P04424_TA265S	CORE	neutral	1.1	1	1.3	0.3	Homo-4-mer
ASL	P04424_TB265S	CORE	neutral	1.1	1	1.3	0.3	Homo-4-mer
ASL	P04424_TC265S	CORE	neutral	1	1.2	0.5	0.3	Homo-4-mer
ASL	P04424_TD265S	CORE	neutral	1.1	1.2	0.5	0.3	Homo-4-mer
ASS1	P00966_EA84Q	CORE	neutral	1.6	1.5	3.1	3	Homo-4-mer
ASS1	P00966_EB84Q	CORE	neutral	2.2	1.5	3.1	3	Homo-4-mer
ASS1	P00966_EC84Q	CORE	neutral	2.5	1.5	3	3	Homo-4-mer

ASS1	P00966_ED84Q	CORE	neutral	2.5	1.5	3	3	Homo-4-mer
ASS1	P00966_FA333Y	CORE	neutral	3.2	1.6	6.2	6.3	Homo-4-mer
ASS1	P00966_FB333Y	CORE	neutral	3.1	1.6	6.2	6.3	Homo-4-mer
ASS1	P00966_FC333Y	CORE	neutral	2.2	1.6	6.2	6.3	Homo-4-mer
ASS1	P00966_FD333Y	CORE	neutral	3.1	1.6	6.2	6.3	Homo-4-mer
ASS1	P00966_GA120L	CORE	neutral	0.6	0.9	0.5	1.8	Homo-4-mer
ASS1	P00966_GB120L	CORE	neutral	-0.3	0.9	0.5	1.8	Homo-4-mer
ASS1	P00966_GC120L	CORE	neutral	1	0.9	0.5	1.8	Homo-4-mer
ASS1	P00966_GD120L	CORE	neutral	-0.1	0.9	0.5	1.8	Homo-4-mer
ASS1	P00966_IA274V	CORE	neutral	1.3	0	0.1	0.1	Homo-4-mer
ASS1	P00966_IA309V	CORE	neutral	1.2	0.6	0.6	0.5	Homo-4-mer
ASS1	P00966_IA360V	CORE	neutral	1.3	0.8	0.1	0.1	Homo-4-mer
ASS1	P00966_IB274V	CORE	neutral	1.4	0	0.1	0.1	Homo-4-mer
ASS1	P00966_IB309V	CORE	neutral	1.1	0.6	0.6	0.5	Homo-4-mer
ASS1	P00966_IB360V	CORE	neutral	1.4	0.8	0.1	0.1	Homo-4-mer
ASS1	P00966_IC274V	CORE	neutral	1.1	0	0.1	0.1	Homo-4-mer
ASS1	P00966_IC309V	CORE	neutral	1.2	0.6	0.6	0.5	Homo-4-mer
ASS1	P00966_IC360V	CORE	neutral	1.4	0.8	0.1	0.1	Homo-4-mer
ASS1	P00966_ID274V	CORE	neutral	1.4	0	0.1	0.1	Homo-4-mer
ASS1	P00966_ID309V	CORE	neutral	1.2	0.6	0.6	0.5	Homo-4-mer
ASS1	P00966_ID360V	CORE	neutral	1.3	0.8	0.1	0.1	Homo-4-mer
ASS1	P00966_KA355R	CORE	neutral	2.8	1.1	-0.4	-0.7	Homo-4-mer
ASS1	P00966_KB355R	CORE	neutral	3.1	1.1	-0.4	-0.7	Homo-4-mer
ASS1	P00966_KC355R	CORE	neutral	2.2	1.1	-0.4	-0.7	Homo-4-mer
ASS1	P00966_KD355R	CORE	neutral	2.4	1.1	-0.4	-0.7	Homo-4-mer
ASS1	P00966_LA405V	CORE	neutral	1.8	0.7	-0.2	1	Homo-4-mer
ASS1	P00966_LB405V	CORE	neutral	1.8	0.7	-0.2	1	Homo-4-mer
ASS1	P00966_LC405V	CORE	neutral	1.7	0.7	-0.2	1	Homo-4-mer
ASS1	P00966_LD405V	CORE	neutral	1.7	0.7	-0.2	1	Homo-4-mer
ASS1	P00966_QA350H	CORE	neutral	1.1	0.2	-1.2	-1.4	Homo-4-mer
ASS1	P00966_QA350L	CORE	neutral	-1.8	0.6	0.7	0.4	Homo-4-mer
ASS1	P00966_QA350R	CORE	neutral	-0.5	1.3	10.6	11	Homo-4-mer
ASS1	P00966_QA350T	CORE	neutral	0.2	1	1.1	1.1	Homo-4-mer
ASS1	P00966_QB350H	CORE	neutral	1.2	0.2	-1.2	-1.4	Homo-4-mer
ASS1	P00966_QB350L	CORE	neutral	-0.9	0.6	0.7	0.4	Homo-4-mer
ASS1	P00966_QB350R	CORE	neutral	-0.3	1.3	10.6	11	Homo-4-mer
ASS1	P00966_QB350T	CORE	neutral	0.5	1	1.1	1.1	Homo-4-mer
ASS1	P00966_QC350H	CORE	neutral	-0.3	0.2	-1.2	-1.4	Homo-4-mer
ASS1	P00966_QC350L	CORE	neutral	-2.3	0.6	0.7	0.4	Homo-4-mer
ASS1	P00966_QC350R	CORE	neutral	-0.7	1.3	10.6	11	Homo-4-mer
ASS1	P00966_QC350T	CORE	neutral	-0.1	1	1.1	1.1	Homo-4-mer
ASS1	P00966_QD350H	CORE	neutral	0.6	0.2	-1.2	-1.4	Homo-4-mer
ASS1	P00966_QD350L	CORE	neutral	-1.7	0.6	0.7	0.4	Homo-4-mer
ASS1	P00966_QD350R	CORE	neutral	-1	1.3	10.6	11	Homo-4-mer
ASS1	P00966_QD350T	CORE	neutral	0.3	1	1.1	1.1	Homo-4-mer
ASS1	P00966_RA344H	CORE	neutral	1.7	0.3	-1.8	-1.5	Homo-4-mer
ASS1	P00966_RA86G	CORE	neutral	3.1	1	-1.6	-1.6	Homo-4-mer
ASS1	P00966_RB344H	CORE	neutral	1.6	0.3	-1.8	-1.5	Homo-4-mer
ASS1	P00966_RB86G	CORE	neutral	2.2	1	-1.6	-1.6	Homo-4-mer
ASS1	P00966_RC344H	CORE	neutral	2	0.3	-1.9	-1.5	Homo-4-mer
ASS1	P00966_RC86G	CORE	neutral	3.3	1	-1.6	-1.6	Homo-4-mer

ASS1	P00966_RD344H	CORE	neutral	1.4	0.3	-1.9	-1.5	Homo-4-mer
ASS1	P00966_RD86G	CORE	neutral	3.4	1	-1.5	-1.6	Homo-4-mer
BCKDHA	P12694_MA294L	CORE	neutral	0	0.7	0.2	-0.2	Homo-2-mer
BCKDHA	P12694-P21953_PA208T	CORE	neutral	2.5	0.8	1.1	1.1	Hetero-4-mer
BCKDHA	P12694-P21953_PB208T	CORE	neutral	2.5	0.8	1.1	1.1	Hetero-4-mer
CBS	P35520_VA90I	CORE	neutral	-0.4	1.5	-0.5	0.8	Homo-2-mer
ETFA	P13804-P38117_IA148V	CORE	neutral	0.5	1	0.8	0.8	Hetero-2-mer
ETFA	P13804-P38117_IA196M	CORE	neutral	-0.4	0.8	-1.3	-0.7	Hetero-2-mer
ETFA	P13804-P38117_IA196V	CORE	neutral	0.9	1.4	0.3	0.5	Hetero-2-mer
ETFA	P13804-P38117_LA154I	CORE	neutral	0.7	1.2	0.1	0	Hetero-2-mer
ETFA	P13804-P38117_LA154V	CORE	neutral	1.2	1.3	0.5	0.5	Hetero-2-mer
ETFA	P13804-P38117_SA197A	CORE	neutral	-1.4	0.5	0.3	0.3	Hetero-2-mer
ETFA	P13804-P38117_YA149S	CORE	neutral	5.5	3.6	11.1	10	Hetero-2-mer
ETFB	P13804-P38117_CB131R	CORE	neutral	-0.3	0.6	16.7	15.4	Hetero-2-mer
ETFB	P13804-P38117_DB128G	CORE	neutral	1.9	1.5	3.6	2.7	Hetero-2-mer
ETFB	P13804-P38117_FB141L	CORE	neutral	0.6	1.8	3.2	3.1	Hetero-2-mer
ETFB	P13804-P38117_IB127L	CORE	neutral	0.2	1	-0.7	-0.6	Hetero-2-mer
ETFB	P13804-P38117_IB166V	CORE	neutral	0.2	0.9	0.2	0.3	Hetero-2-mer
ETFB	P13804-P38117_LB170M	CORE	neutral	0.3	0.1	0	-0.2	Hetero-2-mer
ETFB	P13804-P38117_LB173V	CORE	neutral	0.5	1.1	1	0.8	Hetero-2-mer
ETFB	P13804-P38117_MB137A	CORE	neutral	2.7	0.8	1.8	1.6	Hetero-2-mer
ETFB	P13804-P38117_QB133R	CORE	neutral	2.5	1.8	-0.9	0.4	Hetero-2-mer
ETFB	P13804-P38117_QB136P	CORE	neutral	-0.6	1.6	-1	-0.9	Hetero-2-mer
ETFB	P13804-P38117_RB106Q	CORE	neutral	-0.2	1.2	0	0.5	Hetero-2-mer
ETFB	P13804-P38117_RB254C	CORE	neutral	1.3	1.6	-1.6	-0.5	Hetero-2-mer
GALT	P07902_SB121A	CORE	neutral	-0.3	0.5	0	0	Homo-2-mer
GCDH	Q92947_AA337E	CORE	neutral	5.1	0.9	-2.4	-2	Homo-4-mer
GCDH	Q92947_AA363T	CORE	neutral	1.7	0.4	2.2	1.6	Homo-4-mer
GCDH	Q92947_AA433S	CORE	neutral	1.8	0.9	1.6	0.7	Homo-4-mer
GCDH	Q92947_AB337E	CORE	neutral	6.3	0.9	-2.4	-2	Homo-4-mer
GCDH	Q92947_AB363T	CORE	neutral	1.9	0.4	2.2	1.6	Homo-4-mer
GCDH	Q92947_AB433S	CORE	neutral	1.8	0.9	1.6	0.7	Homo-4-mer
GCDH	Q92947_AC337E	CORE	neutral	6.6	0.9	-2.4	-2	Homo-4-mer
GCDH	Q92947_AC363T	CORE	neutral	1.4	0.4	2.3	1.8	Homo-4-mer
GCDH	Q92947_AC433S	CORE	neutral	1.8	0.9	1.6	0.7	Homo-4-mer
GCDH	Q92947_AD337E	CORE	neutral	3.3	0.9	-2.4	-2	Homo-4-mer
GCDH	Q92947_AD363T	CORE	neutral	1.4	0.4	2.2	1.6	Homo-4-mer
GCDH	Q92947_AD433S	CORE	neutral	1.8	0.9	1.6	0.7	Homo-4-mer
GCDH	Q92947_FA434L	CORE	neutral	1.7	2.3	0.4	0.4	Homo-4-mer
GCDH	Q92947_FB434L	CORE	neutral	1.3	2.3	0.4	0.4	Homo-4-mer
GCDH	Q92947_FC434L	CORE	neutral	1.4	2.3	0.4	0.4	Homo-4-mer
GCDH	Q92947_FD434L	CORE	neutral	1.6	2.3	0.4	0.4	Homo-4-mer
GCDH	Q92947_IA431W	CORE	neutral	0.7	-0.5	-3.5	-2.7	Homo-4-mer
GCDH	Q92947_IB431W	CORE	neutral	1.4	-0.5	-3.5	-2.7	Homo-4-mer
GCDH	Q92947_IC431W	CORE	neutral	0.7	-0.5	-3.5	-2.7	Homo-4-mer
GCDH	Q92947_ID431W	CORE	neutral	0.7	-0.5	-3.5	-2.7	Homo-4-mer
GCDH	Q92947_PA325S	CORE	neutral	4	0.2	0.9	0.8	Homo-4-mer
GCDH	Q92947_PB325S	CORE	neutral	4	0.2	0.9	0.8	Homo-4-mer
GCDH	Q92947_PC325S	CORE	neutral	4	0.2	0.9	0.8	Homo-4-mer
GCDH	Q92947_PD325S	CORE	neutral	4.2	0.2	0.9	0.8	Homo-4-mer
GCDH	Q92947_QA333L	CORE	neutral	-0.2	0.6	-1.5	-0.4	Homo-4-mer

GCDH	Q92947_QB333L	CORE	neutral	-0.2	0.6	-1.5	-0.4	Homo-4-mer
GCDH	Q92947_QC333L	CORE	neutral	0.1	0.6	-1.5	-0.4	Homo-4-mer
GCDH	Q92947_QD333L	CORE	neutral	0.2	0.6	-1.5	-0.4	Homo-4-mer
GCDH	Q92947_TA416V	CORE	neutral	1.2	0.2	0.3	0.3	Homo-4-mer
GCDH	Q92947_TB341I	CORE	neutral	-2	0.6	0.3	-1.6	Homo-4-mer
GCDH	Q92947_TB416V	CORE	neutral	2.1	0.2	0.3	0.3	Homo-4-mer
GCDH	Q92947_TC341I	CORE	neutral	-2.2	0.6	0.3	-1.6	Homo-4-mer
GCDH	Q92947_TC416V	CORE	neutral	1.3	0.2	0.3	0.3	Homo-4-mer
GCDH	Q92947_TD341I	CORE	neutral	-1.3	0.6	0.3	-1.6	Homo-4-mer
GCDH	Q92947_TD416V	CORE	neutral	1.3	0.2	0.2	0.2	Homo-4-mer
HADHA	P40939-HADHB_DG398G	CORE	neutral	-2.1	2.9	4.2	3.2	Hetero-3-mer
HADHA	P40939-HADHB_GG226R	CORE	neutral	10	1.6	10	10.4	Hetero-3-mer
HADHA	P40939-HADHB_LG221I	CORE	neutral	2.4	0.1	0	0	Hetero-3-mer
HADHA	P40939-HADHB_LG225A	CORE	neutral	3.2	1.1	1.3	0.9	Hetero-3-mer
HADHA	P40939-HADHB_QG220A	CORE	neutral	0.7	1.3	-0.6	-0.8	Hetero-3-mer
HADHA	P40939-HADHB_VG222L	CORE	neutral	-1.5	0.7	0.4	0.4	Hetero-3-mer
HBB	P68871-HBA_HB117Q	CORE	neutral	0.7	1.2	3.5	5.2	Hetero-4-mer
HBB	P68871-HBA_HD117Q	CORE	neutral	0.3	1.2	0.8	0.8	Hetero-4-mer
HBB	P68871-HBA_YB146F	CORE	neutral	0.3	1	1.4	0.5	Hetero-4-mer
HBB	P68871-HBA_YD146F	CORE	neutral	0.4	0.7	1.6	0.6	Hetero-4-mer
HBB	P68871-HBAZ_HB117Q	CORE	neutral	0.3	1.2	1.6	1.5	Hetero-4-mer
HBB	P68871-HBAZ_HD117Q	CORE	neutral	0.1	0.8	1.7	1.6	Hetero-4-mer
HBB	P68871-HBAZ_YB146F	CORE	neutral	0.1	0.6	1.8	0.7	Hetero-4-mer
HBB	P68871-HBAZ_YD146F	CORE	neutral	0.3	0.8	1.9	0.8	Hetero-4-mer
HBB	P68871-HP_YL146F	CORE	neutral	0.6	1	1.7	0.7	Hetero-2-mer
HCFC1	P51610_AD1934T	CORE	neutral	-0.1	-0.8	0	0	Hetero-4-mer
HCFC1	P51610_GD1935A	CORE	neutral	3.6	0.2	0	0	Hetero-4-mer
HCFC1	P51610_GD1936S	CORE	neutral	3.1	-0.6	0	0	Hetero-4-mer
HCFC1	P51610_SD1932T	CORE	neutral	0.2	0.1	0	0	Hetero-4-mer
HCFC1	P51610_VA382A	CORE	neutral	2.8	0.9	2.9	1.3	Hetero-4-mer
HCFC1	P51610_VC382A	CORE	neutral	3.1	1.1	1.6	0.4	Hetero-4-mer
HPD	P32754_VA87A	CORE	neutral	1.5	1.4	1.5	1.3	Homo-2-mer
IVD	P26440_KA345R	CORE	neutral	0	0.8	-0.1	0	Homo-4-mer
IVD	P26440_KB345R	CORE	neutral	-0.6	0.7	1.7	1.7	Homo-4-mer
IVD	P26440_KC345R	CORE	neutral	0.1	0.8	0.8	0.9	Homo-4-mer
IVD	P26440_KD345R	CORE	neutral	0.4	0.8	0.4	0.5	Homo-4-mer
MLYCD	O95822_DA174H	CORE	neutral	0.3	-0.4	0.7	1.3	Homo-4-mer
MLYCD	O95822_DC174H	CORE	neutral	0.2	-0.4	1.2	0.8	Homo-4-mer
MLYCD	O95822_DD174H	CORE	neutral	-0.8	-0.2	1.2	1.2	Homo-4-mer
MMAA	Q8IVH4_FA414L	CORE	neutral	0.8	2.8	9.5	7.7	Homo-2-mer
MMADHC	Q9H3L0_AA137V	CORE	neutral	0.2	1	-0.2	0.5	Homo-2-mer
MMADHC	Q9H3L0_AA217T	CORE	neutral	1	0.7	3.5	1.4	Homo-2-mer
MMADHC	Q9H3L0_SA136R	CORE	neutral	-0.6	1.8	-0.7	-0.6	Homo-2-mer
MUT	P22033_AA371T	CORE	neutral	-0.4	1.5	-2.7	-0.1	Homo-2-mer
MUT	P22033_IA372V	CORE	neutral	1.4	1.1	0.5	0.5	Homo-2-mer
MUT	P22033_LA81F	CORE	neutral	-0.9	-0.3	-1.2	-1.2	Homo-2-mer
MUT	P22033_VA368I	CORE	neutral	-1.1	1.5	0.6	0.6	Homo-2-mer
MUT	P22033_VA485I	CORE	neutral	0.1	0.7	2.3	0.2	Homo-2-mer
MUT	P22033_VA88M	CORE	neutral	-0.3	0.3	1.3	0.7	Homo-2-mer
SLC24A13	Q9UJS0_VA255A	CORE	neutral	0.9	1.5	1.5	0.3	Homo-2-mer
TAT	P17735_DA149E	CORE	neutral	-0.4	1.2	0.1	0.4	Homo-2-mer

TAT	P17735_KA98R	CORE	neutral	-0.2	2	1.8	2.6	Homo-2-mer
TGFB1	P01137_VD77I	CORE	neutral	-1	1.1	-0.4	-0.3	Homo-2-mer
TGFB1	P01137-ITGAV-ITGB6_TC240S	CORE	neutral	0.1	0.9	-2	0.5	Hetero-5-mer
ACADM	P11310_KA301R	RIM	neutral	-0.1	1.2	-0.9	-0.9	Homo-4-mer
ACADM	P11310_KB301R	RIM	neutral	0.1	1.1	0.3	0.3	Homo-4-mer
ACADM	P11310_KC301R	RIM	neutral	0.2	1	0.2	0.2	Homo-4-mer
ACADM	P11310_KD301R	RIM	neutral	0.2	1	0.3	0.3	Homo-4-mer
ACADM	P11310_NB421S	RIM	neutral	0.1	-0.4	0	0	Homo-4-mer
ACADM	P11310_ND421S	RIM	neutral	1.2	0.2	1	0.9	Homo-4-mer
ACADM	P11310_SA167P	RIM	neutral	-2.6	-0.4	-0.9	-0.5	Homo-4-mer
ACADM	P11310_SB167P	RIM	neutral	-2.3	-0.6	-1	-0.5	Homo-4-mer
ACADM	P11310_SC167P	RIM	neutral	-2.6	-0.4	-1	-0.5	Homo-4-mer
ACADM	P11310_SD167P	RIM	neutral	-2.4	-0.5	-0.9	-0.5	Homo-4-mer
ACADSB	P45954_IA320M	RIM	neutral	0.3	0.4	0	0	Homo-4-mer
ACADSB	P45954_IB320M	RIM	neutral	0.4	0.4	0	0	Homo-4-mer
ACADSB	P45954_IC320M	RIM	neutral	0.4	0.4	0	0	Homo-4-mer
ACADSB	P45954_ID320M	RIM	neutral	0.4	0.4	0	0	Homo-4-mer
ACADVL	P49748_AA357T	RIM	neutral	-0.4	1.5	-0.4	0	Homo-2-mer
ACADVL	P49748_LA653I	RIM	neutral	1	1.2	1.2	0.9	Homo-2-mer
ACADVL	P49748_PA219S	RIM	neutral	2.9	0	0.1	0.1	Homo-2-mer
ACADVL	P49748_RA531L	RIM	neutral	-0.8	1.1	2.9	-0.5	Homo-2-mer
ACADVL	P49748_SA650N	RIM	neutral	-0.6	1.1	0.2	0.1	Homo-2-mer
ACADVL	P49748_TA364A	RIM	neutral	-0.1	0.5	0.6	0	Homo-2-mer
ACADVL	P49748_TA367I	RIM	neutral	0.3	0.1	0	0	Homo-2-mer
ACADVL	P49748_WA626G	RIM	neutral	1.6	1.1	6	6.7	Homo-2-mer
ACADVL	P49748_WA626Q	RIM	neutral	0.9	1.2	5.2	7.1	Homo-2-mer
ACADVL	P49748_WA626R	RIM	neutral	0.1	1.2	12.8	13.6	Homo-2-mer
ACAT1	P24752_DA299G	RIM	neutral	0.9	0.3	-0.1	-0.1	Homo-4-mer
ACAT1	P24752_DA299S	RIM	neutral	0.7	0.9	-0.1	-0.1	Homo-4-mer
ACAT1	P24752_DB299G	RIM	neutral	1.7	0.3	-0.1	-0.1	Homo-4-mer
ACAT1	P24752_DB299S	RIM	neutral	1.6	0.8	-0.1	-0.1	Homo-4-mer
ACAT1	P24752_DC299G	RIM	neutral	1	0.3	0.9	0.9	Homo-4-mer
ACAT1	P24752_DC299S	RIM	neutral	1	0.9	0.9	0.9	Homo-4-mer
ACAT1	P24752_DD299G	RIM	neutral	1.1	0.3	1.8	1.8	Homo-4-mer
ACAT1	P24752_DD299S	RIM	neutral	1	0.9	1.9	1.9	Homo-4-mer
ACAT1	P24752_IA115V	RIM	neutral	0.8	1.4	0.4	0.1	Homo-4-mer
ACAT1	P24752_IB115V	RIM	neutral	0.9	1.5	0.3	0.2	Homo-4-mer
ACAT1	P24752_IC115V	RIM	neutral	0.9	1.4	0.2	0.2	Homo-4-mer
ACAT1	P24752_ID115V	RIM	neutral	0.4	1.4	-0.2	0	Homo-4-mer
ARG1	P05089_RA205K	RIM	neutral	0.8	1	-5.4	-3	Homo-3-mer
ARG1	P05089_RB205K	RIM	neutral	0.8	1	-5.4	-3	Homo-3-mer
ARG1	P05089_RC205K	RIM	neutral	0.5	1	-5.4	-3	Homo-3-mer
ARG1	P05089_TA253S	RIM	neutral	-0.4	0.6	0.2	0.2	Homo-3-mer
ARG1	P05089_TB253S	RIM	neutral	-0.5	0.6	0.2	0.2	Homo-3-mer
ARG1	P05089_TC253S	RIM	neutral	-0.4	0.6	0.2	0.2	Homo-3-mer
ASL	P04424_AA102E	RIM	neutral	-0.7	-0.1	-1.5	-1.8	Homo-4-mer
ASL	P04424_AA16T	RIM	neutral	0.2	-0.5	-7.1	-6.9	Homo-4-mer
ASL	P04424_AB102E	RIM	neutral	-0.7	-0.1	-1.5	-1.8	Homo-4-mer
ASL	P04424_AB16T	RIM	neutral	0.2	-0.5	-7.1	-6.9	Homo-4-mer
ASL	P04424_AC102E	RIM	neutral	-0.4	-0.3	-0.6	-0.6	Homo-4-mer

ASL	P04424_AC16T	RIM	neutral	-0.1	-0.3	-6.9	-6.8	Homo-4-mer
ASL	P04424_AD102E	RIM	neutral	-0.5	-0.3	-0.7	-0.7	Homo-4-mer
ASL	P04424_AD16T	RIM	neutral	-0.1	-0.3	-6.9	-6.8	Homo-4-mer
ASL	P04424_EA138A	RIM	neutral	0.3	0.1	-0.4	-0.4	Homo-4-mer
ASL	P04424_EB138A	RIM	neutral	0.4	0.1	-0.4	-0.4	Homo-4-mer
ASL	P04424_EC138A	RIM	neutral	0.1	0.1	0.6	0.8	Homo-4-mer
ASL	P04424_ED138A	RIM	neutral	0.1	0.1	0.6	0.8	Homo-4-mer
ASL	P04424_GA210S	RIM	neutral	2	-0.1	0	0	Homo-4-mer
ASL	P04424_GA438A	RIM	neutral	-0.2	0.1	-0.1	-0.1	Homo-4-mer
ASL	P04424_GA438S	RIM	neutral	-0.6	-0.3	0.4	0.4	Homo-4-mer
ASL	P04424_GA438T	RIM	neutral	0.2	-0.3	0.4	0.4	Homo-4-mer
ASL	P04424_GB210S	RIM	neutral	2.1	-0.1	0	0	Homo-4-mer
ASL	P04424_GB438A	RIM	neutral	-0.2	0.1	-0.1	-0.1	Homo-4-mer
ASL	P04424_GB438S	RIM	neutral	-0.5	-0.3	0.4	0.4	Homo-4-mer
ASL	P04424_GB438T	RIM	neutral	0.3	-0.3	0.4	0.4	Homo-4-mer
ASL	P04424_GC210S	RIM	neutral	1.7	0	0.1	0.2	Homo-4-mer
ASL	P04424_GC438A	RIM	neutral	-0.2	0	-0.1	-0.1	Homo-4-mer
ASL	P04424_GC438S	RIM	neutral	-0.6	-0.4	0	0.4	Homo-4-mer
ASL	P04424_GC438T	RIM	neutral	0.2	-0.4	0	0.5	Homo-4-mer
ASL	P04424_GD210S	RIM	neutral	1.8	0	0.1	0.2	Homo-4-mer
ASL	P04424_GD438A	RIM	neutral	-0.2	0	-0.1	-0.1	Homo-4-mer
ASL	P04424_GD438S	RIM	neutral	-0.6	-0.4	0	0.4	Homo-4-mer
ASL	P04424_GD438T	RIM	neutral	0.2	-0.4	0	0.5	Homo-4-mer
ASL	P04424_RA141K	RIM	neutral	-0.2	0.1	0.1	0.1	Homo-4-mer
ASL	P04424_RA379H	RIM	neutral	2.4	0.2	0.8	0.8	Homo-4-mer
ASL	P04424_RB141K	RIM	neutral	-0.2	0.1	0.1	0.1	Homo-4-mer
ASL	P04424_RB379H	RIM	neutral	2.5	0.2	0.8	0.8	Homo-4-mer
ASL	P04424_RC141K	RIM	neutral	-0.4	0.4	-3.6	-3.4	Homo-4-mer
ASL	P04424_RC379H	RIM	neutral	0.7	0	-3.4	-2.4	Homo-4-mer
ASL	P04424_RD141K	RIM	neutral	-0.4	0.4	-3.6	-3.4	Homo-4-mer
ASL	P04424_RD379H	RIM	neutral	0.6	0	-3.4	-2.4	Homo-4-mer
ASS1	P00966_AA214S	RIM	neutral	-0.3	0.2	0.1	0.1	Homo-4-mer
ASS1	P00966_AA214T	RIM	neutral	0.3	0.2	-1.3	-0.6	Homo-4-mer
ASS1	P00966_AB214S	RIM	neutral	-0.8	0.2	0.1	0.1	Homo-4-mer
ASS1	P00966_AB214T	RIM	neutral	-0.1	0.2	-1.3	-0.6	Homo-4-mer
ASS1	P00966_AC214S	RIM	neutral	-0.4	0.2	0.1	0.1	Homo-4-mer
ASS1	P00966_AC214T	RIM	neutral	0.2	0.2	-1.3	-0.6	Homo-4-mer
ASS1	P00966_AD214S	RIM	neutral	-0.4	0.2	0.1	0.1	Homo-4-mer
ASS1	P00966_AD214T	RIM	neutral	0.1	0.2	-1.3	-0.6	Homo-4-mer
ASS1	P00966_DA382N	RIM	neutral	0.5	1.3	3.2	3.2	Homo-4-mer
ASS1	P00966_DB382N	RIM	neutral	-0.2	1.3	3.2	3.2	Homo-4-mer
ASS1	P00966_DC382N	RIM	neutral	0.2	1.3	3.2	3.2	Homo-4-mer
ASS1	P00966_DD382N	RIM	neutral	0	1.3	3.2	3.3	Homo-4-mer
ASS1	P00966_HA336Y	RIM	neutral	-0.7	0.1	1.1	-1.8	Homo-4-mer
ASS1	P00966_HB336Y	RIM	neutral	-0.4	0.1	1.1	-1.8	Homo-4-mer
ASS1	P00966_HC336Y	RIM	neutral	-0.7	0.1	1.1	-1.8	Homo-4-mer
ASS1	P00966_HD336Y	RIM	neutral	-0.9	0.1	1.1	-1.8	Homo-4-mer
ASS1	P00966_KA155Q	RIM	neutral	0.1	-0.1	-0.3	-0.2	Homo-4-mer
ASS1	P00966_KA155R	RIM	neutral	-0.2	-0.1	5	4.2	Homo-4-mer
ASS1	P00966_KA308R	RIM	neutral	-1.7	0.9	-1.5	-1.4	Homo-4-mer
ASS1	P00966_KB155Q	RIM	neutral	0.1	-0.1	-0.3	-0.2	Homo-4-mer

ASS1	P00966_KB155R	RIM	neutral	0	-0.1	5	4.2	Homo-4-mer
ASS1	P00966_KB308R	RIM	neutral	-1.4	0.9	-1.5	-1.4	Homo-4-mer
ASS1	P00966_KC155Q	RIM	neutral	0.3	-0.1	-0.2	-0.2	Homo-4-mer
ASS1	P00966_KC155R	RIM	neutral	0.2	-0.1	5	4.2	Homo-4-mer
ASS1	P00966_KC308R	RIM	neutral	-2.3	0.9	-1.5	-1.4	Homo-4-mer
ASS1	P00966_KD155Q	RIM	neutral	0.2	-0.1	-0.2	-0.2	Homo-4-mer
ASS1	P00966_KD155R	RIM	neutral	-0.3	-0.1	5	4.2	Homo-4-mer
ASS1	P00966_KD308R	RIM	neutral	-0.3	0.9	-1.5	-1.4	Homo-4-mer
ASS1	P00966_NA200S	RIM	neutral	0.1	0	0.2	0.2	Homo-4-mer
ASS1	P00966_NA218D	RIM	neutral	-0.4	-0.1	-0.9	-0.9	Homo-4-mer
ASS1	P00966_NA218S	RIM	neutral	0	0	0.5	0.5	Homo-4-mer
ASS1	P00966_NA378S	RIM	neutral	0.5	-0.5	-1.1	-0.6	Homo-4-mer
ASS1	P00966_NB200S	RIM	neutral	-0.1	0	0.2	0.2	Homo-4-mer
ASS1	P00966_NB218D	RIM	neutral	-0.4	-0.1	-0.9	-0.9	Homo-4-mer
ASS1	P00966_NB218S	RIM	neutral	0.1	0	0.5	0.5	Homo-4-mer
ASS1	P00966_NB378S	RIM	neutral	0.2	-0.5	-1.1	-0.6	Homo-4-mer
ASS1	P00966_NC200S	RIM	neutral	0	0	0.2	0.2	Homo-4-mer
ASS1	P00966_NC218D	RIM	neutral	-0.4	-0.1	-0.9	-0.9	Homo-4-mer
ASS1	P00966_NC218S	RIM	neutral	-0.1	0	0.5	0.5	Homo-4-mer
ASS1	P00966_NC378S	RIM	neutral	0.2	-0.5	-1.1	-0.6	Homo-4-mer
ASS1	P00966_ND200S	RIM	neutral	0.7	0	0.2	0.2	Homo-4-mer
ASS1	P00966_ND218D	RIM	neutral	-0.5	-0.1	-0.9	-0.9	Homo-4-mer
ASS1	P00966_ND218S	RIM	neutral	-0.1	0	0.5	0.5	Homo-4-mer
ASS1	P00966_ND378S	RIM	neutral	0.1	-0.5	-1.1	-0.6	Homo-4-mer
ASS1	P00966_QA201H	RIM	neutral	0.1	0	0	-1.5	Homo-4-mer
ASS1	P00966_QA201P	RIM	neutral	-0.6	0.4	1.1	0.5	Homo-4-mer
ASS1	P00966_QB201H	RIM	neutral	0.2	0	0	-1.5	Homo-4-mer
ASS1	P00966_QB201P	RIM	neutral	-0.9	0.4	1.1	0.5	Homo-4-mer
ASS1	P00966_QC201H	RIM	neutral	0.2	0	0	-1.5	Homo-4-mer
ASS1	P00966_QC201P	RIM	neutral	-0.2	0.4	1.1	0.5	Homo-4-mer
ASS1	P00966_QD201H	RIM	neutral	0.2	0	0	-1.5	Homo-4-mer
ASS1	P00966_QD201P	RIM	neutral	0	0.4	1.1	0.5	Homo-4-mer
ASS1	P00966_RA157C	RIM	neutral	1.3	0.5	-0.6	-0.3	Homo-4-mer
ASS1	P00966_RB157C	RIM	neutral	0.9	0.5	-0.6	-0.3	Homo-4-mer
ASS1	P00966_RC157C	RIM	neutral	0.5	0.5	-0.6	-0.3	Homo-4-mer
ASS1	P00966_RD157C	RIM	neutral	0.5	0.5	-0.6	-0.3	Homo-4-mer
ASS1	P00966_TA386G	RIM	neutral	-0.1	0.3	-0.1	-0.3	Homo-4-mer
ASS1	P00966_TA386I	RIM	neutral	-0.3	0.5	-1.7	-1.4	Homo-4-mer
ASS1	P00966_TA386L	RIM	neutral	-1.8	0.5	-2.3	-2.1	Homo-4-mer
ASS1	P00966_TA386M	RIM	neutral	-1.7	0.2	-1.7	-1.4	Homo-4-mer
ASS1	P00966_TA386S	RIM	neutral	-0.3	0.7	-0.8	-0.9	Homo-4-mer
ASS1	P00966_TA386V	RIM	neutral	0	0.4	-1.5	-1.4	Homo-4-mer
ASS1	P00966_TB386G	RIM	neutral	-0.1	0.3	-0.1	-0.3	Homo-4-mer
ASS1	P00966_TB386I	RIM	neutral	-0.3	0.5	-1.7	-1.4	Homo-4-mer
ASS1	P00966_TB386L	RIM	neutral	-1.8	0.5	-2.3	-2.1	Homo-4-mer
ASS1	P00966_TB386M	RIM	neutral	-1.7	0.2	-1.7	-1.4	Homo-4-mer
ASS1	P00966_TB386S	RIM	neutral	-0.2	0.7	-0.8	-0.9	Homo-4-mer
ASS1	P00966_TB386V	RIM	neutral	0	0.4	-1.5	-1.4	Homo-4-mer
ASS1	P00966_TC386G	RIM	neutral	0.2	0.3	-0.1	-0.3	Homo-4-mer
ASS1	P00966_TC386I	RIM	neutral	-0.1	0.5	-1.7	-1.4	Homo-4-mer
ASS1	P00966_TC386L	RIM	neutral	-1.6	0.5	-2.3	-2.1	Homo-4-mer

ASS1	P00966_TC386M	RIM	neutral	-1.5	0.2	-1.7	-1.4	Homo-4-mer
ASS1	P00966_TC386S	RIM	neutral	-0.1	0.7	-0.8	-0.9	Homo-4-mer
ASS1	P00966_TC386V	RIM	neutral	0.2	0.4	-1.5	-1.4	Homo-4-mer
ASS1	P00966_TD386G	RIM	neutral	0.1	0.3	-0.1	-0.3	Homo-4-mer
ASS1	P00966_TD386I	RIM	neutral	-0.3	0.5	-1.7	-1.4	Homo-4-mer
ASS1	P00966_TD386L	RIM	neutral	-1.7	0.5	-2.3	-2.1	Homo-4-mer
ASS1	P00966_TD386M	RIM	neutral	-1.7	0.2	-1.7	-1.4	Homo-4-mer
ASS1	P00966_TD386S	RIM	neutral	-0.3	0.7	-0.8	-0.9	Homo-4-mer
ASS1	P00966_TD386V	RIM	neutral	0.1	0.4	-1.5	-1.4	Homo-4-mer
ASS1	P00966_YA402F	RIM	neutral	-0.1	0.3	6	4.8	Homo-4-mer
ASS1	P00966_YB402F	RIM	neutral	-0.3	0.3	6	4.8	Homo-4-mer
ASS1	P00966_YC402F	RIM	neutral	-0.5	0.3	6	4.8	Homo-4-mer
ASS1	P00966_YD402F	RIM	neutral	-0.3	0.3	6	4.8	Homo-4-mer
BCKDHA	P12694_IA89V	RIM	neutral	-0.1	0.5	0.1	0.1	Homo-2-mer
BCKDHA	P12694-P21953_QA419R	RIM	neutral	-1.9	0.6	0.1	0.3	Hetero-4-mer
BCKDHA	P12694-P21953_QB419R	RIM	neutral	-2.1	0.6	0.1	0.3	Hetero-4-mer
CBS	P35520_HA501R	RIM	neutral	1.2	1.5	-0.1	-0.1	Homo-2-mer
CBS	P35520_KA485T	RIM	neutral	1.1	0.5	-2.5	-2.5	Homo-2-mer
CBS	P35520_MA505T	RIM	neutral	0.3	0.3	2	1.6	Homo-2-mer
CBS	P35520_QA242R	RIM	neutral	-0.1	0.5	0.6	0.5	Homo-2-mer
CBS	P35520_QA242W	RIM	neutral	-0.2	0.3	-5.5	-5.1	Homo-2-mer
CBS	P35520_SA175T	RIM	neutral	-0.4	0.5	-1.3	-0.2	Homo-2-mer
ETFA	P13804-P38117_DA201E	RIM	neutral	1.4	0.7	-0.8	-1.1	Hetero-2-mer
ETFA	P13804-P38117_EA128D	RIM	neutral	-0.2	1	-0.1	0.4	Hetero-2-mer
ETFA	P13804-P38117_EA128N	RIM	neutral	-0.2	2	2.7	3.2	Hetero-2-mer
ETFA	P13804-P38117_EA195G	RIM	neutral	-0.8	0.5	1.4	1.7	Hetero-2-mer
ETFA	P13804-P38117_IA329L	RIM	neutral	0.1	0.8	-0.9	-1.2	Hetero-2-mer
ETFA	P13804-P38117_KA162R	RIM	neutral	0	0.5	-1.6	-1.6	Hetero-2-mer
ETFA	P13804-P38117_LA204I	RIM	neutral	0.4	1.3	0.4	-0.2	Hetero-2-mer
ETFB	P11310-ETFA-ETFB_ES73V	RIM	neutral	1.1	0.6	-2.8	-2.4	Hetero-6-mer
ETFB	P11310-ETFA-ETFB_YS16F	RIM	neutral	-0.2	0.3	-0.5	0.1	Hetero-6-mer
ETFB	P13804-P38117_DB129E	RIM	neutral	0	0.4	1.1	1.2	Hetero-2-mer
ETFB	P13804-P38117_IB225L	RIM	neutral	-0.4	1.3	-0.4	-0.2	Hetero-2-mer
ETFB	P13804-P38117_IB225V	RIM	neutral	0.9	1.4	0.3	0.3	Hetero-2-mer
ETFB	P13804-P38117_IB252T	RIM	neutral	2.1	0.3	2.3	0.4	Hetero-2-mer
ETFB	P13804-P38117_IB252V	RIM	neutral	0.8	0.8	2.2	0.3	Hetero-2-mer
ETFB	P13804-P38117_KB250R	RIM	neutral	0.1	0.9	-0.7	-0.2	Hetero-2-mer
ETFB	P13804-P38117_LB222I	RIM	neutral	0.2	0.9	0.2	0.6	Hetero-2-mer
ETFB	P13804-P38117_TB234M	RIM	neutral	0.4	-0.8	0.2	0.2	Hetero-2-mer
ETFB	P13804-P38117_TB234V	RIM	neutral	0	-0.1	0.2	0.2	Hetero-2-mer
ETFB	P13804-P38117_TB31M	RIM	neutral	-0.8	0	-0.1	-0.1	Hetero-2-mer
ETFB	P13804-P38117_VB237T	RIM	neutral	0.2	0.6	1.6	1.3	Hetero-2-mer
ETFB	P13804-P38117_YB16F	RIM	neutral	-0.1	0.4	0.3	1.4	Hetero-2-mer
GALT	P07902_AB176S	RIM	neutral	0	0	0	0	Homo-2-mer
GALT	P07902_RB223H	RIM	neutral	0.6	0.1	0.6	0.6	Homo-2-mer
GCDH	Q92947_DA338N	RIM	neutral	-0.6	0.7	3.2	3.3	Homo-4-mer
GCDH	Q92947_DA360E	RIM	neutral	0	0.3	-0.1	-0.1	Homo-4-mer
GCDH	Q92947_DA418Q	RIM	neutral	0.6	0.6	7.9	7.1	Homo-4-mer
GCDH	Q92947_DA418R	RIM	neutral	1.1	-0.1	14.2	13.1	Homo-4-mer

GCDH	Q92947_DB338N	RIM	neutral	-0.4	0.7	3.2	3.3	Homo-4-mer
GCDH	Q92947_DB360E	RIM	neutral	0	0.3	-0.1	-0.1	Homo-4-mer
GCDH	Q92947_DB418Q	RIM	neutral	1.5	0.6	7.9	7.1	Homo-4-mer
GCDH	Q92947_DB418R	RIM	neutral	1.5	-0.1	14.2	13.1	Homo-4-mer
GCDH	Q92947_DC338N	RIM	neutral	-0.6	0.7	3.2	3.3	Homo-4-mer
GCDH	Q92947_DC360E	RIM	neutral	-0.1	0.3	-0.1	-0.1	Homo-4-mer
GCDH	Q92947_DC418Q	RIM	neutral	0.2	0.6	8	7.2	Homo-4-mer
GCDH	Q92947_DC418R	RIM	neutral	-0.1	-0.1	14.3	13.2	Homo-4-mer
GCDH	Q92947_DD338N	RIM	neutral	0.4	0.7	3.2	3.3	Homo-4-mer
GCDH	Q92947_DD360E	RIM	neutral	0	0.3	-0.1	-0.1	Homo-4-mer
GCDH	Q92947_DD418Q	RIM	neutral	-0.3	0.6	7.9	7.2	Homo-4-mer
GCDH	Q92947_DD418R	RIM	neutral	0.6	-0.1	14.2	13.2	Homo-4-mer
GCDH	Q92947_EA47K	RIM	neutral	-0.5	-0.1	-2.7	-2.8	Homo-4-mer
GCDH	Q92947_EA47Q	RIM	neutral	-0.9	0.5	-2.5	-2.4	Homo-4-mer
GCDH	Q92947_EB47K	RIM	neutral	-0.6	-0.1	-2.7	-2.8	Homo-4-mer
GCDH	Q92947_EB47Q	RIM	neutral	-0.4	0.5	-2.5	-2.4	Homo-4-mer
GCDH	Q92947_EC47K	RIM	neutral	-0.3	-0.1	-2.7	-2.7	Homo-4-mer
GCDH	Q92947_EC47Q	RIM	neutral	-0.3	0.5	-2.5	-2.4	Homo-4-mer
GCDH	Q92947_ED47K	RIM	neutral	-0.5	-0.1	-2.7	-2.8	Homo-4-mer
GCDH	Q92947_ED47Q	RIM	neutral	-0.3	0.5	-2.5	-2.4	Homo-4-mer
GCDH	Q92947_HA417V	RIM	neutral	0.7	0.5	-0.4	-0.4	Homo-4-mer
GCDH	Q92947_HB417V	RIM	neutral	-0.1	0.5	-0.4	-0.4	Homo-4-mer
GCDH	Q92947_HC417V	RIM	neutral	-0.2	0.5	-0.4	-0.4	Homo-4-mer
GCDH	Q92947_HD417V	RIM	neutral	-0.8	0.5	-0.4	-0.4	Homo-4-mer
GCDH	Q92947_LA331M	RIM	neutral	0	0.8	2.5	2	Homo-4-mer
GCDH	Q92947_LB331M	RIM	neutral	-0.1	0.8	2.5	2	Homo-4-mer
GCDH	Q92947_LC331M	RIM	neutral	0.4	0.8	2.5	2	Homo-4-mer
GCDH	Q92947_LD331M	RIM	neutral	0	0.8	2.5	2	Homo-4-mer
GCDH	Q92947_MA320I	RIM	neutral	0	0.9	0	0	Homo-4-mer
GCDH	Q92947_MB320I	RIM	neutral	-0.1	0.9	0	0	Homo-4-mer
GCDH	Q92947_MC320I	RIM	neutral	-0.1	0.9	0	0	Homo-4-mer
GCDH	Q92947_MD320I	RIM	neutral	-0.1	0.9	0	0	Homo-4-mer
GCDH	Q92947_QA432N	RIM	neutral	0.4	1	-0.3	-0.3	Homo-4-mer
GCDH	Q92947_QB432N	RIM	neutral	0.3	1	-0.3	-0.3	Homo-4-mer
GCDH	Q92947_QC432N	RIM	neutral	0.1	1	-0.3	-0.3	Homo-4-mer
GCDH	Q92947_QD432N	RIM	neutral	0.1	1	-0.3	-0.3	Homo-4-mer
GCDH	Q92947_RA328K	RIM	neutral	-0.7	1.1	-0.4	-0.3	Homo-4-mer
GCDH	Q92947_RA328L	RIM	neutral	-0.6	1.8	-4.3	-2.8	Homo-4-mer
GCDH	Q92947_RB328K	RIM	neutral	-0.6	1.1	-0.4	-0.3	Homo-4-mer
GCDH	Q92947_RB328L	RIM	neutral	-0.4	1.8	-4.3	-2.8	Homo-4-mer
GCDH	Q92947_RC328K	RIM	neutral	-0.6	1.1	-0.4	-0.3	Homo-4-mer
GCDH	Q92947_RC328L	RIM	neutral	-0.3	1.8	-4.2	-2.8	Homo-4-mer
GCDH	Q92947_RD328K	RIM	neutral	-0.6	1.1	-0.4	-0.3	Homo-4-mer
GCDH	Q92947_RD328L	RIM	neutral	-0.4	1.8	-4.2	-2.8	Homo-4-mer
HADHA	P40939-HADHB_EG223T	RIM	neutral	-1	2.1	3.2	3	Hetero-3-mer
HADHB	P40939-HADHB_KA298R	RIM	neutral	0.6	0.8	0	0	Hetero-3-mer
HADHB	P40939-HADHB_KB268R	RIM	neutral	1.2	1.3	6.1	4.2	Hetero-3-mer
HADHB	P40939-HADHB_RB90Q	RIM	neutral	1.3	0.6	0	0	Hetero-3-mer
HADHB	P40939-HADHB_SA220N	RIM	neutral	-0.1	0.4	-1.2	-1.4	Hetero-3-mer
HADHB	P40939-HADHB_WA397F	RIM	neutral	0	1.5	1.3	1.8	Hetero-3-mer
HADHB	P55084_VA171I	RIM	neutral	0.1	0.1	-0.4	-0.4	Homo-2-mer

HBB	P68871_YB146F	RIM	neutral	0.5	0.6	-2.9	-3.5	Homo-2-mer
HBB	P68871-HBA_EB44D	RIM	neutral	-0.1	0.9	-2.1	0	Hetero-4-mer
HBB	P68871-HBA_EB44K	RIM	neutral	0.1	0.7	3.2	5.3	Hetero-4-mer
HBB	P68871-HBA_ED44D	RIM	neutral	0.4	1.1	-2.1	0.1	Hetero-4-mer
HBB	P68871-HBA_ED44K	RIM	neutral	0.4	0.7	8.8	10.5	Hetero-4-mer
HBB	P68871-HBA_PB126Q	RIM	neutral	1.1	0.9	-0.7	-0.5	Hetero-4-mer
HBB	P68871-HBA_PB126T	RIM	neutral	2	0.7	-0.1	0	Hetero-4-mer
HBB	P68871-HBA_PB126V	RIM	neutral	1.5	0.9	-0.2	-0.1	Hetero-4-mer
HBB	P68871-HBA_PD126Q	RIM	neutral	1	1.3	-0.4	-0.4	Hetero-4-mer
HBB	P68871-HBA_PD126T	RIM	neutral	2.1	1.2	0	0	Hetero-4-mer
HBB	P68871-HBA_PD126V	RIM	neutral	1.5	1.6	-0.1	0	Hetero-4-mer
HBB	P68871-HBA_RB105K	RIM	neutral	-0.2	0.7	0.3	0.3	Hetero-4-mer
HBB	P68871-HBA_RD105K	RIM	neutral	-0.2	0.1	0.2	0.2	Hetero-4-mer
HBB	P68871-HBAZ_PB126Q	RIM	neutral	0.7	0.6	-0.2	0	Hetero-4-mer
HBB	P68871-HBAZ_PB126T	RIM	neutral	1.6	0.5	0.1	0.1	Hetero-4-mer
HBB	P68871-HBAZ_PB126V	RIM	neutral	1.4	0.9	-0.1	0	Hetero-4-mer
HBB	P68871-HBAZ_PD126Q	RIM	neutral	0.9	0.6	-0.3	0	Hetero-4-mer
HBB	P68871-HBAZ_PD126T	RIM	neutral	1.9	0.5	0	0.1	Hetero-4-mer
HBB	P68871-HBAZ_PD126V	RIM	neutral	1.6	0.9	-0.1	0	Hetero-4-mer
HBB	P68871-HBAZ_RB105K	RIM	neutral	0	0.1	-3.9	-3.8	Hetero-4-mer
HBB	P68871-HBAZ_RD105K	RIM	neutral	0.1	0.1	-2.5	-2.5	Hetero-4-mer
HBB	P68871-HP_EL44D	RIM	neutral	1.1	0.6	2	-0.2	Hetero-2-mer
HBB	P68871-HP_EL44K	RIM	neutral	-0.2	0.2	3	1.4	Hetero-2-mer
HBB	P68871-HP_RL105K	RIM	neutral	-0.6	0.4	5.2	4.6	Hetero-2-mer
HCFC1	P51610_AA368S	RIM	neutral	0.3	0	0	0	Hetero-4-mer
HCFC1	P51610_AC368S	RIM	neutral	0.3	0.3	0	0	Hetero-4-mer
HCFC1	P51610_LD1938P	RIM	neutral	0.2	0	0	0	Hetero-4-mer
HMGCL	P35914_TA28A	RIM	neutral	0.3	-0.2	-1	-1.9	Homo-2-mer
HPD	P32754_AA164V	RIM	neutral	-0.2	0	0.1	-0.3	Homo-2-mer
HPD	P32754_DA84N	RIM	neutral	0.4	1.2	-0.1	-0.1	Homo-2-mer
HPD	P32754_EA283Q	RIM	neutral	-0.1	1.8	1.2	1.2	Homo-2-mer
HPD	P32754_KA172Q	RIM	neutral	0.4	0.5	1	1	Homo-2-mer
HPD	P32754_NA79D	RIM	neutral	-1.6	0.4	0.4	0.4	Homo-2-mer
IVD	P26440_AA354P	RIM	neutral	-1.8	0.6	0	0	Homo-4-mer
IVD	P26440_AB354P	RIM	neutral	-1.7	0.6	0	0	Homo-4-mer
IVD	P26440_AC354P	RIM	neutral	-1.7	0.5	0	0	Homo-4-mer
IVD	P26440_AD354P	RIM	neutral	-0.9	0.5	0	0	Homo-4-mer
IVD	P26440_GA170C	RIM	neutral	1.4	-0.4	-0.3	-0.3	Homo-4-mer
IVD	P26440_GB170C	RIM	neutral	1.6	-0.3	-0.3	-0.3	Homo-4-mer
IVD	P26440_GC170C	RIM	neutral	1.4	-0.4	-0.3	-0.3	Homo-4-mer
IVD	P26440_GD170C	RIM	neutral	1.3	-0.3	-0.3	-0.3	Homo-4-mer
MLYCD	O95822_AA119R	RIM	neutral	-1.5	1.1	0.8	0.5	Homo-4-mer
MLYCD	O95822_AB119R	RIM	neutral	-0.7	0.8	3.9	2.6	Homo-4-mer
MLYCD	O95822_AC119R	RIM	neutral	-1.8	1.1	1	0.7	Homo-4-mer
MLYCD	O95822_AD119R	RIM	neutral	-0.7	0.9	3.4	2.5	Homo-4-mer
MLYCD	O95822_GA470S	RIM	neutral	0	-0.4	0	0	Homo-4-mer
MLYCD	O95822_GB470S	RIM	neutral	0.7	-0.4	0	0	Homo-4-mer
MLYCD	O95822_GC470S	RIM	neutral	0	-0.5	0.1	0.1	Homo-4-mer
MLYCD	O95822_GD470S	RIM	neutral	0.6	-0.5	0	0	Homo-4-mer
MMAA	Q8IVH4_HA378Y	RIM	neutral	0	0.9	-3.5	-2.2	Homo-2-mer
MMAA	Q8IVH4_IA398S	RIM	neutral	-0.4	0.3	0.6	0.7	Homo-2-mer

MMAA	Q8IVH4_IA398T	RIM	neutral	1.5	0.3	0.4	0.6	Homo-2-mer
MMAB	Q96EY8_MA239K	RIM	neutral	-0.2	0.3	-0.2	-0.4	Homo-3-mer
MMAB	Q96EY8_MB239K	RIM	neutral	0	0.3	-0.2	-0.6	Homo-3-mer
MMAB	Q96EY8_MC239K	RIM	neutral	0.5	0.5	4.4	2.5	Homo-3-mer
MMACHC	Q9Y4U1_PA42L	RIM	neutral	1.3	0	-1.4	-2.5	Homo-2-mer
MMACHC	Q9Y4U1_QA238E	RIM	neutral	0.5	-0.1	-0.3	-0.3	Homo-2-mer
MMACHC	Q9Y4U1_RA73Q	RIM	neutral	2.1	1.5	2.8	2.1	Homo-2-mer
MMADHC	Q9H3L0_EA133Q	RIM	neutral	0	1.2	-0.6	1.9	Homo-2-mer
MMADHC	Q9H3L0_HA288Y	RIM	neutral	-0.5	0.5	2.3	0.4	Homo-2-mer
MMADHC	Q9H3L0_IA195L	RIM	neutral	0.1	0.4	0	-0.1	Homo-2-mer
MMADHC	Q9H3L0_IA195N	RIM	neutral	0.5	0.4	-0.2	-0.2	Homo-2-mer
MMADHC	Q9H3L0_IA195T	RIM	neutral	0.3	0.1	0	0	Homo-2-mer
MMADHC	Q9H3L0_IA195V	RIM	neutral	0.6	0.4	0	0	Homo-2-mer
MMADHC	Q9H3L0_IA195Y	RIM	neutral	0	-0.2	4	0.3	Homo-2-mer
MMADHC	Q9H3L0_KA203E	RIM	neutral	1.2	0.1	0.6	0.6	Homo-2-mer
MMADHC	Q9H3L0_NA206S	RIM	neutral	0.2	-0.2	0	0	Homo-2-mer
MMADHC	Q9H3L0_RA145K	RIM	neutral	0.3	1.3	6.7	6.3	Homo-2-mer
MMADHC	Q9H3L0_VA199M	RIM	neutral	-0.9	0.7	1.6	0.1	Homo-2-mer
MMADHC	Q9H3L0_YA213F	RIM	neutral	-0.1	0.7	-0.3	-0.3	Homo-2-mer
MUT	P22033_DA498E	RIM	neutral	0.2	0.4	1.8	0.5	Homo-2-mer
MUT	P22033_DA498R	RIM	neutral	0.7	0.1	9.7	7.9	Homo-2-mer
MUT	P22033_EA83G	RIM	neutral	2	0.6	3.4	6.1	Homo-2-mer
MUT	P22033_FA102Y	RIM	neutral	0.7	-0.2	1.2	1.5	Homo-2-mer
MUT	P22033_KA75R	RIM	neutral	-0.3	1.2	-2.3	-1.3	Homo-2-mer
MUT	P22033_KA75S	RIM	neutral	2	0.9	0.8	1.5	Homo-2-mer
MUT	P22033_LA46M	RIM	neutral	0	0.1	1.9	1.5	Homo-2-mer
MUT	P22033_MA79K	RIM	neutral	0.3	0.2	-0.5	-0.5	Homo-2-mer
MUT	P22033_MA79N	RIM	neutral	0.9	0.4	-0.4	-0.4	Homo-2-mer
MUT	P22033_MA79Q	RIM	neutral	0.4	0.2	0	0	Homo-2-mer
MUT	P22033_MA79T	RIM	neutral	1	0.2	0	0	Homo-2-mer
MUT	P22033_MA79V	RIM	neutral	1.1	0.8	-0.2	-0.2	Homo-2-mer
MUT	P22033_PA82A	RIM	neutral	1.2	0.7	0	0	Homo-2-mer
MUT	P22033_PA82L	RIM	neutral	0.6	0.6	0	0	Homo-2-mer
MUT	P22033_RA76E	RIM	neutral	0.5	2.1	4.6	4.7	Homo-2-mer
PAH	P00439_RA413H	RIM	neutral	1.2	0.2	-9.4	-6.9	Homo-2-mer
PAH	P00439_TA236S	RIM	neutral	-0.1	0.7	-0.3	-0.2	Homo-2-mer
PAH	P00439-D1_RA53H	RIM	neutral	0.8	0.3	-3.1	-3.3	Homo-2-mer
SLC24A13	Q9UJS0_AA173G	RIM	neutral	0.2	0.4	0	0	Homo-2-mer
SLC24A13	Q9UJS0_AA173S	RIM	neutral	0.3	-0.1	0	0	Homo-2-mer
SLC24A13	Q9UJS0_KA140R	RIM	neutral	0	0.5	4.7	4	Homo-2-mer
SLC24A13	Q9UJS0_KA653R	RIM	neutral	0.1	0.4	-2.5	-2.8	Homo-2-mer
SLC24A13	Q9UJS0_RA174K	RIM	neutral	-0.9	1.6	-1.2	-1	Homo-2-mer
SLC24A13	Q9UJS0_RA174N	RIM	neutral	0.4	2.2	-3.5	-2.4	Homo-2-mer
SLC24A13	Q9UJS0_RA174S	RIM	neutral	1.1	2.1	-4.7	-3.3	Homo-2-mer
SLC24A13	Q9UJS0_RA177K	RIM	neutral	0	0.4	-2.2	-2.2	Homo-2-mer
SLC24A13	Q9UJS0_TA147A	RIM	neutral	0.1	0.3	0	0.4	Homo-2-mer
SLC24A13	Q9UJS0_TA147N	RIM	neutral	-0.6	0.8	2.5	2.3	Homo-2-mer
SLC24A13	Q9UJS0_TA175A	RIM	neutral	0.4	0.7	-0.1	-0.1	Homo-2-mer
SLC24A13	Q9UJS0_TA175M	RIM	neutral	0	0.2	-0.6	-0.7	Homo-2-mer
SLC24A13	Q9UJS0_TA175N	RIM	neutral	0.3	1	-0.4	-0.4	Homo-2-mer
SLC24A13	Q9UJS0_TA211I	RIM	neutral	-0.3	0	-1.4	-1.1	Homo-2-mer

SLC24A13	Q9UJS0_TA211R	RIM	neutral	-0.7	-0.3	-1.7	-1.5	Homo-2-mer
SLC24A13	Q9UJS0_VA168L	RIM	neutral	-1.2	0.9	-0.1	-1.6	Homo-2-mer
SLC24A13	Q9UJS0_VA168M	RIM	neutral	-1.6	0.6	0.7	-0.1	Homo-2-mer
TAT	P17735_RA315C	RIM	neutral	2.6	1.9	-1.6	0.1	Homo-2-mer
TGFB1	P01137_ED67D	RIM	neutral	-0.2	0.9	0	-0.1	Homo-2-mer
TGFB1	P01137_RD87Q	RIM	neutral	-1.1	0.2	-4.1	-3.8	Homo-2-mer
TGFB1	P01137_SD64N	RIM	neutral	-0.3	0.4	0.1	0.1	Homo-2-mer
TGFB1	P01137-ITGAV-ITGB6_GC242S	RIM	neutral	0.5	-0.4	1.2	0.4	Hetero-5-mer
TGFB1	P01137-ITGAV-ITGB6_GD242S	RIM	neutral	0	0	-0.8	-0.7	Hetero-5-mer
TGFB1	P01137-ITGAV-ITGB6_QC126H	RIM	neutral	0.7	-0.3	2	1.5	Hetero-5-mer
TGFB1	P01137-ITGAV-ITGB6_QC126K	RIM	neutral	-0.1	0.3	5.2	5.3	Hetero-5-mer
TGFB1	P01137-ITGAV-ITGB6_QC126R	RIM	neutral	-0.1	0.3	-1.1	-1.2	Hetero-5-mer
TGFB1	P01137-ITGAV-ITGB6_QC126S	RIM	neutral	0.5	0.6	1.4	1.1	Hetero-5-mer
TGFB1	P01137-ITGAV-ITGB6_TC128P	RIM	neutral	-2.7	0.2	0.3	0.2	Hetero-5-mer
TGFB1	P01137-ITGAV-ITGB6_TC128S	RIM	neutral	-0.2	0.4	0.8	0.7	Hetero-5-mer
ACADM	P11310_AA318E	CORE	pathogenic	0.2	1.9	-0.9	0.5	Homo-4-mer
ACADM	P11310_AB318E	CORE	pathogenic	0.1	2	-1.2	0.2	Homo-4-mer
ACADM	P11310_AC318E	CORE	pathogenic	0.2	1.9	-0.7	0.9	Homo-4-mer
ACADM	P11310_AD318E	CORE	pathogenic	0.4	2	-1	0.7	Homo-4-mer
ACADM	P11310_DA168A	CORE	pathogenic	1.8	2.1	1.9	1.2	Homo-4-mer
ACADM	P11310_DA168V	CORE	pathogenic	2.7	1.8	-1.4	2.1	Homo-4-mer
ACADM	P11310_DA297G	CORE	pathogenic	2.7	0.4	1.3	0.9	Homo-4-mer
ACADM	P11310_DB168A	CORE	pathogenic	1.4	2.1	0.6	0.3	Homo-4-mer
ACADM	P11310_DB168V	CORE	pathogenic	2.7	1.8	1.1	1.1	Homo-4-mer
ACADM	P11310_DB297G	CORE	pathogenic	4.9	1	-0.1	0.6	Homo-4-mer
ACADM	P11310_DC168A	CORE	pathogenic	1.7	2.1	1.2	0.9	Homo-4-mer
ACADM	P11310_DC168V	CORE	pathogenic	2.5	1.9	-1.8	1.6	Homo-4-mer
ACADM	P11310_DC297G	CORE	pathogenic	4.3	1	0.1	0.8	Homo-4-mer
ACADM	P11310_DD168A	CORE	pathogenic	1.3	2.1	1	0.7	Homo-4-mer
ACADM	P11310_DD168V	CORE	pathogenic	2.8	1.9	-1.9	1.6	Homo-4-mer
ACADM	P11310_DD297G	CORE	pathogenic	4.6	1	3.9	0.8	Homo-4-mer
ACADM	P11310_IA356F	CORE	pathogenic	9.8	0.5	-3.5	-3	Homo-4-mer
ACADM	P11310_IA356T	CORE	pathogenic	2.8	0.7	0.9	0.7	Homo-4-mer
ACADM	P11310_IB356F	CORE	pathogenic	6.8	0.3	-3.4	-2.9	Homo-4-mer
ACADM	P11310_IB356T	CORE	pathogenic	2.6	0.6	0.9	0.7	Homo-4-mer
ACADM	P11310_IB416T	CORE	pathogenic	1.8	0.7	2	2	Homo-4-mer
ACADM	P11310_IC356F	CORE	pathogenic	4	0.5	-3.3	-2.7	Homo-4-mer
ACADM	P11310_IC356T	CORE	pathogenic	2.7	0.8	0.9	0.7	Homo-4-mer
ACADM	P11310_IC416T	CORE	pathogenic	2.3	0.7	2	1.9	Homo-4-mer
ACADM	P11310_ID356F	CORE	pathogenic	9.3	0.5	-3.5	-2.9	Homo-4-mer
ACADM	P11310_ID356T	CORE	pathogenic	2.7	0.7	0.8	0.6	Homo-4-mer
ACADM	P11310_ID416T	CORE	pathogenic	2	0.5	1.4	1.9	Homo-4-mer
ACADM	P11310_KA395R	CORE	pathogenic	5.7	1.9	-0.2	0.3	Homo-4-mer
ACADM	P11310_KB395R	CORE	pathogenic	5	1.9	-0.1	0.3	Homo-4-mer

ACADM	P11310_KC395R	CORE	pathogenic	5.1	1.9	0	0.5	Homo-4-mer
ACADM	P11310_KD395R	CORE	pathogenic	5.2	1.8	-0.6	-0.1	Homo-4-mer
ACADM	P11310_MA328V	CORE	pathogenic	1.3	0.9	1.2	1.1	Homo-4-mer
ACADM	P11310_MB328V	CORE	pathogenic	1.6	0.9	1.3	1.2	Homo-4-mer
ACADM	P11310_MC328V	CORE	pathogenic	1.3	0.9	1.2	1.1	Homo-4-mer
ACADM	P11310_MD328V	CORE	pathogenic	1.2	0.9	1.3	1.2	Homo-4-mer
ACADM	P11310_NA379K	CORE	pathogenic	-0.9	1	2.4	-0.8	Homo-4-mer
ACADM	P11310_NB379K	CORE	pathogenic	-0.9	1	2.5	-0.5	Homo-4-mer
ACADM	P11310_NC379K	CORE	pathogenic	-0.3	1	2.5	-0.6	Homo-4-mer
ACADM	P11310_ND379K	CORE	pathogenic	-1	1.2	2.5	-0.6	Homo-4-mer
ACADM	P11310-ETFA-ETFB_ED43K	CORE	pathogenic	2.8	1	7.1	10.8	Hetero-6-mer
ACADM	P11310-ETFA-ETFB_GD85C	CORE	pathogenic	7.6	0.6	0.2	1.3	Hetero-6-mer
ACADM	P11310-ETFA-ETFB_GD85R	CORE	pathogenic	9.4	2.5	6.5	7.9	Hetero-6-mer
ACADM	P11310-ETFA-ETFB_KA395R	CORE	pathogenic	-0.1	-0.2	-6.6	-6.7	Hetero-6-mer
ACADM	P11310-ETFA-ETFB_LD84F	CORE	pathogenic	-0.5	0.3	-3.9	-2.1	Hetero-6-mer
ACADM	P11310-ETFA-ETFB_ND379K	CORE	pathogenic	0.9	0.9	8.2	10.3	Hetero-6-mer
ACADS	P16219_AE315V	CORE	pathogenic	3.1	2.2	-0.9	0.4	Homo-4-mer
ACADS	P16219_AF315V	CORE	pathogenic	2.7	2.1	-0.9	0.3	Homo-4-mer
ACADS	P16219_AG315V	CORE	pathogenic	3.7	2.1	-1	0.2	Homo-4-mer
ACADS	P16219_AH315V	CORE	pathogenic	3.9	2.2	-0.3	0.6	Homo-4-mer
ACADS	P16219_EE228K	CORE	pathogenic	-0.7	0.9	-1.5	-2	Homo-4-mer
ACADS	P16219_EF228K	CORE	pathogenic	-1.3	0.8	-1.5	-1.5	Homo-4-mer
ACADS	P16219_EG228K	CORE	pathogenic	-1.4	1.1	-0.3	-1	Homo-4-mer
ACADS	P16219_EH228K	CORE	pathogenic	-1.6	1.1	-0.5	-0.4	Homo-4-mer
ACADS	P16219_GE368S	CORE	pathogenic	3.3	0	1.1	0.8	Homo-4-mer
ACADS	P16219_GF368S	CORE	pathogenic	3.4	0	0.8	0.6	Homo-4-mer
ACADS	P16219_GG368S	CORE	pathogenic	3.2	0	0.9	0.7	Homo-4-mer
ACADS	P16219_GH368S	CORE	pathogenic	3.8	0	0.6	0.7	Homo-4-mer
ACADS	P16219_IE390M	CORE	pathogenic	0.3	0.6	-1	1.1	Homo-4-mer
ACADS	P16219_IF390M	CORE	pathogenic	0.4	0.7	-4.2	1.2	Homo-4-mer
ACADS	P16219_IG390M	CORE	pathogenic	0.2	0.7	-0.9	1.1	Homo-4-mer
ACADS	P16219_IH390M	CORE	pathogenic	0.3	0.7	-0.9	1.1	Homo-4-mer
ACADS	P16219_ME370V	CORE	pathogenic	1.2	1.2	-0.1	-0.1	Homo-4-mer
ACADS	P16219_MF370V	CORE	pathogenic	1.7	1.1	-0.1	-0.1	Homo-4-mer
ACADS	P16219_MG370V	CORE	pathogenic	1.6	1.1	0.3	0	Homo-4-mer
ACADS	P16219_MH370V	CORE	pathogenic	1.6	1.1	0.2	-0.1	Homo-4-mer
ACADS	P16219_QE365H	CORE	pathogenic	4.6	1	1.9	-1.8	Homo-4-mer
ACADS	P16219_QF365H	CORE	pathogenic	2.1	1	-5.1	-6.1	Homo-4-mer
ACADS	P16219_QG365H	CORE	pathogenic	4	0.9	-1.5	-1.4	Homo-4-mer
ACADS	P16219_QH365H	CORE	pathogenic	5.6	1.1	-8	-5.9	Homo-4-mer
ACADS	P16219_RE330C	CORE	pathogenic	1.3	1.2	3.1	2.6	Homo-4-mer
ACADS	P16219_RE330H	CORE	pathogenic	0.7	0.9	4	3.6	Homo-4-mer
ACADS	P16219_RE386C	CORE	pathogenic	3.9	2.5	-1.5	-2.4	Homo-4-mer
ACADS	P16219_RE386H	CORE	pathogenic	5	2.4	-3.9	-2.7	Homo-4-mer
ACADS	P16219_RF330C	CORE	pathogenic	1	0.9	0.8	2.5	Homo-4-mer
ACADS	P16219_RF330H	CORE	pathogenic	0.7	0.8	1.4	3.2	Homo-4-mer

ACADS	P16219_RF386C	CORE	pathogenic	4.6	2.7	-1.4	-2.3	Homo-4-mer
ACADS	P16219_RF386H	CORE	pathogenic	6.8	2	-6.4	-1.6	Homo-4-mer
ACADS	P16219_RG330C	CORE	pathogenic	1.4	1.1	0.6	2.4	Homo-4-mer
ACADS	P16219_RG330H	CORE	pathogenic	0.9	0.9	1.6	3.4	Homo-4-mer
ACADS	P16219_RG386C	CORE	pathogenic	4.6	2.5	-0.4	-2.6	Homo-4-mer
ACADS	P16219_RG386H	CORE	pathogenic	7.1	2.3	-2.4	-2.7	Homo-4-mer
ACADS	P16219_RH330C	CORE	pathogenic	1.2	1.3	2.8	2.4	Homo-4-mer
ACADS	P16219_RH330H	CORE	pathogenic	0.8	1	3.7	3.4	Homo-4-mer
ACADS	P16219_RH386C	CORE	pathogenic	3.7	2.5	1.6	-2.5	Homo-4-mer
ACADS	P16219_RH386H	CORE	pathogenic	8.4	2.2	-0.2	-2.4	Homo-4-mer
ACADSB	P45954_EA387K	CORE	pathogenic	-0.1	2	2	7.6	Homo-4-mer
ACADSB	P45954_EB387K	CORE	pathogenic	-0.3	1.9	6.7	10.7	Homo-4-mer
ACADSB	P45954_EC387K	CORE	pathogenic	-0.1	1.8	6.5	10.8	Homo-4-mer
ACADSB	P45954_ED387K	CORE	pathogenic	0.5	1.8	1.4	6.8	Homo-4-mer
ACADVL	P49748_DA431G	CORE	pathogenic	4.5	3.3	4.8	5.2	Homo-2-mer
ACADVL	P49748_DA466Y	CORE	pathogenic	3.9	1.1	-3.6	-3.4	Homo-2-mer
ACADVL	P49748_GA439D	CORE	pathogenic	1.4	0.3	-4	0.1	Homo-2-mer
ACADVL	P49748_IA457N	CORE	pathogenic	3.5	1.4	1.9	0.4	Homo-2-mer
ACADVL	P49748_KA299M	CORE	pathogenic	0.1	2.4	-0.2	-0.2	Homo-2-mer
ACADVL	P49748_KA299N	CORE	pathogenic	3.2	1.3	0.6	0.6	Homo-2-mer
ACADVL	P49748_LA641F	CORE	pathogenic	8.3	1.1	-6.4	-4.6	Homo-2-mer
ACADVL	P49748_MA443R	CORE	pathogenic	1.9	1.4	0	2	Homo-2-mer
ACADVL	P49748_MA443T	CORE	pathogenic	2.8	0.6	1.5	1.1	Homo-2-mer
ACADVL	P49748_QA368P	CORE	pathogenic	5.2	1.5	1	1.3	Homo-2-mer
ACADVL	P49748_RA162H	CORE	pathogenic	3.5	1.3	-11.4	-5.4	Homo-2-mer
ACADVL	P49748_RA456H	CORE	pathogenic	4.3	1.9	0	4.4	Homo-2-mer
ACADVL	P49748_RA567Q	CORE	pathogenic	2.1	0.1	-0.9	1.5	Homo-2-mer
ACAT1	P24752_QA145E	CORE	pathogenic	2.4	0.4	0.1	0.1	Homo-4-mer
ACAT1	P24752_QB145E	CORE	pathogenic	2.7	0.3	0.3	0.4	Homo-4-mer
ACAT1	P24752_QC145E	CORE	pathogenic	2.5	0.4	-0.4	-0.4	Homo-4-mer
ACAT1	P24752_QD145E	CORE	pathogenic	1.9	0.3	-0.2	-0.2	Homo-4-mer
ARG1	P05089_RA308Q	CORE	pathogenic	3	1.5	1.2	2.9	Homo-3-mer
ARG1	P05089_RB308Q	CORE	pathogenic	3.1	1.5	1.2	2.9	Homo-3-mer
ARG1	P05089_RC308Q	CORE	pathogenic	3.1	1.5	1.2	2.9	Homo-3-mer
ARG1	P05089_TA134I	CORE	pathogenic	1	0.3	-1.1	0	Homo-3-mer
ARG1	P05089_TB134I	CORE	pathogenic	1	0.3	-1.1	0	Homo-3-mer
ARG1	P05089_TC134I	CORE	pathogenic	1	0.3	-1.1	0	Homo-3-mer
ASL	P04424_AA205V	CORE	pathogenic	2.3	1	-7.2	-1.5	Homo-4-mer
ASL	P04424_AB205V	CORE	pathogenic	3.1	1	-7.2	-1.5	Homo-4-mer
ASL	P04424_AC205V	CORE	pathogenic	3.4	1.4	-7	-1.2	Homo-4-mer
ASL	P04424_AD205V	CORE	pathogenic	1.5	1.4	-7	-1.2	Homo-4-mer
ASL	P04424_DA145G	CORE	pathogenic	0	0	-0.3	-0.3	Homo-4-mer
ASL	P04424_DA231E	CORE	pathogenic	0.9	1.3	-2.5	-2.3	Homo-4-mer
ASL	P04424_DA237N	CORE	pathogenic	2.5	1.9	10.2	10.2	Homo-4-mer
ASL	P04424_DA324A	CORE	pathogenic	1	2.3	1.8	3.3	Homo-4-mer
ASL	P04424_DB145G	CORE	pathogenic	0.1	0	-0.3	-0.3	Homo-4-mer
ASL	P04424_DB231E	CORE	pathogenic	1.9	1.3	-2.5	-2.3	Homo-4-mer
ASL	P04424_DB237N	CORE	pathogenic	1.6	1.9	10.2	10.2	Homo-4-mer
ASL	P04424_DB324A	CORE	pathogenic	0.7	2.3	1.8	3.3	Homo-4-mer
ASL	P04424_DC145G	CORE	pathogenic	0.3	1.1	1.5	0.9	Homo-4-mer
ASL	P04424_DC231E	CORE	pathogenic	1.7	1.1	-2.4	-2.2	Homo-4-mer

ASL	P04424_DC237N	CORE	pathogenic	-0.9	1.8	10.1	10.3	Homo-4-mer
ASL	P04424_DC324A	CORE	pathogenic	0.6	2.1	1.3	2.9	Homo-4-mer
ASL	P04424_DD145G	CORE	pathogenic	0.1	1.1	1.5	0.9	Homo-4-mer
ASL	P04424_DD231E	CORE	pathogenic	2.8	1.1	-2.4	-2.2	Homo-4-mer
ASL	P04424_DD237N	CORE	pathogenic	0.1	1.8	10.1	10.4	Homo-4-mer
ASL	P04424_DD324A	CORE	pathogenic	1	2.1	1.4	3	Homo-4-mer
ASL	P04424_EA189G	CORE	pathogenic	1.8	2.9	-0.4	0.1	Homo-4-mer
ASL	P04424_EA241K	CORE	pathogenic	3.2	1.9	6.2	6.3	Homo-4-mer
ASL	P04424_EB189G	CORE	pathogenic	1.3	2.9	-0.4	0.1	Homo-4-mer
ASL	P04424_EB241K	CORE	pathogenic	3.2	1.9	6.2	6.3	Homo-4-mer
ASL	P04424_EC189G	CORE	pathogenic	1.4	3.1	-0.1	0.6	Homo-4-mer
ASL	P04424_EC241K	CORE	pathogenic	3.3	1.5	5.5	7.6	Homo-4-mer
ASL	P04424_ED189G	CORE	pathogenic	1.7	3.1	-0.1	0.6	Homo-4-mer
ASL	P04424_ED241K	CORE	pathogenic	3.7	1.5	5.5	7.6	Homo-4-mer
ASL	P04424_GA301R	CORE	pathogenic	15.3	2.9	6	5.4	Homo-4-mer
ASL	P04424_GA309R	CORE	pathogenic	5.8	1.9	3.2	4	Homo-4-mer
ASL	P04424_GA316E	CORE	pathogenic	6.3	1.3	1.9	-3.3	Homo-4-mer
ASL	P04424_GB301R	CORE	pathogenic	14.4	2.9	6	5.4	Homo-4-mer
ASL	P04424_GB309R	CORE	pathogenic	5.1	1.9	3.2	4	Homo-4-mer
ASL	P04424_GB316E	CORE	pathogenic	7.6	1.3	1.9	-3.3	Homo-4-mer
ASL	P04424_GC301R	CORE	pathogenic	15.2	2.9	8.5	8.9	Homo-4-mer
ASL	P04424_GC309R	CORE	pathogenic	6.4	2	2.5	3	Homo-4-mer
ASL	P04424_GC316E	CORE	pathogenic	6.7	2.2	-0.8	-0.6	Homo-4-mer
ASL	P04424_GD301R	CORE	pathogenic	10.9	2.9	8.5	8.8	Homo-4-mer
ASL	P04424_GD309R	CORE	pathogenic	3.9	2	2.5	3	Homo-4-mer
ASL	P04424_GD316E	CORE	pathogenic	6.5	2.2	-0.8	-0.7	Homo-4-mer
ASL	P04424_HA388Q	CORE	pathogenic	0.1	0.5	3	2.8	Homo-4-mer
ASL	P04424_HB388Q	CORE	pathogenic	0.1	0.5	3	2.8	Homo-4-mer
ASL	P04424_HC388Q	CORE	pathogenic	-0.1	1.1	3.2	3.1	Homo-4-mer
ASL	P04424_HD388Q	CORE	pathogenic	0	1.1	3.2	3.1	Homo-4-mer
ASL	P04424_LA227P	CORE	pathogenic	1.1	0.6	4.2	1.3	Homo-4-mer
ASL	P04424_LA295P	CORE	pathogenic	6.3	0.8	8.2	1.8	Homo-4-mer
ASL	P04424_LB227P	CORE	pathogenic	1.3	0.6	4.2	1.3	Homo-4-mer
ASL	P04424_LB295P	CORE	pathogenic	6.2	0.8	8.2	1.8	Homo-4-mer
ASL	P04424_LC227P	CORE	pathogenic	1.3	0.2	4	1.6	Homo-4-mer
ASL	P04424_LC295P	CORE	pathogenic	6.4	0.7	1.4	1.6	Homo-4-mer
ASL	P04424_LD227P	CORE	pathogenic	1.3	0.2	4	1.6	Homo-4-mer
ASL	P04424_LD295P	CORE	pathogenic	6.4	0.7	1.4	1.6	Homo-4-mer
ASL	P04424_PA208R	CORE	pathogenic	7.8	1.7	3.1	7.6	Homo-4-mer
ASL	P04424_PA208T	CORE	pathogenic	4.5	1.7	-6.7	-0.1	Homo-4-mer
ASL	P04424_PB208R	CORE	pathogenic	8.8	1.7	3.1	7.6	Homo-4-mer
ASL	P04424_PB208T	CORE	pathogenic	5.5	1.7	-6.7	-0.1	Homo-4-mer
ASL	P04424_PC208R	CORE	pathogenic	10.4	2.1	5.1	6.3	Homo-4-mer
ASL	P04424_PC208T	CORE	pathogenic	3	2	-3.2	-0.2	Homo-4-mer
ASL	P04424_PD208R	CORE	pathogenic	10.3	2.1	5.1	6.3	Homo-4-mer
ASL	P04424_PD208T	CORE	pathogenic	3.1	2	-3.2	-0.2	Homo-4-mer
ASL	P04424_QA326H	CORE	pathogenic	1.5	-0.2	-3.4	-2.7	Homo-4-mer
ASL	P04424_QA326L	CORE	pathogenic	-0.1	0.2	0	0	Homo-4-mer
ASL	P04424_QB326H	CORE	pathogenic	1	-0.2	-3.4	-2.7	Homo-4-mer
ASL	P04424_QB326L	CORE	pathogenic	-0.6	0.2	0	0	Homo-4-mer
ASL	P04424_QC326H	CORE	pathogenic	0.3	-0.2	-3.1	-2.4	Homo-4-mer

ASL	P04424_QC326L	CORE	pathogenic	0	-0.1	0.4	0.4	Homo-4-mer
ASL	P04424_QD326H	CORE	pathogenic	0.8	-0.2	-3.1	-2.4	Homo-4-mer
ASL	P04424_QD326L	CORE	pathogenic	-0.6	-0.1	0.4	0.4	Homo-4-mer
ASL	P04424_RA182G	CORE	pathogenic	4.9	0.8	5.4	7.2	Homo-4-mer
ASL	P04424_RA182Q	CORE	pathogenic	3.6	0.3	6.8	6.1	Homo-4-mer
ASL	P04424_RA186Q	CORE	pathogenic	2	-0.3	4.6	4	Homo-4-mer
ASL	P04424_RA186W	CORE	pathogenic	2.8	0.4	-3.1	-3.5	Homo-4-mer
ASL	P04424_RA193Q	CORE	pathogenic	2.8	0.5	-1	0.4	Homo-4-mer
ASL	P04424_RA193W	CORE	pathogenic	1.4	0.3	-7.6	-7.7	Homo-4-mer
ASL	P04424_RA297Q	CORE	pathogenic	4.5	1.5	-0.7	1	Homo-4-mer
ASL	P04424_RA297W	CORE	pathogenic	5.5	0.3	-8.7	-7.8	Homo-4-mer
ASL	P04424_RA306W	CORE	pathogenic	-3.2	0.8	-14.6	-14.6	Homo-4-mer
ASL	P04424_RB182G	CORE	pathogenic	5.2	0.8	5.4	7.2	Homo-4-mer
ASL	P04424_RB182Q	CORE	pathogenic	3.9	0.3	6.8	6.1	Homo-4-mer
ASL	P04424_RB186Q	CORE	pathogenic	3	-0.3	4.6	4	Homo-4-mer
ASL	P04424_RB186W	CORE	pathogenic	3.3	0.4	-3.1	-3.5	Homo-4-mer
ASL	P04424_RB193Q	CORE	pathogenic	2.7	0.5	-1	0.4	Homo-4-mer
ASL	P04424_RB193W	CORE	pathogenic	0.9	0.3	-7.6	-7.7	Homo-4-mer
ASL	P04424_RB297Q	CORE	pathogenic	4.1	1.5	-0.7	1	Homo-4-mer
ASL	P04424_RB297W	CORE	pathogenic	6.2	0.3	-8.7	-7.8	Homo-4-mer
ASL	P04424_RB306W	CORE	pathogenic	-2.6	0.8	-14.6	-14.6	Homo-4-mer
ASL	P04424_RC12Q	CORE	pathogenic	1.6	1.9	3.5	3.2	Homo-4-mer
ASL	P04424_RC182G	CORE	pathogenic	4.8	0.6	3.2	7.2	Homo-4-mer
ASL	P04424_RC182Q	CORE	pathogenic	3.5	0.2	4.3	5.8	Homo-4-mer
ASL	P04424_RC186Q	CORE	pathogenic	1.7	-0.4	4.1	3.6	Homo-4-mer
ASL	P04424_RC186W	CORE	pathogenic	1.3	0.4	-2.7	-3.7	Homo-4-mer
ASL	P04424_RC193Q	CORE	pathogenic	3.1	0.9	0.9	0.9	Homo-4-mer
ASL	P04424_RC193W	CORE	pathogenic	0.2	0.6	-11.2	-9.1	Homo-4-mer
ASL	P04424_RC297Q	CORE	pathogenic	5.5	1.2	1.4	4	Homo-4-mer
ASL	P04424_RC297W	CORE	pathogenic	6.8	0.2	-9.3	-8.4	Homo-4-mer
ASL	P04424_RC306W	CORE	pathogenic	-2.4	0.7	-14.3	-14.3	Homo-4-mer
ASL	P04424_RD12Q	CORE	pathogenic	1.7	1.9	3.5	3.2	Homo-4-mer
ASL	P04424_RD182G	CORE	pathogenic	4.8	0.6	3.2	7.2	Homo-4-mer
ASL	P04424_RD182Q	CORE	pathogenic	3.5	0.2	4.3	5.8	Homo-4-mer
ASL	P04424_RD186Q	CORE	pathogenic	1.7	-0.4	4.1	3.6	Homo-4-mer
ASL	P04424_RD186W	CORE	pathogenic	1.3	0.4	-2.7	-3.7	Homo-4-mer
ASL	P04424_RD193Q	CORE	pathogenic	2.5	0.9	0.9	0.8	Homo-4-mer
ASL	P04424_RD193W	CORE	pathogenic	0.6	0.6	-11.2	-9.2	Homo-4-mer
ASL	P04424_RD297Q	CORE	pathogenic	5	1.2	1.4	4	Homo-4-mer
ASL	P04424_RD297W	CORE	pathogenic	7	0.2	-9.3	-8.4	Homo-4-mer
ASL	P04424_RD306W	CORE	pathogenic	-2.8	0.7	-14.3	-14.3	Homo-4-mer
ASL	P04424_TA159N	CORE	pathogenic	1.6	0.9	1	0.3	Homo-4-mer
ASL	P04424_TB159N	CORE	pathogenic	0.5	0.9	1	0.3	Homo-4-mer
ASL	P04424_TC159N	CORE	pathogenic	1.6	1	0.7	0.8	Homo-4-mer
ASL	P04424_TD159N	CORE	pathogenic	1.3	1	0.7	0.8	Homo-4-mer
ASL	P04424_VA178M	CORE	pathogenic	-0.6	0.6	-0.5	0.6	Homo-4-mer
ASL	P04424_VB178M	CORE	pathogenic	-0.3	0.6	-0.5	0.6	Homo-4-mer
ASL	P04424_VC178M	CORE	pathogenic	-1	0.3	-0.1	0.1	Homo-4-mer
ASL	P04424_VD178M	CORE	pathogenic	-0.5	0.3	-0.1	0.1	Homo-4-mer
ASS1	P00966_GA390R	CORE	pathogenic	8.9	1.2	4.9	5	Homo-4-mer
ASS1	P00966_GB390R	CORE	pathogenic	7	1.2	4.9	5	Homo-4-mer

ASS1	P00966_GC390R	CORE	pathogenic	8.5	1.2	5	5.1	Homo-4-mer
ASS1	P00966_GD390R	CORE	pathogenic	9.1	1.2	5	5.1	Homo-4-mer
ASS1	P00966_IA394N	CORE	pathogenic	1.8	0.7	4.8	2.6	Homo-4-mer
ASS1	P00966_IB394N	CORE	pathogenic	1.9	0.7	4.8	2.6	Homo-4-mer
ASS1	P00966_IC394N	CORE	pathogenic	2.1	0.7	4.8	2.6	Homo-4-mer
ASS1	P00966_ID394N	CORE	pathogenic	1.9	0.7	4.8	2.6	Homo-4-mer
ASS1	P00966_KA277T	CORE	pathogenic	3.9	1.3	0.4	1.1	Homo-4-mer
ASS1	P00966_KB277T	CORE	pathogenic	3.8	1.3	0.4	1.1	Homo-4-mer
ASS1	P00966_KC277T	CORE	pathogenic	3.8	1.3	0.4	1.1	Homo-4-mer
ASS1	P00966_KD277T	CORE	pathogenic	3.8	1.3	0.4	1.1	Homo-4-mer
ASS1	P00966_RA265C	CORE	pathogenic	0.7	1.7	0.1	-0.1	Homo-4-mer
ASS1	P00966_RA265H	CORE	pathogenic	3.9	1.5	-2.6	-1.8	Homo-4-mer
ASS1	P00966_RA304W	CORE	pathogenic	0.5	0.9	-8.7	-7.4	Homo-4-mer
ASS1	P00966_RA307C	CORE	pathogenic	1.3	0.6	1.4	2.6	Homo-4-mer
ASS1	P00966_RA363G	CORE	pathogenic	5.7	1.1	-4.5	-4	Homo-4-mer
ASS1	P00966_RA363L	CORE	pathogenic	3.7	0.9	-6.3	-5.8	Homo-4-mer
ASS1	P00966_RA363Q	CORE	pathogenic	4.9	1.4	-5.4	-4.9	Homo-4-mer
ASS1	P00966_RA363W	CORE	pathogenic	6.4	0.8	-13.2	-11.8	Homo-4-mer
ASS1	P00966_RA86C	CORE	pathogenic	1.7	1.1	-3.2	-2.6	Homo-4-mer
ASS1	P00966_RA86H	CORE	pathogenic	1.3	0.8	-1.9	-1.8	Homo-4-mer
ASS1	P00966_RB265C	CORE	pathogenic	1.1	1.7	0.1	-0.1	Homo-4-mer
ASS1	P00966_RB265H	CORE	pathogenic	4.1	1.5	-2.6	-1.8	Homo-4-mer
ASS1	P00966_RB304W	CORE	pathogenic	1.5	0.9	-8.7	-7.4	Homo-4-mer
ASS1	P00966_RB307C	CORE	pathogenic	2.2	0.6	1.4	2.6	Homo-4-mer
ASS1	P00966_RB363G	CORE	pathogenic	4.8	1.1	-4.5	-4	Homo-4-mer
ASS1	P00966_RB363L	CORE	pathogenic	1.3	0.9	-6.3	-5.8	Homo-4-mer
ASS1	P00966_RB363Q	CORE	pathogenic	5.4	1.4	-5.4	-4.9	Homo-4-mer
ASS1	P00966_RB363W	CORE	pathogenic	5.7	0.8	-13.2	-11.8	Homo-4-mer
ASS1	P00966_RB86C	CORE	pathogenic	0.4	1.1	-3.2	-2.6	Homo-4-mer
ASS1	P00966_RB86H	CORE	pathogenic	2.4	0.8	-1.9	-1.8	Homo-4-mer
ASS1	P00966_RC265C	CORE	pathogenic	0.9	1.7	0.1	0	Homo-4-mer
ASS1	P00966_RC265H	CORE	pathogenic	2.6	1.5	-2.6	-1.8	Homo-4-mer
ASS1	P00966_RC304W	CORE	pathogenic	1.2	0.9	-8.7	-7.5	Homo-4-mer
ASS1	P00966_RC307C	CORE	pathogenic	2.6	0.6	1.4	2.6	Homo-4-mer
ASS1	P00966_RC363G	CORE	pathogenic	5.3	1.1	-4.5	-4	Homo-4-mer
ASS1	P00966_RC363L	CORE	pathogenic	1.3	0.9	-6.3	-5.8	Homo-4-mer
ASS1	P00966_RC363Q	CORE	pathogenic	4.2	1.4	-5.4	-4.9	Homo-4-mer
ASS1	P00966_RC363W	CORE	pathogenic	5.2	0.8	-13.2	-11.8	Homo-4-mer
ASS1	P00966_RC86C	CORE	pathogenic	1	1.1	-3.1	-2.6	Homo-4-mer
ASS1	P00966_RC86H	CORE	pathogenic	2.3	0.8	-1.9	-1.8	Homo-4-mer
ASS1	P00966_RD265C	CORE	pathogenic	1	1.7	0.1	0	Homo-4-mer
ASS1	P00966_RD265H	CORE	pathogenic	3.4	1.5	-2.6	-1.8	Homo-4-mer
ASS1	P00966_RD304W	CORE	pathogenic	1.5	0.9	-8.7	-7.5	Homo-4-mer
ASS1	P00966_RD307C	CORE	pathogenic	1.9	0.6	1.4	2.6	Homo-4-mer
ASS1	P00966_RD363G	CORE	pathogenic	5	1.1	-4.5	-4	Homo-4-mer
ASS1	P00966_RD363L	CORE	pathogenic	2.1	0.9	-6.3	-5.8	Homo-4-mer
ASS1	P00966_RD363Q	CORE	pathogenic	4.5	1.4	-5.4	-4.9	Homo-4-mer
ASS1	P00966_RD363W	CORE	pathogenic	6.5	0.8	-13.2	-11.8	Homo-4-mer
ASS1	P00966_RD86C	CORE	pathogenic	1.5	1.1	-3.1	-2.6	Homo-4-mer
ASS1	P00966_RD86H	CORE	pathogenic	2.4	0.8	-1.9	-1.7	Homo-4-mer
ASS1	P00966_TA389I	CORE	pathogenic	2.3	0.7	0.1	0.1	Homo-4-mer

ASS1	P00966_TB389I	CORE	pathogenic	2.1	0.7	0.1	0.1	Homo-4-mer
ASS1	P00966_TC389I	CORE	pathogenic	1.9	0.7	0.1	0.1	Homo-4-mer
ASS1	P00966_TD389I	CORE	pathogenic	2.2	0.7	0.1	0.1	Homo-4-mer
ASS1	P00966_VA269M	CORE	pathogenic	3.1	1.1	-5.3	1.6	Homo-4-mer
ASS1	P00966_VB269M	CORE	pathogenic	3.3	1.1	-5.3	1.6	Homo-4-mer
ASS1	P00966_VC269M	CORE	pathogenic	1.3	1.1	-5.3	1.6	Homo-4-mer
ASS1	P00966_VD269M	CORE	pathogenic	2.5	1.1	-5.3	1.6	Homo-4-mer
ASS1	P00966_YA359D	CORE	pathogenic	3.1	1.2	4.8	3.7	Homo-4-mer
ASS1	P00966_YB359D	CORE	pathogenic	2.9	1.2	4.8	3.7	Homo-4-mer
ASS1	P00966_YC359D	CORE	pathogenic	3.3	1.2	4.8	3.7	Homo-4-mer
ASS1	P00966_YD359D	CORE	pathogenic	3.5	1.2	4.8	3.7	Homo-4-mer
BCKDHA	P12694_GA290R	CORE	pathogenic	1	1.7	2.1	2.9	Homo-2-mer
BCKDHA	P12694_HA247R	CORE	pathogenic	3.2	2.2	12.4	12.6	Homo-2-mer
BCKDHA	P12694-P21953_GA188W	CORE	pathogenic	55.1	1.2	0.1	0.1	Hetero-4-mer
BCKDHA	P12694-P21953_GB188W	CORE	pathogenic	66	1.2	0.1	0.1	Hetero-4-mer
BCKDHA	P12694-P21953_QA190K	CORE	pathogenic	0.8	1.3	8.2	10.4	Hetero-4-mer
BCKDHA	P12694-P21953_QB190K	CORE	pathogenic	2.4	1.3	8.2	10.5	Hetero-4-mer
BCKDHA	P12694-P21953_TA211M	CORE	pathogenic	1.4	0.8	0.3	0.2	Hetero-4-mer
BCKDHA	P12694-P21953_TB211M	CORE	pathogenic	0.9	0.8	0.3	0.2	Hetero-4-mer
BCKDHA	P12694-P21953_VA412M	CORE	pathogenic	4	0.8	0.5	0	Hetero-4-mer
BCKDHA	P12694-P21953_VA412M	CORE	pathogenic	4	0.8	0.5	0	Hetero-4-mer
BCKDHA	P12694-P21953_VA412M	CORE	pathogenic	3.5	0.8	0.5	0	Hetero-4-mer
BCKDHA	P12694-P21953_VA412M	CORE	pathogenic	3.5	0.8	0.5	0	Hetero-4-mer
BCKDHA	P12694-P21953_VB412M	CORE	pathogenic	4.1	0.8	0.5	0	Hetero-4-mer
BCKDHA	P12694-P21953_VB412M	CORE	pathogenic	4.1	0.8	0.5	0	Hetero-4-mer
BCKDHA	P12694-P21953_VB412M	CORE	pathogenic	2.1	0.8	0.5	0	Hetero-4-mer
BCKDHA	P12694-P21953_VB412M	CORE	pathogenic	2.1	0.8	0.5	0	Hetero-4-mer
BCKDHA	P12694-P21953_YA413H	CORE	pathogenic	4.2	2.2	6.2	5.9	Hetero-4-mer
BCKDHA	P12694-P21953_YA438H	CORE	pathogenic	4.8	1.4	2.4	2	Hetero-4-mer
BCKDHA	P12694-P21953_YA438N	CORE	pathogenic	4.9	3	7.1	4.1	Hetero-4-mer
BCKDHA	P12694-P21953_YB413C	CORE	pathogenic	4.3	1.6	4.9	5	Hetero-4-mer
BCKDHA	P12694-P21953_YB413H	CORE	pathogenic	4.3	2.2	6.2	5.9	Hetero-4-mer
BCKDHA	P12694-P21953_YB438H	CORE	pathogenic	3.9	1.4	2.4	1.9	Hetero-4-mer
BCKDHA	P12694-P21953_YB438N	CORE	pathogenic	5.9	3	7	4.1	Hetero-4-mer
BCKDHB	P12694-P21953_GC128E	CORE	pathogenic	15.9	1.9	1.2	5.7	Hetero-4-mer
BCKDHB	P12694-P21953_GC135R	CORE	pathogenic	37	3	11.2	13.7	Hetero-4-mer
BCKDHB	P12694-P21953_GC172W	CORE	pathogenic	21.9	1.7	0	0.5	Hetero-4-mer
BCKDHB	P12694-P21953_GD128E	CORE	pathogenic	12.9	1.9	1.2	5.7	Hetero-4-mer
BCKDHB	P12694-P21953_GD135R	CORE	pathogenic	43.8	3	11.2	13.7	Hetero-4-mer
BCKDHB	P12694-P21953_GD172W	CORE	pathogenic	22	1.7	-0.1	0.5	Hetero-4-mer
BCKDHB	P12694-P21953_LC194F	CORE	pathogenic	6.5	0.3	-2	-1.7	Hetero-4-mer
BCKDHB	P12694-P21953_LD194F	CORE	pathogenic	6.1	0.3	-2	-1.7	Hetero-4-mer
BCKDHB	P12694-P21953_PC123L	CORE	pathogenic	9	1.8	4.1	1.9	Hetero-4-mer
BCKDHB	P12694-P21953_PC123L	CORE	pathogenic	9	1.8	4.1	1.9	Hetero-4-mer
BCKDHB	P12694-P21953_PC123L	CORE	pathogenic	8.9	1.8	4.1	1.9	Hetero-4-mer
BCKDHB	P12694-P21953_PC123L	CORE	pathogenic	8.9	1.8	4.1	1.9	Hetero-4-mer
BCKDHB	P12694-P21953_PD123L	CORE	pathogenic	5.2	1.8	4.1	1.9	Hetero-4-mer
BCKDHB	P12694-P21953_PD123L	CORE	pathogenic	5.2	1.8	4.1	1.9	Hetero-4-mer
BCKDHB	P12694-P21953_PD123L	CORE	pathogenic	5.3	1.8	4.1	1.9	Hetero-4-mer
BCKDHB	P12694-P21953_PD123L	CORE	pathogenic	5.3	1.8	4.1	1.9	Hetero-4-mer
BCKDHB	P12694-P21953_QC159K	CORE	pathogenic	0.6	2.3	-0.8	-0.4	Hetero-4-mer

BCKDHB	P12694-P21953_QD159K	CORE	pathogenic	-0.7	2.3	-0.8	-0.4	Hetero-4-mer
BCKDHB	P12694-P21953_RC168C	CORE	pathogenic	1	0.1	-3	-3.8	Hetero-4-mer
BCKDHB	P12694-P21953_RC168H	CORE	pathogenic	4.2	0.8	-3.9	-5.4	Hetero-4-mer
BCKDHB	P12694-P21953_RC170C	CORE	pathogenic	3	0.7	-1.2	-1.5	Hetero-4-mer
BCKDHB	P12694-P21953_RC170H	CORE	pathogenic	5.4	0.8	-1.4	-1.7	Hetero-4-mer
BCKDHB	P12694-P21953_RD168C	CORE	pathogenic	1.1	0.1	-3	-3.8	Hetero-4-mer
BCKDHB	P12694-P21953_RD168H	CORE	pathogenic	4.4	0.8	-3.9	-5.4	Hetero-4-mer
BCKDHB	P12694-P21953_RD170C	CORE	pathogenic	2.9	0.7	-1.2	-1.5	Hetero-4-mer
BCKDHB	P12694-P21953_RD170H	CORE	pathogenic	5.1	0.8	-1.4	-1.7	Hetero-4-mer
BCKDHB	P21953_ED340K	CORE	pathogenic	2.6	1.7	11.6	14.5	Homo-2-mer
BCKDHB	P21953_HD206R	CORE	pathogenic	0.9	2.8	3.8	3.5	Homo-2-mer
BCKDHB	P21953_HD206Y	CORE	pathogenic	1.9	2.2	0.4	-0.2	Homo-2-mer
BCKDHB	P21953_LD194F	CORE	pathogenic	1.3	0.3	-8.6	-3.6	Homo-2-mer
BCKDHB	P21953_PD356L	CORE	pathogenic	2.2	2	-4	0	Homo-2-mer
BCKDHB	P21953_QD346R	CORE	pathogenic	5.1	1.7	6.9	7	Homo-2-mer
BCKDHB	P21953_SD339L	CORE	pathogenic	3	1.3	1	-0.7	Homo-2-mer
CBS	P35520_AA114V	CORE	pathogenic	-0.7	0.4	-0.3	-0.3	Homo-2-mer
CBS	P35520_LA540Q	CORE	pathogenic	1.2	0.7	0.6	0.4	Homo-2-mer
CBS	P35520_PA88S	CORE	pathogenic	3.3	0.5	0	0.3	Homo-2-mer
CBS	P35520_RA379Q	CORE	pathogenic	1.6	0.5	2	2.7	Homo-2-mer
CBS	P35520_RA379W	CORE	pathogenic	2.1	0.6	-7.7	-6.9	Homo-2-mer
CBS	P35520_SA466L	CORE	pathogenic	2.3	1	-1.3	-1	Homo-2-mer
ETFA	P13804-P38117_RA122K	CORE	pathogenic	2.2	1.3	-0.8	-1.5	Hetero-2-mer
ETFA	P13804-P38117_TA266M	CORE	pathogenic	-2	-0.5	-1.3	-1.5	Hetero-2-mer
ETFB	P13804-P38117_DB128N	CORE	pathogenic	0.4	1.9	2.3	2.8	Hetero-2-mer
FCGR2A	P12318_QA163K	CORE	pathogenic	-0.8	0.1	0.3	0.3	Homo-2-mer
GALT	P07902_AB101D	CORE	pathogenic	2.8	3.8	-4.8	-4.1	Homo-2-mer
GALT	P07902_FB117S	CORE	pathogenic	4.5	3.5	19.1	16.7	Homo-2-mer
GALT	P07902_FB194L	CORE	pathogenic	2.9	2.9	5.9	7	Homo-2-mer
GALT	P07902_HB31N	CORE	pathogenic	3.6	1.2	2.7	2.1	Homo-2-mer
GALT	P07902_HB31R	CORE	pathogenic	6.4	2.4	2.7	2.1	Homo-2-mer
GALT	P07902_IB198M	CORE	pathogenic	0	0.9	0.6	0	Homo-2-mer
GALT	P07902_IB198T	CORE	pathogenic	1.7	0.6	1.8	1.2	Homo-2-mer
GALT	P07902_IB32N	CORE	pathogenic	3.2	1.6	3.1	2	Homo-2-mer
GALT	P07902_LB116I	CORE	pathogenic	2.2	0.4	0.8	1.9	Homo-2-mer
GALT	P07902_LB116P	CORE	pathogenic	3.8	0.5	4.3	2.7	Homo-2-mer
GALT	P07902_LB116V	CORE	pathogenic	3.1	0.5	2.2	2	Homo-2-mer
GALT	P07902_LB342I	CORE	pathogenic	0.5	2.3	2.1	2.4	Homo-2-mer
GALT	P07902_PB36L	CORE	pathogenic	1.8	1.1	0.9	1	Homo-2-mer
GALT	P07902_QB30H	CORE	pathogenic	2.9	1.2	0.7	0.5	Homo-2-mer
GALT	P07902_RB201C	CORE	pathogenic	2.4	0.3	-6.1	-3.1	Homo-2-mer
GALT	P07902_RB201H	CORE	pathogenic	1.9	0	-8.7	-6.8	Homo-2-mer
GALT	P07902_VB337I	CORE	pathogenic	1.4	2.1	1.3	0.8	Homo-2-mer
GALT	P07902_YB209C	CORE	pathogenic	4	1.7	3.9	2.4	Homo-2-mer
GALT	P07902_YB209S	CORE	pathogenic	4.4	1.4	4.3	2.8	Homo-2-mer
GALT	P07902_YB323D	CORE	pathogenic	3.7	3.2	5.2	5.8	Homo-2-mer
GALT	P07902_YB323H	CORE	pathogenic	1.6	1.2	2.9	3.5	Homo-2-mer
GALT	P07902_YB339C	CORE	pathogenic	3.8	1.7	7.1	3.9	Homo-2-mer
GALT	P07902_YB34N	CORE	pathogenic	4.5	1.4	2.2	3.3	Homo-2-mer
GCDH	Q92947_AA349T	CORE	pathogenic	0.3	1.2	0.1	1.6	Homo-4-mer
GCDH	Q92947_AA433E	CORE	pathogenic	5.1	2.5	-0.7	-2.4	Homo-4-mer

GCDH	Q92947_AA433V	CORE	pathogenic	2.1	0.9	2.2	0.2	Homo-4-mer
GCDH	Q92947_AB349T	CORE	pathogenic	0.3	1.2	0.1	1.6	Homo-4-mer
GCDH	Q92947_AB433E	CORE	pathogenic	5.1	2.5	-0.7	-2.4	Homo-4-mer
GCDH	Q92947_AB433V	CORE	pathogenic	1.6	0.9	2.2	0.2	Homo-4-mer
GCDH	Q92947_AC349T	CORE	pathogenic	0.4	1.2	0.1	1.6	Homo-4-mer
GCDH	Q92947_AC433E	CORE	pathogenic	5.8	2.5	-0.8	-2.5	Homo-4-mer
GCDH	Q92947_AC433V	CORE	pathogenic	1.5	0.9	2.2	0.2	Homo-4-mer
GCDH	Q92947_AD349T	CORE	pathogenic	1.5	1.2	0.1	1.6	Homo-4-mer
GCDH	Q92947_AD433E	CORE	pathogenic	7.1	2.5	-0.7	-2.5	Homo-4-mer
GCDH	Q92947_AD433V	CORE	pathogenic	2.3	0.9	2.2	0.2	Homo-4-mer
GCDH	Q92947_EA365K	CORE	pathogenic	0.5	2	-0.8	-1.3	Homo-4-mer
GCDH	Q92947_EB365K	CORE	pathogenic	0.4	2	-0.7	-1.2	Homo-4-mer
GCDH	Q92947_EC365K	CORE	pathogenic	0.8	2	0.8	0.3	Homo-4-mer
GCDH	Q92947_ED365K	CORE	pathogenic	0.6	2	-0.7	-1.2	Homo-4-mer
GCDH	Q92947_GA390A	CORE	pathogenic	3.2	0.8	-0.3	-0.2	Homo-4-mer
GCDH	Q92947_GA390R	CORE	pathogenic	1.8	1.4	3.1	2.9	Homo-4-mer
GCDH	Q92947_GA390V	CORE	pathogenic	4.1	0.9	-0.7	-0.5	Homo-4-mer
GCDH	Q92947_GA390W	CORE	pathogenic	1.2	0.2	-5.5	-5.5	Homo-4-mer
GCDH	Q92947_GA425V	CORE	pathogenic	11.6	1.9	2.9	2.1	Homo-4-mer
GCDH	Q92947_GB390A	CORE	pathogenic	3.2	0.8	-0.3	-0.2	Homo-4-mer
GCDH	Q92947_GB390R	CORE	pathogenic	1.9	1.4	3.1	2.9	Homo-4-mer
GCDH	Q92947_GB390V	CORE	pathogenic	3.9	0.9	-0.7	-0.5	Homo-4-mer
GCDH	Q92947_GB390W	CORE	pathogenic	1	0.2	-5.5	-5.5	Homo-4-mer
GCDH	Q92947_GB425V	CORE	pathogenic	11.9	1.9	2.9	2.1	Homo-4-mer
GCDH	Q92947_GC390A	CORE	pathogenic	3.2	0.8	-0.3	-0.2	Homo-4-mer
GCDH	Q92947_GC390R	CORE	pathogenic	1.5	1.4	3	2.9	Homo-4-mer
GCDH	Q92947_GC390V	CORE	pathogenic	4.1	0.9	-0.7	-0.5	Homo-4-mer
GCDH	Q92947_GC390W	CORE	pathogenic	1.6	0.2	-5.5	-5.5	Homo-4-mer
GCDH	Q92947_GC425V	CORE	pathogenic	11.3	1.9	2.9	2.1	Homo-4-mer
GCDH	Q92947_GD390A	CORE	pathogenic	3.2	0.8	-0.3	-0.2	Homo-4-mer
GCDH	Q92947_GD390R	CORE	pathogenic	2	1.4	3	2.9	Homo-4-mer
GCDH	Q92947_GD390V	CORE	pathogenic	4.1	0.9	-0.7	-0.5	Homo-4-mer
GCDH	Q92947_GD390W	CORE	pathogenic	1.1	0.2	-5.5	-5.5	Homo-4-mer
GCDH	Q92947_GD425V	CORE	pathogenic	12.3	1.9	2.9	2.1	Homo-4-mer
GCDH	Q92947_MA405V	CORE	pathogenic	3.9	1.3	4.2	0.2	Homo-4-mer
GCDH	Q92947_MB405V	CORE	pathogenic	4	1.3	4.2	0.2	Homo-4-mer
GCDH	Q92947_MC405V	CORE	pathogenic	4	1.3	4.2	0.2	Homo-4-mer
GCDH	Q92947_MD405V	CORE	pathogenic	4	1.3	4.2	0.2	Homo-4-mer
GCDH	Q92947_NA392D	CORE	pathogenic	-1.2	1.2	0.4	0.4	Homo-4-mer
GCDH	Q92947_NA392S	CORE	pathogenic	1.2	0.7	1	0.5	Homo-4-mer
GCDH	Q92947_NB392D	CORE	pathogenic	-0.9	1.2	0.4	0.4	Homo-4-mer
GCDH	Q92947_NB392S	CORE	pathogenic	1.2	0.7	1	0.5	Homo-4-mer
GCDH	Q92947_NC392D	CORE	pathogenic	-0.7	1.2	0.4	0.4	Homo-4-mer
GCDH	Q92947_NC392S	CORE	pathogenic	0.3	0.7	1	0.5	Homo-4-mer
GCDH	Q92947_ND392D	CORE	pathogenic	-0.9	1.2	0.4	0.4	Homo-4-mer
GCDH	Q92947_ND392S	CORE	pathogenic	0.4	0.7	1	0.5	Homo-4-mer
GCDH	Q92947_QA321H	CORE	pathogenic	2.7	1.5	0.7	0.6	Homo-4-mer
GCDH	Q92947_QA333E	CORE	pathogenic	2.3	1.1	-1.8	-1.6	Homo-4-mer
GCDH	Q92947_QB321H	CORE	pathogenic	2.4	1.5	0.7	0.6	Homo-4-mer
GCDH	Q92947_QB333E	CORE	pathogenic	2.3	1.1	-1.8	-1.6	Homo-4-mer
GCDH	Q92947_QC321H	CORE	pathogenic	1.9	1.5	0.7	0.6	Homo-4-mer

GCDH	Q92947_QC333E	CORE	pathogenic	2.6	1.1	-1.8	-1.6	Homo-4-mer
GCDH	Q92947_QD321H	CORE	pathogenic	2.1	1.5	0.7	0.6	Homo-4-mer
GCDH	Q92947_QD333E	CORE	pathogenic	2.5	1.1	-1.8	-1.6	Homo-4-mer
GCDH	Q92947_RA313Q	CORE	pathogenic	3.5	0.8	0.2	0.3	Homo-4-mer
GCDH	Q92947_RA313W	CORE	pathogenic	3.8	0.9	0.5	0.9	Homo-4-mer
GCDH	Q92947_RA372K	CORE	pathogenic	-0.1	-0.4	-3.6	-2.8	Homo-4-mer
GCDH	Q92947_RA383C	CORE	pathogenic	1.2	1.1	-1.8	-2.1	Homo-4-mer
GCDH	Q92947_RA383H	CORE	pathogenic	0.8	1.4	-3.3	-3.5	Homo-4-mer
GCDH	Q92947_RA386G	CORE	pathogenic	3.8	0.8	-5.9	-4	Homo-4-mer
GCDH	Q92947_RA386Q	CORE	pathogenic	4.1	-0.1	-6.1	-4.7	Homo-4-mer
GCDH	Q92947_RA88C	CORE	pathogenic	0.5	1	-1.3	-1.6	Homo-4-mer
GCDH	Q92947_RA88C	CORE	pathogenic	0.5	1	-1.3	-1.6	Homo-4-mer
GCDH	Q92947_RA88C	CORE	pathogenic	0.5	1	-1.3	-1.6	Homo-4-mer
GCDH	Q92947_RA88C	CORE	pathogenic	0.5	1	-1.3	-1.6	Homo-4-mer
GCDH	Q92947_RA88H	CORE	pathogenic	1.1	1.1	-3.7	-4.7	Homo-4-mer
GCDH	Q92947_RB313Q	CORE	pathogenic	2.7	0.8	0.2	0.3	Homo-4-mer
GCDH	Q92947_RB313W	CORE	pathogenic	3.1	0.9	0.5	0.9	Homo-4-mer
GCDH	Q92947_RB372K	CORE	pathogenic	1.1	-0.4	-3.6	-2.8	Homo-4-mer
GCDH	Q92947_RB383C	CORE	pathogenic	0.4	1.1	-1.8	-2.1	Homo-4-mer
GCDH	Q92947_RB383H	CORE	pathogenic	0.3	1.4	-3.3	-3.5	Homo-4-mer
GCDH	Q92947_RB386G	CORE	pathogenic	2.5	0.8	-5.9	-4	Homo-4-mer
GCDH	Q92947_RB386Q	CORE	pathogenic	2.1	-0.1	-6.1	-4.7	Homo-4-mer
GCDH	Q92947_RB88C	CORE	pathogenic	0.3	1	-1.3	-1.6	Homo-4-mer
GCDH	Q92947_RB88C	CORE	pathogenic	0.3	1	-1.3	-1.6	Homo-4-mer
GCDH	Q92947_RB88C	CORE	pathogenic	0.3	1	-1.3	-1.6	Homo-4-mer
GCDH	Q92947_RB88C	CORE	pathogenic	0.3	1	-1.3	-1.6	Homo-4-mer
GCDH	Q92947_RB88H	CORE	pathogenic	1.3	1.1	-3.7	-4.7	Homo-4-mer
GCDH	Q92947_RC313Q	CORE	pathogenic	2.8	0.8	0.2	0.3	Homo-4-mer
GCDH	Q92947_RC313W	CORE	pathogenic	1.5	0.9	0.5	0.9	Homo-4-mer
GCDH	Q92947_RC372K	CORE	pathogenic	0.7	-0.4	-3.5	-2.8	Homo-4-mer
GCDH	Q92947_RC383C	CORE	pathogenic	0.8	1.1	-1.8	-2.1	Homo-4-mer
GCDH	Q92947_RC383H	CORE	pathogenic	0.3	1.4	-3.3	-3.5	Homo-4-mer
GCDH	Q92947_RC386G	CORE	pathogenic	2.6	0.8	-5.9	-3.9	Homo-4-mer
GCDH	Q92947_RC386Q	CORE	pathogenic	2.1	-0.1	-6.1	-4.7	Homo-4-mer
GCDH	Q92947_RC88C	CORE	pathogenic	0.6	1	-1.3	-1.6	Homo-4-mer
GCDH	Q92947_RC88C	CORE	pathogenic	0.6	1	-1.3	-1.6	Homo-4-mer
GCDH	Q92947_RC88C	CORE	pathogenic	0.5	1	-1.3	-1.6	Homo-4-mer
GCDH	Q92947_RC88C	CORE	pathogenic	0.5	1	-1.3	-1.6	Homo-4-mer
GCDH	Q92947_RC88H	CORE	pathogenic	1.2	1.1	-3.7	-4.7	Homo-4-mer
GCDH	Q92947_RD313Q	CORE	pathogenic	2.6	0.8	0.2	0.3	Homo-4-mer
GCDH	Q92947_RD313W	CORE	pathogenic	0.9	0.9	0.6	0.9	Homo-4-mer
GCDH	Q92947_RD372K	CORE	pathogenic	0.9	-0.4	-3.6	-2.8	Homo-4-mer
GCDH	Q92947_RD383C	CORE	pathogenic	1.4	1.1	-1.8	-2.1	Homo-4-mer
GCDH	Q92947_RD383H	CORE	pathogenic	0.5	1.4	-3.3	-3.5	Homo-4-mer
GCDH	Q92947_RD386G	CORE	pathogenic	4.3	0.8	-5.9	-3.9	Homo-4-mer
GCDH	Q92947_RD386Q	CORE	pathogenic	2.4	-0.1	-6.1	-4.7	Homo-4-mer
GCDH	Q92947_RD88C	CORE	pathogenic	0.4	1	-1.3	-1.6	Homo-4-mer
GCDH	Q92947_RD88C	CORE	pathogenic	0.4	1	-1.3	-1.6	Homo-4-mer
GCDH	Q92947_RD88C	CORE	pathogenic	0.6	1	-1.3	-1.6	Homo-4-mer
GCDH	Q92947_RD88C	CORE	pathogenic	0.6	1	-1.3	-1.6	Homo-4-mer
GCDH	Q92947_RD88H	CORE	pathogenic	1.2	1.1	-3.7	-4.7	Homo-4-mer

GCDH	Q92947_TA341I	CORE	pathogenic	-1.5	0.6	0.3	-1.6	Homo-4-mer
GCDH	Q92947_TA341P	CORE	pathogenic	5.2	0.5	-0.8	-0.9	Homo-4-mer
GCDH	Q92947_TA416I	CORE	pathogenic	1.2	0.2	0.2	0.2	Homo-4-mer
GCDH	Q92947_TA429M	CORE	pathogenic	-0.7	-0.2	0.2	0.4	Homo-4-mer
GCDH	Q92947_TB341I	CORE	pathogenic	-2	0.6	0.3	-1.6	Homo-4-mer
GCDH	Q92947_TB341P	CORE	pathogenic	4.9	0.5	-0.8	-0.9	Homo-4-mer
GCDH	Q92947_TB416I	CORE	pathogenic	2	0.2	0.2	0.2	Homo-4-mer
GCDH	Q92947_TB429M	CORE	pathogenic	0.1	-0.2	0.2	0.4	Homo-4-mer
GCDH	Q92947_TC341I	CORE	pathogenic	-0.9	0.6	0.3	-1.6	Homo-4-mer
GCDH	Q92947_TC341P	CORE	pathogenic	4.6	0.5	-0.8	-0.9	Homo-4-mer
GCDH	Q92947_TC416I	CORE	pathogenic	1.2	0.2	0.2	0.2	Homo-4-mer
GCDH	Q92947_TC429M	CORE	pathogenic	-0.7	-0.2	0.2	0.4	Homo-4-mer
GCDH	Q92947_TD341I	CORE	pathogenic	-0.3	0.6	0.3	-1.6	Homo-4-mer
GCDH	Q92947_TD341P	CORE	pathogenic	5.2	0.5	-0.8	-0.9	Homo-4-mer
GCDH	Q92947_TD416I	CORE	pathogenic	1.2	0.2	0.2	0.2	Homo-4-mer
GCDH	Q92947_TD429M	CORE	pathogenic	-0.7	-0.2	0.2	0.4	Homo-4-mer
GCDH	Q92947_WA50C	CORE	pathogenic	3.1	3	1.7	1.8	Homo-4-mer
GCDH	Q92947_WA50R	CORE	pathogenic	2.2	4.3	-0.4	0.4	Homo-4-mer
GCDH	Q92947_WB50C	CORE	pathogenic	3.1	3	1.7	1.8	Homo-4-mer
GCDH	Q92947_WB50R	CORE	pathogenic	1.5	4.3	-0.4	0.4	Homo-4-mer
GCDH	Q92947_WC50C	CORE	pathogenic	3.3	3	1.7	1.8	Homo-4-mer
GCDH	Q92947_WC50R	CORE	pathogenic	1.2	4.3	-0.4	0.4	Homo-4-mer
GCDH	Q92947_WD50C	CORE	pathogenic	3.1	3	1.7	1.8	Homo-4-mer
GCDH	Q92947_WD50R	CORE	pathogenic	1.2	4.3	-0.4	0.4	Homo-4-mer
HADHA	P40939-HADHB_RG235W	CORE	pathogenic	2.6	1	-1.7	-2.1	Hetero-3-mer
HADHB	P55084_RA117G	CORE	pathogenic	2.5	-0.5	0.9	0.3	Homo-2-mer
HADHB	P55084_TA133P	CORE	pathogenic	6.2	2	-0.3	-0.3	Homo-2-mer
HBB	P68871_NB140D	CORE	pathogenic	1.8	0.7	0.3	0.4	Homo-2-mer
HBB	P68871_NB140S	CORE	pathogenic	0.7	0.3	0.2	0.3	Homo-2-mer
HBB	P68871_NB140T	CORE	pathogenic	-0.3	0.4	-1	-0.3	Homo-2-mer
HBB	P68871_NB140Y	CORE	pathogenic	-0.3	0.6	-4.3	-4.3	Homo-2-mer
HBB	P68871-HBA_AB116D	CORE	pathogenic	6.8	2.4	-2	-1	Hetero-4-mer
HBB	P68871-HBA_AB129D	CORE	pathogenic	3.9	1.9	1.3	1	Hetero-4-mer
HBB	P68871-HBA_AB129P	CORE	pathogenic	5.3	2.5	0.5	0.5	Hetero-4-mer
HBB	P68871-HBA_AB129V	CORE	pathogenic	1.6	2.5	0.2	0	Hetero-4-mer
HBB	P68871-HBA_AD116D	CORE	pathogenic	8.1	2.3	-4.2	-3.2	Hetero-4-mer
HBB	P68871-HBA_AD116P	CORE	pathogenic	7.1	2.6	0	0	Hetero-4-mer
HBB	P68871-HBA_AD116V	CORE	pathogenic	5.4	2.6	-1.2	0.3	Hetero-4-mer
HBB	P68871-HBA_AD129D	CORE	pathogenic	3.8	1.3	0.9	0.7	Hetero-4-mer
HBB	P68871-HBA_AD129P	CORE	pathogenic	4.8	2.6	0.8	0.8	Hetero-4-mer
HBB	P68871-HBA_AD129V	CORE	pathogenic	1.2	2.6	0.4	0.2	Hetero-4-mer
HBB	P68871-HBA_CB113F	CORE	pathogenic	0.5	1.1	-7.8	-7.9	Hetero-4-mer
HBB	P68871-HBA_CB113R	CORE	pathogenic	-0.1	2.5	6.9	3	Hetero-4-mer
HBB	P68871-HBA_CB113W	CORE	pathogenic	1.6	1.5	-3.6	-3.3	Hetero-4-mer
HBB	P68871-HBA_CB113Y	CORE	pathogenic	6.1	1.2	-2.1	-2.3	Hetero-4-mer
HBB	P68871-HBA_CD113F	CORE	pathogenic	0.6	0.9	-8	-8	Hetero-4-mer
HBB	P68871-HBA_CD113R	CORE	pathogenic	-1.5	2.4	7.2	3.3	Hetero-4-mer
HBB	P68871-HBA_CD113W	CORE	pathogenic	0.3	1.5	-4.2	-4	Hetero-4-mer
HBB	P68871-HBA_CD113Y	CORE	pathogenic	2.7	1.2	-2.5	-2.5	Hetero-4-mer
HBB	P68871-HBA_DB100A	CORE	pathogenic	5.5	2.7	1.8	1.8	Hetero-4-mer
HBB	P68871-HBA_DB100E	CORE	pathogenic	3.7	1.7	0.2	0.2	Hetero-4-mer

HBB	P68871-HBA_DB100G	CORE	pathogenic	6.4	2.7	2.6	2.5	Hetero-4-mer
HBB	P68871-HBA_DB100H	CORE	pathogenic	6.4	2.1	-0.7	-1.1	Hetero-4-mer
HBB	P68871-HBA_DB100N	CORE	pathogenic	3.3	2.3	6.3	6.1	Hetero-4-mer
HBB	P68871-HBA_DB100V	CORE	pathogenic	6.3	2.6	1.1	0.9	Hetero-4-mer
HBB	P68871-HBA_DB100Y	CORE	pathogenic	5.1	2.9	-4.9	-4.5	Hetero-4-mer
HBB	P68871-HBA_DD100A	CORE	pathogenic	5.2	2.2	1.4	1.3	Hetero-4-mer
HBB	P68871-HBA_DD100E	CORE	pathogenic	3.2	1.4	-0.3	-0.4	Hetero-4-mer
HBB	P68871-HBA_DD100G	CORE	pathogenic	6	2.3	2.2	2.1	Hetero-4-mer
HBB	P68871-HBA_DD100H	CORE	pathogenic	4	1.7	-2.2	-2.6	Hetero-4-mer
HBB	P68871-HBA_DD100N	CORE	pathogenic	3.1	1.9	5.5	5.3	Hetero-4-mer
HBB	P68871-HBA_DD100V	CORE	pathogenic	5.4	2.1	0.7	0.5	Hetero-4-mer
HBB	P68871-HBA_DD100Y	CORE	pathogenic	4	2.5	-6.7	-6.3	Hetero-4-mer
HBB	P68871-HBA_EB102A	CORE	pathogenic	-0.4	1.3	2.8	2.3	Hetero-4-mer
HBB	P68871-HBA_EB102D	CORE	pathogenic	1.3	1	3.2	0.8	Hetero-4-mer
HBB	P68871-HBA_EB102G	CORE	pathogenic	0.8	1.3	3.7	2.8	Hetero-4-mer
HBB	P68871-HBA_EB102K	CORE	pathogenic	-1	1.5	8.9	8	Hetero-4-mer
HBB	P68871-HBA_EB102Q	CORE	pathogenic	-0.9	1.3	1	1.1	Hetero-4-mer
HBB	P68871-HBA_ED102A	CORE	pathogenic	-0.2	1.6	1.9	1.4	Hetero-4-mer
HBB	P68871-HBA_ED102D	CORE	pathogenic	1.8	1.4	2.4	-0.1	Hetero-4-mer
HBB	P68871-HBA_ED102G	CORE	pathogenic	0.9	1.6	2.8	1.9	Hetero-4-mer
HBB	P68871-HBA_ED102K	CORE	pathogenic	-0.7	1.9	3.9	4.2	Hetero-4-mer
HBB	P68871-HBA_ED102Q	CORE	pathogenic	-0.7	1.7	0	0.1	Hetero-4-mer
HBB	P68871-HBA_GB120A	CORE	pathogenic	2.8	1.4	-0.2	-0.3	Hetero-4-mer
HBB	P68871-HBA_GB120D	CORE	pathogenic	4	2.2	1.3	1.7	Hetero-4-mer
HBB	P68871-HBA_GD120A	CORE	pathogenic	2.8	0.8	-0.1	-0.2	Hetero-4-mer
HBB	P68871-HBA_GD120D	CORE	pathogenic	3.9	1.7	1	1.3	Hetero-4-mer
HBB	P68871-HBA_HB117L	CORE	pathogenic	-1.5	1.4	0.2	0.3	Hetero-4-mer
HBB	P68871-HBA_HB117P	CORE	pathogenic	7.4	1.4	3	3.8	Hetero-4-mer
HBB	P68871-HBA_HB117Y	CORE	pathogenic	-0.9	0	-0.9	-1.1	Hetero-4-mer
HBB	P68871-HBA_HD117L	CORE	pathogenic	-2.1	1.3	-1	-0.9	Hetero-4-mer
HBB	P68871-HBA_HD117P	CORE	pathogenic	7.3	1.3	0.6	1.2	Hetero-4-mer
HBB	P68871-HBA_HD117Y	CORE	pathogenic	0.2	1	-3.9	-3.8	Hetero-4-mer
HBB	P68871-HBA_NB109D	CORE	pathogenic	0.7	1.4	-1.1	-1.4	Hetero-4-mer
HBB	P68871-HBA_NB109I	CORE	pathogenic	-1.2	1.8	0.9	0.6	Hetero-4-mer
HBB	P68871-HBA_NB109K	CORE	pathogenic	-0.8	2	5.7	6	Hetero-4-mer
HBB	P68871-HBA_ND109D	CORE	pathogenic	0.8	0.6	-1.6	-1.6	Hetero-4-mer
HBB	P68871-HBA_ND109I	CORE	pathogenic	-0.8	0.6	0.5	0.5	Hetero-4-mer
HBB	P68871-HBA_ND109K	CORE	pathogenic	-0.5	1	4	4.5	Hetero-4-mer
HBB	P68871-HBA_PB125L	CORE	pathogenic	3	1.6	1.4	1.6	Hetero-4-mer
HBB	P68871-HBA_PB125Q	CORE	pathogenic	3.3	1.6	3.8	4	Hetero-4-mer
HBB	P68871-HBA_PB125R	CORE	pathogenic	5.1	2.1	7.4	7.6	Hetero-4-mer
HBB	P68871-HBA_PB37A	CORE	pathogenic	2.4	0.9	0.4	0.6	Hetero-4-mer
HBB	P68871-HBA_PB37H	CORE	pathogenic	5.8	1.1	1.7	1.9	Hetero-4-mer
HBB	P68871-HBA_PB37R	CORE	pathogenic	2	2.5	6.2	7.1	Hetero-4-mer
HBB	P68871-HBA_PB37S	CORE	pathogenic	2.5	0.9	2.1	1.6	Hetero-4-mer
HBB	P68871-HBA_PB37T	CORE	pathogenic	3.1	1.1	1.9	1.4	Hetero-4-mer
HBB	P68871-HBA_PD125L	CORE	pathogenic	2.5	2.4	1.6	2.9	Hetero-4-mer
HBB	P68871-HBA_PD125Q	CORE	pathogenic	2.7	1.9	5.2	6.2	Hetero-4-mer
HBB	P68871-HBA_PD125R	CORE	pathogenic	3.6	2.5	2.8	3.5	Hetero-4-mer
HBB	P68871-HBA_PD37A	CORE	pathogenic	3.1	0.7	0.4	0.5	Hetero-4-mer
HBB	P68871-HBA_PD37H	CORE	pathogenic	6.3	1	-1.5	-1.3	Hetero-4-mer

HBB	P68871-HBA_PD37R	CORE	pathogenic	2.3	2.4	7.2	8	Hetero-4-mer
HBB	P68871-HBA_PD37S	CORE	pathogenic	3.2	0.8	2.1	1.7	Hetero-4-mer
HBB	P68871-HBA_PD37T	CORE	pathogenic	3.4	1	1.5	1.4	Hetero-4-mer
HBB	P68871-HBA_QB128E	CORE	pathogenic	0.4	2.3	-2.2	-1.8	Hetero-4-mer
HBB	P68871-HBA_QB128K	CORE	pathogenic	1	3.3	10.2	12.7	Hetero-4-mer
HBB	P68871-HBA_QB128P	CORE	pathogenic	6.2	3	3	2.2	Hetero-4-mer
HBB	P68871-HBA_QB128R	CORE	pathogenic	6.1	3.2	14.1	11	Hetero-4-mer
HBB	P68871-HBA_QB132K	CORE	pathogenic	1.1	3.5	7	7.1	Hetero-4-mer
HBB	P68871-HBA_QB132P	CORE	pathogenic	3.8	2.9	-0.7	1.3	Hetero-4-mer
HBB	P68871-HBA_QB132R	CORE	pathogenic	-0.2	3.5	-3.7	2.8	Hetero-4-mer
HBB	P68871-HBA_QD128K	CORE	pathogenic	0.9	2.6	10.1	12.5	Hetero-4-mer
HBB	P68871-HBA_QD128P	CORE	pathogenic	7.2	2.2	3.1	2.3	Hetero-4-mer
HBB	P68871-HBA_QD128R	CORE	pathogenic	7.6	2.9	9.9	11.2	Hetero-4-mer
HBB	P68871-HBA_QD132K	CORE	pathogenic	0.7	1.6	7.1	6.9	Hetero-4-mer
HBB	P68871-HBA_QD132P	CORE	pathogenic	3.7	1.1	-0.6	1.2	Hetero-4-mer
HBB	P68871-HBA_QD132R	CORE	pathogenic	-0.2	1.5	-3.6	2.7	Hetero-4-mer
HBB	P68871-HBA_RB31S	CORE	pathogenic	4.1	1.8	-1.5	-2.4	Hetero-4-mer
HBB	P68871-HBA_RB41K	CORE	pathogenic	0.5	2	4.7	4.6	Hetero-4-mer
HBB	P68871-HBA_RD31S	CORE	pathogenic	3.1	0.6	-1.2	-2.6	Hetero-4-mer
HBB	P68871-HBA_RD41K	CORE	pathogenic	0.8	2.2	5	4.6	Hetero-4-mer
HBB	P68871-HBA_VB35D	CORE	pathogenic	4.7	1.9	-4.7	-0.2	Hetero-4-mer
HBB	P68871-HBA_VB35F	CORE	pathogenic	14.9	1	-6.7	-7.4	Hetero-4-mer
HBB	P68871-HBA_VB35I	CORE	pathogenic	0.3	2	-3.6	1	Hetero-4-mer
HBB	P68871-HBA_VB35L	CORE	pathogenic	1.3	2	-5.2	-0.6	Hetero-4-mer
HBB	P68871-HBA_VD35D	CORE	pathogenic	5.3	1.9	-4.7	-0.3	Hetero-4-mer
HBB	P68871-HBA_VD35F	CORE	pathogenic	17.9	1	-3	-0.3	Hetero-4-mer
HBB	P68871-HBA_VD35I	CORE	pathogenic	1.2	2	0.7	0.8	Hetero-4-mer
HBB	P68871-HBA_VD35L	CORE	pathogenic	2.2	2	-5.2	-0.8	Hetero-4-mer
HBB	P68871-HBA_WB38C	CORE	pathogenic	5.9	3.7	7.6	9.2	Hetero-4-mer
HBB	P68871-HBA_WB38R	CORE	pathogenic	3.2	4.2	12.2	12.4	Hetero-4-mer
HBB	P68871-HBA_WB38S	CORE	pathogenic	7.1	4.6	9.1	10.7	Hetero-4-mer
HBB	P68871-HBA_WD38C	CORE	pathogenic	6.1	3.6	7.5	9.2	Hetero-4-mer
HBB	P68871-HBA_WD38R	CORE	pathogenic	6.7	4	12	12.3	Hetero-4-mer
HBB	P68871-HBA_WD38S	CORE	pathogenic	7.3	4.4	9	10.6	Hetero-4-mer
HBB	P68871-HBA_YB146C	CORE	pathogenic	4.2	2.5	2.5	1.6	Hetero-4-mer
HBB	P68871-HBA_YB146D	CORE	pathogenic	6.8	2.5	0.8	-0.2	Hetero-4-mer
HBB	P68871-HBA_YB146H	CORE	pathogenic	3.3	1	1.6	0.7	Hetero-4-mer
HBB	P68871-HBA_YB146N	CORE	pathogenic	4.4	2	2.4	1.4	Hetero-4-mer
HBB	P68871-HBA_YD146C	CORE	pathogenic	4.2	2.2	2.7	1.7	Hetero-4-mer
HBB	P68871-HBA_YD146D	CORE	pathogenic	6.9	2.2	1	0	Hetero-4-mer
HBB	P68871-HBA_YD146H	CORE	pathogenic	3.5	1	1.8	0.8	Hetero-4-mer
HBB	P68871-HBA_YD146N	CORE	pathogenic	4.5	2.1	2.5	1.5	Hetero-4-mer
HBB	P68871-HBAZ_AB116D	CORE	pathogenic	5.3	2	-1.8	-0.7	Hetero-4-mer
HBB	P68871-HBAZ_AB129D	CORE	pathogenic	3.2	1.4	1.2	1	Hetero-4-mer
HBB	P68871-HBAZ_AB129P	CORE	pathogenic	4.2	2.1	0.2	0.2	Hetero-4-mer
HBB	P68871-HBAZ_AB129V	CORE	pathogenic	2	2.1	-0.4	-0.6	Hetero-4-mer
HBB	P68871-HBAZ_AD116D	CORE	pathogenic	5.8	1.9	-1.4	-1.3	Hetero-4-mer
HBB	P68871-HBAZ_AD116P	CORE	pathogenic	8.3	2.7	-0.1	-0.1	Hetero-4-mer
HBB	P68871-HBAZ_AD116V	CORE	pathogenic	4.8	2.7	-0.3	-0.3	Hetero-4-mer
HBB	P68871-HBAZ_AD129D	CORE	pathogenic	1.8	1.4	1.2	1	Hetero-4-mer
HBB	P68871-HBAZ_AD129P	CORE	pathogenic	4.3	2	0.2	0.2	Hetero-4-mer

HBB	P68871-HBAZ_AD129V	CORE	pathogenic	1.5	2	-0.4	-0.6	Hetero-4-mer
HBB	P68871-HBAZ_CB113F	CORE	pathogenic	2.5	1	-7.4	-7.5	Hetero-4-mer
HBB	P68871-HBAZ_CB113R	CORE	pathogenic	0.3	2.5	6.7	2.8	Hetero-4-mer
HBB	P68871-HBAZ_CB113W	CORE	pathogenic	3.8	1.5	-2.9	-2.4	Hetero-4-mer
HBB	P68871-HBAZ_CB113Y	CORE	pathogenic	7.3	1.2	-1.7	-1.8	Hetero-4-mer
HBB	P68871-HBAZ_CD113F	CORE	pathogenic	1.9	0.9	-7.7	-7.8	Hetero-4-mer
HBB	P68871-HBAZ_CD113R	CORE	pathogenic	1.5	2.3	7	3.1	Hetero-4-mer
HBB	P68871-HBAZ_CD113W	CORE	pathogenic	5.6	1.4	-0.7	-2.5	Hetero-4-mer
HBB	P68871-HBAZ_CD113Y	CORE	pathogenic	2	1.1	-2.1	-2.2	Hetero-4-mer
HBB	P68871-HBAZ_DB100A	CORE	pathogenic	2.1	2.2	2.1	2	Hetero-4-mer
HBB	P68871-HBAZ_DB100E	CORE	pathogenic	2.8	1.3	-0.4	-0.3	Hetero-4-mer
HBB	P68871-HBAZ_DB100G	CORE	pathogenic	3.5	2.3	2.8	2.6	Hetero-4-mer
HBB	P68871-HBAZ_DB100H	CORE	pathogenic	1.9	1.5	4.4	4.4	Hetero-4-mer
HBB	P68871-HBAZ_DB100N	CORE	pathogenic	1.5	2.5	4.1	4	Hetero-4-mer
HBB	P68871-HBAZ_DB100V	CORE	pathogenic	1.7	1.9	1.3	1.5	Hetero-4-mer
HBB	P68871-HBAZ_DB100Y	CORE	pathogenic	4.9	2.5	-7	-6.3	Hetero-4-mer
HBB	P68871-HBAZ_DD100A	CORE	pathogenic	3.5	2.1	1.2	1.1	Hetero-4-mer
HBB	P68871-HBAZ_DD100E	CORE	pathogenic	5.2	1.4	-1.2	-1.1	Hetero-4-mer
HBB	P68871-HBAZ_DD100G	CORE	pathogenic	3	2.3	1.9	1.7	Hetero-4-mer
HBB	P68871-HBAZ_DD100H	CORE	pathogenic	6.6	1.5	4.1	4.5	Hetero-4-mer
HBB	P68871-HBAZ_DD100N	CORE	pathogenic	4	2.6	4.4	4.3	Hetero-4-mer
HBB	P68871-HBAZ_DD100V	CORE	pathogenic	5.5	1.9	0.4	0.6	Hetero-4-mer
HBB	P68871-HBAZ_DD100Y	CORE	pathogenic	6	2.5	-7.2	-6.2	Hetero-4-mer
HBB	P68871-HBAZ_GB120A	CORE	pathogenic	2.3	1.2	-0.1	-0.2	Hetero-4-mer
HBB	P68871-HBAZ_GB120D	CORE	pathogenic	3.9	1.1	1	1.2	Hetero-4-mer
HBB	P68871-HBAZ_GD120A	CORE	pathogenic	2.3	1.3	-0.1	-0.2	Hetero-4-mer
HBB	P68871-HBAZ_GD120D	CORE	pathogenic	5.1	1.1	1.2	1.4	Hetero-4-mer
HBB	P68871-HBAZ_HB117L	CORE	pathogenic	-1.3	1.5	-0.6	-0.5	Hetero-4-mer
HBB	P68871-HBAZ_HB117P	CORE	pathogenic	9.7	1.5	1.2	2	Hetero-4-mer
HBB	P68871-HBAZ_HB117Y	CORE	pathogenic	-0.5	0	-2.7	-2.7	Hetero-4-mer
HBB	P68871-HBAZ_HD117L	CORE	pathogenic	-1.6	1.3	-0.3	-0.2	Hetero-4-mer
HBB	P68871-HBAZ_HD117P	CORE	pathogenic	9.2	1.3	1.3	2.1	Hetero-4-mer
HBB	P68871-HBAZ_HD117Y	CORE	pathogenic	0	-0.1	-2.2	-2.1	Hetero-4-mer
HBB	P68871-HBAZ_NB109D	CORE	pathogenic	0	0.4	-0.8	-0.9	Hetero-4-mer
HBB	P68871-HBAZ_NB109I	CORE	pathogenic	-0.5	0.6	1	0.8	Hetero-4-mer
HBB	P68871-HBAZ_NB109K	CORE	pathogenic	-0.6	1.1	4.1	4.5	Hetero-4-mer
HBB	P68871-HBAZ_ND109D	CORE	pathogenic	0.2	0.3	-0.8	-0.8	Hetero-4-mer
HBB	P68871-HBAZ_ND109I	CORE	pathogenic	-0.4	0.3	0.9	0.7	Hetero-4-mer
HBB	P68871-HBAZ_ND109K	CORE	pathogenic	-0.8	0.6	1.7	1.8	Hetero-4-mer
HBB	P68871-HBAZ_PB101A	CORE	pathogenic	3.3	1.1	1.3	0.6	Hetero-4-mer
HBB	P68871-HBAZ_PB101L	CORE	pathogenic	2.8	1.2	1.2	0.4	Hetero-4-mer
HBB	P68871-HBAZ_PB101R	CORE	pathogenic	3.9	1	6.1	4.8	Hetero-4-mer
HBB	P68871-HBAZ_PB125L	CORE	pathogenic	4.3	1.3	0.6	0.8	Hetero-4-mer
HBB	P68871-HBAZ_PB125Q	CORE	pathogenic	2.8	0.6	2	2.2	Hetero-4-mer
HBB	P68871-HBAZ_PB125R	CORE	pathogenic	3.6	2.3	10	10.9	Hetero-4-mer
HBB	P68871-HBAZ_PD101A	CORE	pathogenic	3	1.1	1.4	0.6	Hetero-4-mer
HBB	P68871-HBAZ_PD101L	CORE	pathogenic	2.2	1.3	1.3	0.5	Hetero-4-mer
HBB	P68871-HBAZ_PD101R	CORE	pathogenic	3.3	0.9	4.2	3.3	Hetero-4-mer
HBB	P68871-HBAZ_PD125L	CORE	pathogenic	4	1.3	-0.2	0.9	Hetero-4-mer
HBB	P68871-HBAZ_PD125Q	CORE	pathogenic	2.7	0.6	1.2	2.2	Hetero-4-mer
HBB	P68871-HBAZ_PD125R	CORE	pathogenic	4.2	2.3	8.4	9.4	Hetero-4-mer

HBB	P68871-HBAZ_QB128E	CORE	pathogenic	-0.3	1.4	-1.3	-0.9	Hetero-4-mer
HBB	P68871-HBAZ_QB128K	CORE	pathogenic	0.7	2.5	10.5	12.9	Hetero-4-mer
HBB	P68871-HBAZ_QB128P	CORE	pathogenic	7.4	1.9	3.8	3	Hetero-4-mer
HBB	P68871-HBAZ_QB128R	CORE	pathogenic	4.8	2.6	7.2	11	Hetero-4-mer
HBB	P68871-HBAZ_QB132K	CORE	pathogenic	0	1.3	0.4	4.1	Hetero-4-mer
HBB	P68871-HBAZ_QB132P	CORE	pathogenic	3.9	0.7	0.3	2.1	Hetero-4-mer
HBB	P68871-HBAZ_QB132R	CORE	pathogenic	-0.7	1.2	-1.3	2.5	Hetero-4-mer
HBB	P68871-HBAZ_QD128K	CORE	pathogenic	0.3	2.4	10.1	13.1	Hetero-4-mer
HBB	P68871-HBAZ_QD128P	CORE	pathogenic	7	1.8	3.4	3.1	Hetero-4-mer
HBB	P68871-HBAZ_QD128R	CORE	pathogenic	1.4	2.5	6.8	9.6	Hetero-4-mer
HBB	P68871-HBAZ_QD132K	CORE	pathogenic	0.2	1.5	1	4.7	Hetero-4-mer
HBB	P68871-HBAZ_QD132P	CORE	pathogenic	3	0.9	0.3	2.2	Hetero-4-mer
HBB	P68871-HBAZ_QD132R	CORE	pathogenic	0.4	1.3	-1.2	2.6	Hetero-4-mer
HBB	P68871-HBAZ_RB31S	CORE	pathogenic	2.8	0.7	-2.3	-1.2	Hetero-4-mer
HBB	P68871-HBAZ_RD31S	CORE	pathogenic	2.9	1	-1.9	-0.6	Hetero-4-mer
HBB	P68871-HBAZ_VB35D	CORE	pathogenic	5.3	1	-4.7	0.4	Hetero-4-mer
HBB	P68871-HBAZ_VB35F	CORE	pathogenic	17.4	0.7	-4	-0.9	Hetero-4-mer
HBB	P68871-HBAZ_VB35I	CORE	pathogenic	-0.1	1.5	0.5	1.3	Hetero-4-mer
HBB	P68871-HBAZ_VB35L	CORE	pathogenic	2.1	1.5	-2.1	-1.4	Hetero-4-mer
HBB	P68871-HBAZ_VD35D	CORE	pathogenic	5.9	1.1	-4.6	0.6	Hetero-4-mer
HBB	P68871-HBAZ_VD35F	CORE	pathogenic	17.6	0.6	-3.7	-0.6	Hetero-4-mer
HBB	P68871-HBAZ_VD35I	CORE	pathogenic	1.1	1.4	0.5	1.2	Hetero-4-mer
HBB	P68871-HBAZ_VD35L	CORE	pathogenic	4	1.4	-6	-1.3	Hetero-4-mer
HBB	P68871-HBAZ_WB38C	CORE	pathogenic	4.6	3.3	5.6	8	Hetero-4-mer
HBB	P68871-HBAZ_WB38R	CORE	pathogenic	3.8	4.1	17.4	17.7	Hetero-4-mer
HBB	P68871-HBAZ_WB38S	CORE	pathogenic	5.4	3.9	6.5	8.9	Hetero-4-mer
HBB	P68871-HBAZ_WD38C	CORE	pathogenic	4.1	3.4	8	8.3	Hetero-4-mer
HBB	P68871-HBAZ_WD38R	CORE	pathogenic	5.3	3.7	15.1	15.3	Hetero-4-mer
HBB	P68871-HBAZ_WD38S	CORE	pathogenic	5.3	3.9	9.5	9.9	Hetero-4-mer
HBB	P68871-HBAZ_YB146C	CORE	pathogenic	4.5	1.7	2.7	1.6	Hetero-4-mer
HBB	P68871-HBAZ_YB146D	CORE	pathogenic	7.2	1.9	1.4	0.4	Hetero-4-mer
HBB	P68871-HBAZ_YB146H	CORE	pathogenic	3.1	1.2	3.3	2.3	Hetero-4-mer
HBB	P68871-HBAZ_YB146N	CORE	pathogenic	4.7	2.2	2.8	1.7	Hetero-4-mer
HBB	P68871-HBAZ_YD146C	CORE	pathogenic	4.3	1.7	2.7	1.7	Hetero-4-mer
HBB	P68871-HBAZ_YD146D	CORE	pathogenic	7	1.7	1.1	0.1	Hetero-4-mer
HBB	P68871-HBAZ_YD146H	CORE	pathogenic	3.5	1.1	2.9	1.9	Hetero-4-mer
HBB	P68871-HBAZ_YD146N	CORE	pathogenic	4	2	3	1.9	Hetero-4-mer
HBB	P68871-HP_DL100A	CORE	pathogenic	2.1	1.5	3.2	3.9	Hetero-2-mer
HBB	P68871-HP_DL100E	CORE	pathogenic	2.2	1.3	1.1	0.5	Hetero-2-mer
HBB	P68871-HP_DL100G	CORE	pathogenic	3	1.6	4.1	4.6	Hetero-2-mer
HBB	P68871-HP_DL100H	CORE	pathogenic	1.8	0.7	1.4	1.3	Hetero-2-mer
HBB	P68871-HP_DL100N	CORE	pathogenic	1.9	1.8	6.8	6.8	Hetero-2-mer
HBB	P68871-HP_DL100V	CORE	pathogenic	2.1	1.2	3.1	3.6	Hetero-2-mer
HBB	P68871-HP_DL100Y	CORE	pathogenic	2.9	1.8	0.6	0.7	Hetero-2-mer
HBB	P68871-HP_EL102A	CORE	pathogenic	2	2.2	3.6	4.7	Hetero-2-mer
HBB	P68871-HP_EL102D	CORE	pathogenic	2.2	0.8	0.3	0.4	Hetero-2-mer
HBB	P68871-HP_EL102G	CORE	pathogenic	2.3	2.3	3.4	5.2	Hetero-2-mer
HBB	P68871-HP_EL102K	CORE	pathogenic	2.9	2.1	25	20.8	Hetero-2-mer
HBB	P68871-HP_EL102Q	CORE	pathogenic	1.6	1.7	7.1	6.3	Hetero-2-mer
HBB	P68871-HP_PL101A	CORE	pathogenic	2.8	0.8	3	0.6	Hetero-2-mer
HBB	P68871-HP_PL101L	CORE	pathogenic	3	0.9	-0.8	0.4	Hetero-2-mer

HBB	P68871-HP_PL101R	CORE	pathogenic	3.7	0.6	1.6	0.4	Hetero-2-mer
HBB	P68871-HP_PL37A	CORE	pathogenic	2.1	1.6	0.6	0.4	Hetero-2-mer
HBB	P68871-HP_PL37H	CORE	pathogenic	3.4	1.4	-1.5	-1.5	Hetero-2-mer
HBB	P68871-HP_PL37R	CORE	pathogenic	1.1	2.8	1.8	2	Hetero-2-mer
HBB	P68871-HP_PL37S	CORE	pathogenic	2	1.3	2.3	2.1	Hetero-2-mer
HBB	P68871-HP_PL37T	CORE	pathogenic	2.4	1.5	1.9	1.8	Hetero-2-mer
HBB	P68871-HP_WL38C	CORE	pathogenic	4.4	2.4	8.4	8.4	Hetero-2-mer
HBB	P68871-HP_WL38R	CORE	pathogenic	6	3.6	12.3	12.1	Hetero-2-mer
HBB	P68871-HP_WL38S	CORE	pathogenic	4.4	3	9.4	9.3	Hetero-2-mer
HBB	P68871-HP_YL146C	CORE	pathogenic	1.5	2.6	3.5	2.3	Hetero-2-mer
HBB	P68871-HP_YL146D	CORE	pathogenic	4.2	4	2.8	1.6	Hetero-2-mer
HBB	P68871-HP_YL146H	CORE	pathogenic	4.4	2	2	0.9	Hetero-2-mer
HBB	P68871-HP_YL146N	CORE	pathogenic	4.1	3.8	3.6	2.3	Hetero-2-mer
HCFC1	P51610_RB2016W	CORE	pathogenic	-0.7	0.3	-0.5	-0.5	Hetero-4-mer
IVD	P26440_AA314V	CORE	pathogenic	-0.6	0.9	-0.2	-0.2	Homo-4-mer
IVD	P26440_AB314V	CORE	pathogenic	-0.8	1	-0.2	-0.2	Homo-4-mer
IVD	P26440_AC314V	CORE	pathogenic	-0.2	1.1	-0.2	-0.2	Homo-4-mer
IVD	P26440_AD314V	CORE	pathogenic	-1	1.1	-0.1	-0.2	Homo-4-mer
IVD	P26440_EA411K	CORE	pathogenic	2.8	1.1	14.1	14.9	Homo-4-mer
IVD	P26440_EB411K	CORE	pathogenic	3.4	0.7	12.3	13.2	Homo-4-mer
IVD	P26440_EC411K	CORE	pathogenic	2.3	0.9	13.4	14.4	Homo-4-mer
IVD	P26440_ED411K	CORE	pathogenic	3	1.6	12.1	14	Homo-4-mer
IVD	P26440_LA246P	CORE	pathogenic	8.8	1.9	-0.5	0.8	Homo-4-mer
IVD	P26440_LB246P	CORE	pathogenic	8.1	2.2	-1.1	0.8	Homo-4-mer
IVD	P26440_LC246P	CORE	pathogenic	8	2.1	-0.5	1.1	Homo-4-mer
IVD	P26440_LD246P	CORE	pathogenic	7.2	1.9	0	1.2	Homo-4-mer
IVD	P26440_RA398Q	CORE	pathogenic	1.7	-0.2	-1	-1	Homo-4-mer
IVD	P26440_RA414L	CORE	pathogenic	0.1	1.1	-1.1	0.3	Homo-4-mer
IVD	P26440_RA414W	CORE	pathogenic	2.2	0.5	-4.4	-4.1	Homo-4-mer
IVD	P26440_RB398Q	CORE	pathogenic	1.8	-0.3	1.5	2	Homo-4-mer
IVD	P26440_RB414L	CORE	pathogenic	-0.3	1.1	-2.1	-0.6	Homo-4-mer
IVD	P26440_RB414W	CORE	pathogenic	1.8	0.6	-5.6	-5.2	Homo-4-mer
IVD	P26440_RC398Q	CORE	pathogenic	2.4	-0.5	-0.5	-0.1	Homo-4-mer
IVD	P26440_RC414L	CORE	pathogenic	0.4	1.3	-3.2	-1.8	Homo-4-mer
IVD	P26440_RC414W	CORE	pathogenic	1.3	0.6	-6.7	-6.3	Homo-4-mer
IVD	P26440_RD398Q	CORE	pathogenic	3	-0.5	-0.4	0	Homo-4-mer
IVD	P26440_RD414L	CORE	pathogenic	0.3	1.3	-0.9	-0.3	Homo-4-mer
IVD	P26440_RD414W	CORE	pathogenic	2.7	0.7	-5.2	-4.8	Homo-4-mer
MLYCD	O95822_HA142N	CORE	pathogenic	0.9	0.2	5	3	Homo-4-mer
MLYCD	O95822_HB142N	CORE	pathogenic	0.1	0.3	-1	1.2	Homo-4-mer
MLYCD	O95822_HC142N	CORE	pathogenic	0.2	0.2	1.8	2.8	Homo-4-mer
MLYCD	O95822_HD142N	CORE	pathogenic	0	0.3	-1.1	1.1	Homo-4-mer
MMAB	Q96EY8_RA186Q	CORE	pathogenic	0.9	-0.4	0.9	0.9	Homo-3-mer
MMAB	Q96EY8_RA186W	CORE	pathogenic	2.6	0.1	0.6	0.6	Homo-3-mer
MMAB	Q96EY8_RA191Q	CORE	pathogenic	1.3	0	-4.1	-4.1	Homo-3-mer
MMAB	Q96EY8_RA191W	CORE	pathogenic	1.1	0.1	-10.7	-9.5	Homo-3-mer
MMAB	Q96EY8_RB186Q	CORE	pathogenic	0.3	-0.4	2	2	Homo-3-mer
MMAB	Q96EY8_RB186W	CORE	pathogenic	-0.7	0.1	-0.4	0	Homo-3-mer
MMAB	Q96EY8_RB191Q	CORE	pathogenic	1.3	-0.4	-4.1	-4.1	Homo-3-mer
MMAB	Q96EY8_RB191W	CORE	pathogenic	2.8	0.1	-8.3	-6.8	Homo-3-mer
MMAB	Q96EY8_RC186Q	CORE	pathogenic	1.3	-0.5	1.1	1.1	Homo-3-mer

MMAB	Q96EY8_RC186W	CORE	pathogenic	2.2	-0.1	1.5	-1.2	Homo-3-mer
MMAB	Q96EY8_RC191Q	CORE	pathogenic	-0.5	-0.3	-3.3	-3.4	Homo-3-mer
MMAB	Q96EY8_RC191W	CORE	pathogenic	-0.6	-0.1	-8.9	-8	Homo-3-mer
MMAB	Q96EY8_SA180W	CORE	pathogenic	23.1	2.1	-4.8	-5.6	Homo-3-mer
MMAB	Q96EY8_SB180W	CORE	pathogenic	27.9	2	-4	-3.8	Homo-3-mer
MMAB	Q96EY8_SC180W	CORE	pathogenic	34	2.2	-1.9	-3.9	Homo-3-mer
MUT	P22033_GA426E	CORE	pathogenic	11.7	2.3	-4.3	-1.5	Homo-2-mer
MUT	P22033_GA426R	CORE	pathogenic	10.5	2.7	4.3	5	Homo-2-mer
MUT	P22033_GA427D	CORE	pathogenic	24.5	3.1	1.5	3.2	Homo-2-mer
MUT	P22033_GA87E	CORE	pathogenic	14.4	1.4	-3.4	-1.3	Homo-2-mer
MUT	P22033_NA407Y	CORE	pathogenic	-2.7	1.9	-4.7	-4.4	Homo-2-mer
MUT	P22033_PA424L	CORE	pathogenic	9.3	1.2	0.8	2.8	Homo-2-mer
MUT	P22033_RA403Q	CORE	pathogenic	0.4	-0.8	6.1	7.1	Homo-2-mer
MUT	P22033_VA368D	CORE	pathogenic	3.4	1.1	2.1	2.8	Homo-2-mer
MUT	P22033_YA316C	CORE	pathogenic	4.9	2.6	3	2.3	Homo-2-mer
PAH	P00439_GA239A	CORE	pathogenic	7.2	0.9	-2.1	0.8	Homo-2-mer
PAH	P00439_GA239D	CORE	pathogenic	8.6	0.2	-5.8	1.1	Homo-2-mer
PAH	P00439_GA239S	CORE	pathogenic	8.7	0.4	-5.7	0.9	Homo-2-mer
PAH	P00439_GA239V	CORE	pathogenic	15.6	1.1	-4.8	2.3	Homo-2-mer
PAH	P00439_PA416Q	CORE	pathogenic	3.4	1.1	-7.5	0	Homo-2-mer
PAH	P00439_QA235P	CORE	pathogenic	5	0.8	-0.1	-0.1	Homo-2-mer
PAH	P00439_TA418N	CORE	pathogenic	0.6	1.2	0.3	0.2	Homo-2-mer
PAH	P00439_TA418P	CORE	pathogenic	0.1	1.8	1	1.2	Homo-2-mer
PAH	P00439_YA417C	CORE	pathogenic	1.7	1.9	7.6	7.2	Homo-2-mer
PAH	P00439_YA417D	CORE	pathogenic	1.8	2.4	5.2	4.8	Homo-2-mer
PAH	P00439_YA417H	CORE	pathogenic	0.6	1.4	4.6	4.7	Homo-2-mer
PAH	P00439_YA417N	CORE	pathogenic	1.3	2.2	12.9	12.9	Homo-2-mer
PAH	P00439-D1_GA46R	CORE	pathogenic	5	2.3	-1.5	-4.5	Homo-2-mer
PAH	P00439-D1_GA46S	CORE	pathogenic	5.2	0.8	2.2	-0.4	Homo-2-mer
PAH	P00439-D1_IA65M	CORE	pathogenic	-0.5	1.4	1	0.9	Homo-2-mer
PAH	P00439-D1_IA65N	CORE	pathogenic	4.5	1.1	0.2	1	Homo-2-mer
PAH	P00439-D1_IA65S	CORE	pathogenic	3.8	1.3	1.6	1.4	Homo-2-mer
PAH	P00439-D1_IA65T	CORE	pathogenic	3	1.3	0.1	0.1	Homo-2-mer
PAH	P00439-D1_IA65V	CORE	pathogenic	1.1	1.7	-0.1	-0.2	Homo-2-mer
PAH	P00439-D1_LA48S	CORE	pathogenic	3.6	1.9	0.8	0.8	Homo-2-mer
PAH	P00439-D1_LA52S	CORE	pathogenic	4.3	0.3	2.2	1.4	Homo-2-mer
PAH	P00439-D1_LA62P	CORE	pathogenic	4.7	1.8	-0.4	-0.4	Homo-2-mer
PAH	P00439-D1_NA61D	CORE	pathogenic	3.1	3.2	1.9	2.2	Homo-2-mer
PAH	P00439-D1_NA61K	CORE	pathogenic	0.9	2.7	-2.6	-2.2	Homo-2-mer
PAH	P00439-D1_SA67P	CORE	pathogenic	5.6	1.7	-1.7	-1.7	Homo-2-mer
SLC24A13	Q9UJS0_GA139R	CORE	pathogenic	11.5	2.7	2.8	1.9	Homo-2-mer
SLC24A13	Q9UJS0_GA176V	CORE	pathogenic	11.2	1.9	-0.4	0.7	Homo-2-mer
SLC24A13	Q9UJS0_YA148C	CORE	pathogenic	3.8	1.1	1.9	1.2	Homo-2-mer
TGFB1	P01137_CD223G	CORE	pathogenic	3.9	1.1	0.5	0.7	Homo-2-mer
TGFB1	P01137_CD223R	CORE	pathogenic	3.8	1.1	3.1	3.9	Homo-2-mer
TGFB1	P01137_HD222D	CORE	pathogenic	2.2	0.9	2.1	1.8	Homo-2-mer
TGFB1	P01137_YD81H	CORE	pathogenic	2.7	1.5	5.5	5	Homo-2-mer
ACADM	P11310_EA43K	RIM	pathogenic	1.3	0	1.5	1.5	Homo-4-mer
ACADM	P11310_EB43K	RIM	pathogenic	1.2	0	1.6	1.7	Homo-4-mer
ACADM	P11310_EC43K	RIM	pathogenic	1.3	0.2	2.8	1.8	Homo-4-mer
ACADM	P11310_ED43K	RIM	pathogenic	1.5	-0.1	2.4	2.4	Homo-4-mer

ACADM	P11310_FA309C	RIM	pathogenic	2.6	0.7	8.2	8.1	Homo-4-mer
ACADM	P11310_FB309C	RIM	pathogenic	2.4	0.5	8.6	8.6	Homo-4-mer
ACADM	P11310_FC309C	RIM	pathogenic	2.2	0.7	8.3	8.2	Homo-4-mer
ACADM	P11310_FD309C	RIM	pathogenic	2.2	0.7	8.2	8.2	Homo-4-mer
ACADM	P11310_GA310R	RIM	pathogenic	2.2	0.2	5.2	3.6	Homo-4-mer
ACADM	P11310_GB310R	RIM	pathogenic	3	0.1	0.7	0.7	Homo-4-mer
ACADM	P11310_GC310R	RIM	pathogenic	2.1	0.2	1.5	1.3	Homo-4-mer
ACADM	P11310_GD310R	RIM	pathogenic	1.9	0.1	3.1	2.4	Homo-4-mer
ACADM	P11310_LA409F	RIM	pathogenic	1.8	0	0.2	-1.5	Homo-4-mer
ACADM	P11310_LB409F	RIM	pathogenic	1.7	0.1	0.2	-1.5	Homo-4-mer
ACADM	P11310_LC409F	RIM	pathogenic	1.1	0.1	0.6	-1.1	Homo-4-mer
ACADM	P11310_LD409F	RIM	pathogenic	1.1	0	0.6	-1.1	Homo-4-mer
ACADM	P11310_RA243Q	RIM	pathogenic	1.8	0	-3.3	-2.2	Homo-4-mer
ACADM	P11310_RB243Q	RIM	pathogenic	2	0	-3.8	-2.7	Homo-4-mer
ACADM	P11310_RC243Q	RIM	pathogenic	1.6	0	-3	-2	Homo-4-mer
ACADM	P11310_RD243Q	RIM	pathogenic	1.8	0	-3.5	-2.4	Homo-4-mer
ACADM	P11310_SA167P	RIM	pathogenic	-2.6	-0.4	-0.9	-0.5	Homo-4-mer
ACADM	P11310_SB167P	RIM	pathogenic	-2.3	-0.6	-1	-0.5	Homo-4-mer
ACADM	P11310_SC167P	RIM	pathogenic	-2.6	-0.4	-1	-0.5	Homo-4-mer
ACADM	P11310_SD167P	RIM	pathogenic	-2.4	-0.5	-0.9	-0.5	Homo-4-mer
ACADM	P11310-ETFA-ETFB_NC186K	RIM	pathogenic	-0.3	0.7	17.4	14.5	Hetero-6-mer
ACADM	P11310-ETFA-ETFB_PC211S	RIM	pathogenic	2	0.1	-0.1	-0.2	Hetero-6-mer
ACADM	P11310-ETFA-ETFB_RD53C	RIM	pathogenic	1.1	0.2	-10	-4.5	Hetero-6-mer
ACADM	P11310-ETFA-ETFB_RD53H	RIM	pathogenic	0.1	0.2	-11.1	-5.4	Hetero-6-mer
ACADM	P11310-ETFA-ETFB_RD80G	RIM	pathogenic	2	0.6	-0.1	-1.8	Hetero-6-mer
ACADS	P16219_RE411W	RIM	pathogenic	1.5	0.6	-5.9	-1.4	Homo-4-mer
ACADS	P16219_RF411W	RIM	pathogenic	1.6	0.5	-0.8	1.8	Homo-4-mer
ACADS	P16219_RG411W	RIM	pathogenic	1.3	0.7	-5.4	-3.4	Homo-4-mer
ACADS	P16219_RH411W	RIM	pathogenic	1.3	0.6	-4	-0.7	Homo-4-mer
ACADVL	P49748_GA222R	RIM	pathogenic	0.3	1.4	12.8	9.9	Homo-2-mer
ACADVL	P49748_RA366C	RIM	pathogenic	0.9	1	-1.6	-1.2	Homo-2-mer
ACADVL	P49748_RA366H	RIM	pathogenic	1.3	1	-2.9	-2.6	Homo-2-mer
ACADVL	P49748_RA385W	RIM	pathogenic	-1.5	1	-1.7	-3.3	Homo-2-mer
ACADVL	P49748_RA531L	RIM	pathogenic	-0.9	1.1	2.9	-0.5	Homo-2-mer
ACADVL	P49748_RA531W	RIM	pathogenic	-0.5	0.9	3.1	-1.2	Homo-2-mer
ACADVL	P49748_RA615Q	RIM	pathogenic	0.3	-0.7	2.3	2.8	Homo-2-mer
ACAT1	P24752_HA144P	RIM	pathogenic	8.2	0.9	-0.4	-0.4	Homo-4-mer
ACAT1	P24752_HB144P	RIM	pathogenic	8.5	1.1	0.1	0	Homo-4-mer
ACAT1	P24752_HC144P	RIM	pathogenic	7.3	1	-0.4	-0.2	Homo-4-mer
ACAT1	P24752_HD144P	RIM	pathogenic	8.4	1	-0.4	-0.9	Homo-4-mer
ACAT1	P24752_NA158D	RIM	pathogenic	0.5	-0.2	-1.7	-1.7	Homo-4-mer
ACAT1	P24752_NA158S	RIM	pathogenic	0.5	-0.1	0	0	Homo-4-mer
ACAT1	P24752_NB158D	RIM	pathogenic	0.7	-0.2	-1.8	-1.8	Homo-4-mer
ACAT1	P24752_NB158S	RIM	pathogenic	0.9	-0.1	0.1	0.1	Homo-4-mer
ACAT1	P24752_NC158D	RIM	pathogenic	0.3	-0.2	-1.8	-1.8	Homo-4-mer
ACAT1	P24752_NC158S	RIM	pathogenic	0.8	-0.2	0.1	0.1	Homo-4-mer
ACAT1	P24752_ND158D	RIM	pathogenic	0.9	-0.2	-1.9	-1.9	Homo-4-mer

ACAT1	P24752_ND158S	RIM	pathogenic	0.9	-0.1	0	0	Homo-4-mer
ASL	P04424_HA160N	RIM	pathogenic	0.9	1.7	8.2	7.4	Homo-4-mer
ASL	P04424_HB160N	RIM	pathogenic	1	1.7	8.2	7.4	Homo-4-mer
ASL	P04424_HC160N	RIM	pathogenic	0.1	2.2	8.4	7.8	Homo-4-mer
ASL	P04424_HD160N	RIM	pathogenic	0.1	2.2	8.4	7.8	Homo-4-mer
ASL	P04424_KA315E	RIM	pathogenic	5.9	1.4	-2.1	-2.7	Homo-4-mer
ASL	P04424_KB315E	RIM	pathogenic	5.9	1.4	-2.1	-2.7	Homo-4-mer
ASL	P04424_KC315E	RIM	pathogenic	5.5	1.6	-1.1	-0.7	Homo-4-mer
ASL	P04424_KD315E	RIM	pathogenic	5.5	1.6	-1.2	-0.8	Homo-4-mer
ASL	P04424_RA379C	RIM	pathogenic	1	0.1	-0.2	-0.2	Homo-4-mer
ASL	P04424_RA385C	RIM	pathogenic	-0.2	0.5	-0.7	-0.9	Homo-4-mer
ASL	P04424_RA385H	RIM	pathogenic	0.7	0.5	-0.1	-0.3	Homo-4-mer
ASL	P04424_RA385L	RIM	pathogenic	-0.2	0.9	-0.7	-0.8	Homo-4-mer
ASL	P04424_RB379C	RIM	pathogenic	1.3	0.1	-0.2	-0.2	Homo-4-mer
ASL	P04424_RB385C	RIM	pathogenic	-0.1	0.5	-0.7	-0.9	Homo-4-mer
ASL	P04424_RB385H	RIM	pathogenic	0.2	0.5	-0.1	-0.3	Homo-4-mer
ASL	P04424_RB385L	RIM	pathogenic	0	0.9	-0.7	-0.8	Homo-4-mer
ASL	P04424_RC379C	RIM	pathogenic	0.3	0.3	-0.3	-0.4	Homo-4-mer
ASL	P04424_RC385C	RIM	pathogenic	-0.6	1.2	-3.5	-4	Homo-4-mer
ASL	P04424_RC385H	RIM	pathogenic	0	0.4	-2.5	-3.1	Homo-4-mer
ASL	P04424_RC385L	RIM	pathogenic	-0.8	1.3	-3	-3.4	Homo-4-mer
ASL	P04424_RD379C	RIM	pathogenic	0.5	0.3	-0.3	-0.4	Homo-4-mer
ASL	P04424_RD385C	RIM	pathogenic	-0.3	1.2	-3.5	-4	Homo-4-mer
ASL	P04424_RD385H	RIM	pathogenic	0.5	0.4	-2.5	-3.1	Homo-4-mer
ASL	P04424_RD385L	RIM	pathogenic	-0.6	1.3	-3	-3.3	Homo-4-mer
ASS1	P00966_RA157C	RIM	pathogenic	1.3	0.5	-0.6	-0.3	Homo-4-mer
ASS1	P00966_RA157H	RIM	pathogenic	0.8	0.4	-1.2	0.3	Homo-4-mer
ASS1	P00966_RB157C	RIM	pathogenic	0.8	0.5	-0.6	-0.3	Homo-4-mer
ASS1	P00966_RB157H	RIM	pathogenic	0.5	0.4	-1.2	0.3	Homo-4-mer
ASS1	P00966_RC157C	RIM	pathogenic	0.9	0.5	-0.6	-0.3	Homo-4-mer
ASS1	P00966_RC157H	RIM	pathogenic	0.1	0.4	-1.2	0.3	Homo-4-mer
ASS1	P00966_RD157C	RIM	pathogenic	1.2	0.5	-0.6	-0.3	Homo-4-mer
ASS1	P00966_RD157H	RIM	pathogenic	0.2	0.4	-1.2	0.3	Homo-4-mer
BCKDHA	P12694_GA244E	RIM	pathogenic	2.3	-0.7	2.3	2	Homo-2-mer
BCKDHA	P12694-P21953_GA244E	RIM	pathogenic	17.1	1.9	2.5	2	Hetero-4-mer
BCKDHA	P12694-P21953_GB244E	RIM	pathogenic	16.7	1.9	2.5	2	Hetero-4-mer
BCKDHA	P12694-P21953_HA247R	RIM	pathogenic	3	2.2	-2	-2	Hetero-4-mer
BCKDHA	P12694-P21953_HB247R	RIM	pathogenic	5.3	2.2	-2	-2	Hetero-4-mer
BCKDHA	P12694-P21953_VA412M	RIM	pathogenic	4	0.8	0.5	0	Hetero-4-mer
BCKDHA	P12694-P21953_VA412M	RIM	pathogenic	4	0.8	0.5	0	Hetero-4-mer
BCKDHA	P12694-P21953_VA412M	RIM	pathogenic	3.5	0.8	0.5	0	Hetero-4-mer
BCKDHA	P12694-P21953_VA412M	RIM	pathogenic	3.5	0.8	0.5	0	Hetero-4-mer
BCKDHA	P12694-P21953_VB412M	RIM	pathogenic	4.1	0.8	0.5	0	Hetero-4-mer
BCKDHA	P12694-P21953_VB412M	RIM	pathogenic	4.1	0.8	0.5	0	Hetero-4-mer
BCKDHA	P12694-P21953_VB412M	RIM	pathogenic	2.1	0.8	0.5	0	Hetero-4-mer
BCKDHA	P12694-P21953_VB412M	RIM	pathogenic	2.1	0.8	0.5	0	Hetero-4-mer
BCKDHB	P12694-P21953_PC123L	RIM	pathogenic	9	1.8	4.1	1.9	Hetero-4-mer
BCKDHB	P12694-P21953_PC123L	RIM	pathogenic	9	1.8	4.1	1.9	Hetero-4-mer
BCKDHB	P12694-P21953_PC123L	RIM	pathogenic	8.9	1.8	4.1	1.9	Hetero-4-mer
BCKDHB	P12694-P21953_PC123L	RIM	pathogenic	8.9	1.8	4.1	1.9	Hetero-4-mer
BCKDHB	P12694-P21953_PD123L	RIM	pathogenic	5.2	1.8	4.1	1.9	Hetero-4-mer

BCKDHB	P12694-P21953_PD123L	RIM	pathogenic	5.2	1.8	4.1	1.9	Hetero-4-mer
BCKDHB	P12694-P21953_PD123L	RIM	pathogenic	5.3	1.8	4.1	1.9	Hetero-4-mer
BCKDHB	P12694-P21953_PD123L	RIM	pathogenic	5.3	1.8	4.1	1.9	Hetero-4-mer
BCKDHB	P21953_RD170C	RIM	pathogenic	0.8	0.7	0	-0.9	Homo-2-mer
BCKDHB	P21953_RD170H	RIM	pathogenic	0.9	0.4	-0.9	0.5	Homo-2-mer
CBS	P35520_DA198V	RIM	pathogenic	1.7	1	-1.1	0.7	Homo-2-mer
CBS	P35520_NA93Y	RIM	pathogenic	1.5	0.8	-4.2	-3.2	Homo-2-mer
CBS	P35520_PA200L	RIM	pathogenic	1.7	0.4	0.7	0.3	Homo-2-mer
CBS	P35520_PA78R	RIM	pathogenic	1.6	1.2	1.4	0.9	Homo-2-mer
CBS	P35520_RA336C	RIM	pathogenic	3.8	0.6	0.2	0.2	Homo-2-mer
CBS	P35520_RA336H	RIM	pathogenic	3.7	0.2	-2.1	-1.9	Homo-2-mer
ETFA	P11310-ETFA-ETFB_RR249C	RIM	pathogenic	0.6	0.6	-4.2	-2.2	Hetero-6-mer
ETFA	P11310-ETFA-ETFB_TR266M	RIM	pathogenic	-0.5	-0.3	2.7	0.2	Hetero-6-mer
ETFA	P13804-P38117_RA249C	RIM	pathogenic	1.1	0.7	0.7	0.4	Hetero-2-mer
ETFB	P11310-ETFA-ETFB_RS191C	RIM	pathogenic	2.8	0.5	0.9	0.9	Hetero-6-mer
ETFB	P13804-P38117_AB17P	RIM	pathogenic	3.8	0.7	0.3	0.2	Hetero-2-mer
FCGR2A	P12318-IGHG1_QC163K	RIM	pathogenic	-0.3	0.5	-0.1	-0.1	Hetero-3-mer
GALT	P07902_AB176V	RIM	pathogenic	0.3	0.3	-0.1	-0.1	Homo-2-mer
GALT	P07902_DB113N	RIM	pathogenic	0.2	1.4	-0.6	3.1	Homo-2-mer
GALT	P07902_GB179D	RIM	pathogenic	7.6	-0.3	-2.2	-1.1	Homo-2-mer
GALT	P07902_GB179S	RIM	pathogenic	6.9	0.2	-0.1	0.9	Homo-2-mer
GALT	P07902_QB103R	RIM	pathogenic	0.2	1.3	1.3	1	Homo-2-mer
GALT	P07902_QB118H	RIM	pathogenic	0.7	-0.3	0.8	0.5	Homo-2-mer
GALT	P07902_RB204P	RIM	pathogenic	4.3	0.4	2.3	2.3	Homo-2-mer
GALT	P07902_RB223S	RIM	pathogenic	1	-0.3	0.5	0.5	Homo-2-mer
GALT	P07902_RB48C	RIM	pathogenic	0.2	0.3	-0.7	-0.5	Homo-2-mer
GALT	P07902_TB23A	RIM	pathogenic	0	0.8	1	1	Homo-2-mer
GCDH	Q92947_GA185A	RIM	pathogenic	0.5	1.1	1.5	-0.5	Homo-4-mer
GCDH	Q92947_GB185A	RIM	pathogenic	0.5	1.1	1.5	-0.5	Homo-4-mer
GCDH	Q92947_GC185A	RIM	pathogenic	0.4	1.1	1.5	-0.5	Homo-4-mer
GCDH	Q92947_GD185A	RIM	pathogenic	0.5	1.1	1.5	-0.5	Homo-4-mer
GCDH	Q92947_KA361E	RIM	pathogenic	2.2	0.7	4.4	6.3	Homo-4-mer
GCDH	Q92947_KB361E	RIM	pathogenic	2.1	0.7	4.5	6.3	Homo-4-mer
GCDH	Q92947_KC361E	RIM	pathogenic	2.1	0.7	3.5	5.4	Homo-4-mer
GCDH	Q92947_KD361E	RIM	pathogenic	2	0.7	4.5	6.3	Homo-4-mer
GCDH	Q92947_QA59P	RIM	pathogenic	4.2	0.4	-1	-0.7	Homo-4-mer
GCDH	Q92947_QB59P	RIM	pathogenic	5.5	0.4	-1	-0.7	Homo-4-mer
GCDH	Q92947_QC59P	RIM	pathogenic	4.3	0.4	-1	-0.7	Homo-4-mer
GCDH	Q92947_QD59P	RIM	pathogenic	4.1	0.4	-1	-0.7	Homo-4-mer
GCDH	Q92947_RA88C	RIM	pathogenic	0.5	1	-1.3	-1.6	Homo-4-mer
GCDH	Q92947_RA88C	RIM	pathogenic	0.5	1	-1.3	-1.6	Homo-4-mer
GCDH	Q92947_RA88C	RIM	pathogenic	0.5	1	-1.3	-1.6	Homo-4-mer
GCDH	Q92947_RA88C	RIM	pathogenic	0.5	1	-1.3	-1.6	Homo-4-mer
GCDH	Q92947_RB88C	RIM	pathogenic	0.3	1	-1.3	-1.6	Homo-4-mer
GCDH	Q92947_RB88C	RIM	pathogenic	0.3	1	-1.3	-1.6	Homo-4-mer
GCDH	Q92947_RB88C	RIM	pathogenic	0.3	1	-1.3	-1.6	Homo-4-mer
GCDH	Q92947_RB88C	RIM	pathogenic	0.3	1	-1.3	-1.6	Homo-4-mer
GCDH	Q92947_RC88C	RIM	pathogenic	0.6	1	-1.3	-1.6	Homo-4-mer
GCDH	Q92947_RC88C	RIM	pathogenic	0.6	1	-1.3	-1.6	Homo-4-mer

GCDH	Q92947_RC88C	RIM	pathogenic	0.5	1	-1.3	-1.6	Homo-4-mer
GCDH	Q92947_RC88C	RIM	pathogenic	0.5	1	-1.3	-1.6	Homo-4-mer
GCDH	Q92947_RD88C	RIM	pathogenic	0.4	1	-1.3	-1.6	Homo-4-mer
GCDH	Q92947_RD88C	RIM	pathogenic	0.4	1	-1.3	-1.6	Homo-4-mer
GCDH	Q92947_RD88C	RIM	pathogenic	0.6	1	-1.3	-1.6	Homo-4-mer
GCDH	Q92947_RD88C	RIM	pathogenic	0.6	1	-1.3	-1.6	Homo-4-mer
GCDH	Q92947_SA186C	RIM	pathogenic	-1	-0.3	-0.1	0	Homo-4-mer
GCDH	Q92947_SB186C	RIM	pathogenic	-0.7	-0.3	-0.1	0	Homo-4-mer
GCDH	Q92947_SC186C	RIM	pathogenic	-0.3	-0.3	-0.1	0	Homo-4-mer
GCDH	Q92947_SD186C	RIM	pathogenic	-0.8	-0.3	-0.1	0	Homo-4-mer
HADHB	P40939-HADHB_LB121P	RIM	pathogenic	4.1	0.6	0	0	Hetero-3-mer
HADHB	P40939-HADHB_RA229L	RIM	pathogenic	2.4	0.9	-3.6	-1.5	Hetero-3-mer
HADHB	P55084_LA121P	RIM	pathogenic	2.8	0.4	0.8	0.7	Homo-2-mer
HADHB	P55084_NA114D	RIM	pathogenic	0.7	1.1	-2.7	-2.8	Homo-2-mer
HADHB	P55084_NA114S	RIM	pathogenic	2	1.9	0.1	0.1	Homo-2-mer
HADHB	P55084_RA174C	RIM	pathogenic	-3.1	0.8	-16.2	-15.4	Homo-2-mer
HBB	P68871_HB147D	RIM	pathogenic	-0.1	1.9	-0.7	0.1	Homo-2-mer
HBB	P68871_HB147L	RIM	pathogenic	-1.2	1.2	0.6	0.4	Homo-2-mer
HBB	P68871_HB147P	RIM	pathogenic	1.7	1.2	2	1.8	Homo-2-mer
HBB	P68871_HB147Q	RIM	pathogenic	-0.1	0.9	2.4	2	Homo-2-mer
HBB	P68871_HB147Y	RIM	pathogenic	-2.2	-0.1	-2.2	-1.9	Homo-2-mer
HBB	P68871_KB83E	RIM	pathogenic	-0.8	0.5	-1.6	-2.2	Homo-2-mer
HBB	P68871_KB83M	RIM	pathogenic	-0.3	0	-2.1	-2.6	Homo-2-mer
HBB	P68871_KB83Q	RIM	pathogenic	0	0.6	-1	-1.5	Homo-2-mer
HBB	P68871_KB83R	RIM	pathogenic	0.1	0.8	-0.6	0.2	Homo-2-mer
HBB	P68871_KB83T	RIM	pathogenic	0.3	0.6	-2	-2.5	Homo-2-mer
HBB	P68871_YB146C	RIM	pathogenic	2.7	1	0.1	0.1	Homo-2-mer
HBB	P68871_YB146D	RIM	pathogenic	3.5	1.6	-1	-1	Homo-2-mer
HBB	P68871_YB146H	RIM	pathogenic	2	0.5	0	0	Homo-2-mer
HBB	P68871_YB146N	RIM	pathogenic	3.3	1.9	-0.1	-0.1	Homo-2-mer
HBB	P68871-HBA_EB27A	RIM	pathogenic	1.3	0.6	0.9	0.9	Hetero-4-mer
HBB	P68871-HBA_EB27K	RIM	pathogenic	0.9	0.2	0.6	0.6	Hetero-4-mer
HBB	P68871-HBA_EB44Q	RIM	pathogenic	0.2	1.3	2.6	4.7	Hetero-4-mer
HBB	P68871-HBA_ED27A	RIM	pathogenic	0.9	0.4	0.6	0.6	Hetero-4-mer
HBB	P68871-HBA_ED27K	RIM	pathogenic	0.7	0.2	1.1	1.1	Hetero-4-mer
HBB	P68871-HBA_ED44Q	RIM	pathogenic	0.1	1.4	3.5	5.2	Hetero-4-mer
HBB	P68871-HBA_HB147D	RIM	pathogenic	0.1	1.9	-2.9	-2.7	Hetero-4-mer
HBB	P68871-HBA_HB147L	RIM	pathogenic	0.5	1.8	-1.2	-0.9	Hetero-4-mer
HBB	P68871-HBA_HB147P	RIM	pathogenic	0.4	1.8	-1.3	-1	Hetero-4-mer
HBB	P68871-HBA_HB147Q	RIM	pathogenic	-0.1	1.4	-1.2	-0.9	Hetero-4-mer
HBB	P68871-HBA_HB147Y	RIM	pathogenic	-1.2	1.2	-1.4	-1.9	Hetero-4-mer
HBB	P68871-HBA_HB98L	RIM	pathogenic	-0.3	2.1	0.5	0.4	Hetero-4-mer
HBB	P68871-HBA_HB98N	RIM	pathogenic	0.6	1.6	2.6	2.4	Hetero-4-mer
HBB	P68871-HBA_HB98P	RIM	pathogenic	5.3	2.2	1.1	0.6	Hetero-4-mer
HBB	P68871-HBA_HB98Q	RIM	pathogenic	0.7	1.7	1.8	1.7	Hetero-4-mer
HBB	P68871-HBA_HB98Y	RIM	pathogenic	-0.2	1.1	-3.4	-3.3	Hetero-4-mer
HBB	P68871-HBA_HD147D	RIM	pathogenic	0.5	1.6	-2.6	-2.6	Hetero-4-mer
HBB	P68871-HBA_HD147L	RIM	pathogenic	0.7	1.5	-0.9	-0.7	Hetero-4-mer
HBB	P68871-HBA_HD147P	RIM	pathogenic	0.7	1.5	-1.1	-0.8	Hetero-4-mer
HBB	P68871-HBA_HD147Q	RIM	pathogenic	0.2	1.2	-0.9	-0.7	Hetero-4-mer
HBB	P68871-HBA_HD147Y	RIM	pathogenic	-1	0.9	-1.1	-1.7	Hetero-4-mer

HBB	P68871-HBA_KB133N	RIM	pathogenic	1.4	-0.5	-0.2	-0.2	Hetero-4-mer
HBB	P68871-HBA_KB133Q	RIM	pathogenic	0.6	-0.4	-0.1	-0.1	Hetero-4-mer
HBB	P68871-HBA_KB133T	RIM	pathogenic	1.3	-0.4	-0.1	-0.1	Hetero-4-mer
HBB	P68871-HBA_KD133N	RIM	pathogenic	0.9	0.2	-0.2	-0.2	Hetero-4-mer
HBB	P68871-HBA_KD133Q	RIM	pathogenic	0.5	0.3	-0.7	-0.7	Hetero-4-mer
HBB	P68871-HBA_KD133T	RIM	pathogenic	1.7	0.4	-0.1	-0.1	Hetero-4-mer
HBB	P68871-HBA_PB52R	RIM	pathogenic	2.5	1.6	5.9	8.1	Hetero-4-mer
HBB	P68871-HBA_PD52R	RIM	pathogenic	2.1	1.5	5.7	7.9	Hetero-4-mer
HBB	P68871-HBA_RB105S	RIM	pathogenic	0.2	0.7	-1.3	-1.3	Hetero-4-mer
HBB	P68871-HBA_RB105T	RIM	pathogenic	0.4	0.7	-1.3	-1.3	Hetero-4-mer
HBB	P68871-HBA_RB105W	RIM	pathogenic	-0.8	0.3	-4	-3.8	Hetero-4-mer
HBB	P68871-HBA_RD105S	RIM	pathogenic	0.1	-0.3	-1.1	-1.1	Hetero-4-mer
HBB	P68871-HBA_RD105T	RIM	pathogenic	-0.2	-0.2	-1.1	-1.1	Hetero-4-mer
HBB	P68871-HBA_RD105W	RIM	pathogenic	-1.2	0	-3.6	-3.4	Hetero-4-mer
HBB	P68871-HBA_TB124I	RIM	pathogenic	-0.1	-0.4	0.4	0.4	Hetero-4-mer
HBB	P68871-HBA_TD124I	RIM	pathogenic	0.6	0.1	0.5	0.5	Hetero-4-mer
HBB	P68871-HBAZ_EB102A	RIM	pathogenic	-1.4	1.1	2.4	2.1	Hetero-4-mer
HBB	P68871-HBAZ_EB102D	RIM	pathogenic	2.1	1.2	-0.9	-1.1	Hetero-4-mer
HBB	P68871-HBAZ_EB102G	RIM	pathogenic	-0.4	1.3	2.5	2.6	Hetero-4-mer
HBB	P68871-HBAZ_EB102K	RIM	pathogenic	-1.8	1.1	4.5	4.3	Hetero-4-mer
HBB	P68871-HBAZ_EB102Q	RIM	pathogenic	-1.7	1.8	1.2	0.9	Hetero-4-mer
HBB	P68871-HBAZ_ED102A	RIM	pathogenic	-0.7	0.7	1.4	1.3	Hetero-4-mer
HBB	P68871-HBAZ_ED102D	RIM	pathogenic	1.7	1.1	-1.1	-0.9	Hetero-4-mer
HBB	P68871-HBAZ_ED102G	RIM	pathogenic	0.2	0.9	0.9	1.8	Hetero-4-mer
HBB	P68871-HBAZ_ED102K	RIM	pathogenic	-1.2	0.5	2.9	3	Hetero-4-mer
HBB	P68871-HBAZ_ED102Q	RIM	pathogenic	-1	1.2	0.4	0.5	Hetero-4-mer
HBB	P68871-HBAZ_HB147D	RIM	pathogenic	0.1	1.6	-1.9	-1.8	Hetero-4-mer
HBB	P68871-HBAZ_HB147L	RIM	pathogenic	0.3	1.6	-0.7	-0.5	Hetero-4-mer
HBB	P68871-HBAZ_HB147P	RIM	pathogenic	-0.1	1.6	-0.6	-0.5	Hetero-4-mer
HBB	P68871-HBAZ_HB147Q	RIM	pathogenic	0.1	1.2	-0.5	-0.4	Hetero-4-mer
HBB	P68871-HBAZ_HB147Y	RIM	pathogenic	-1	1	-1.5	-1.6	Hetero-4-mer
HBB	P68871-HBAZ_HB98L	RIM	pathogenic	-0.5	0.9	1.1	0.7	Hetero-4-mer
HBB	P68871-HBAZ_HB98N	RIM	pathogenic	-0.7	0.2	2.8	2.3	Hetero-4-mer
HBB	P68871-HBAZ_HB98P	RIM	pathogenic	5.3	0.8	2.1	1.3	Hetero-4-mer
HBB	P68871-HBAZ_HB98Q	RIM	pathogenic	-0.2	0.3	2.7	2.1	Hetero-4-mer
HBB	P68871-HBAZ_HB98Y	RIM	pathogenic	-0.4	0	-3.4	-3.5	Hetero-4-mer
HBB	P68871-HBAZ_HD147D	RIM	pathogenic	0.8	1.6	-1.8	-1.9	Hetero-4-mer
HBB	P68871-HBAZ_HD147L	RIM	pathogenic	1.1	1.6	-0.7	-0.6	Hetero-4-mer
HBB	P68871-HBAZ_HD147P	RIM	pathogenic	0.7	1.6	-0.7	-0.6	Hetero-4-mer
HBB	P68871-HBAZ_HD147Q	RIM	pathogenic	0.2	1.2	-0.5	-0.4	Hetero-4-mer
HBB	P68871-HBAZ_HD147Y	RIM	pathogenic	-0.5	1	-1	-1.6	Hetero-4-mer
HBB	P68871-HBAZ_HD98L	RIM	pathogenic	0.1	0.8	0.5	0.4	Hetero-4-mer
HBB	P68871-HBAZ_HD98N	RIM	pathogenic	0.1	0.3	1.7	1.5	Hetero-4-mer
HBB	P68871-HBAZ_HD98P	RIM	pathogenic	4	0.7	0.8	0.4	Hetero-4-mer
HBB	P68871-HBAZ_HD98Q	RIM	pathogenic	0.4	0.3	1.1	0.9	Hetero-4-mer
HBB	P68871-HBAZ_HD98Y	RIM	pathogenic	0.1	-0.2	-4.3	-4.6	Hetero-4-mer
HBB	P68871-HBAZ_MB56K	RIM	pathogenic	1.4	0.2	6.5	6.5	Hetero-4-mer
HBB	P68871-HBAZ_MD56K	RIM	pathogenic	1.1	0.3	3.1	1.8	Hetero-4-mer
HBB	P68871-HBAZ_PB37A	RIM	pathogenic	2.4	0.7	0.1	0.1	Hetero-4-mer
HBB	P68871-HBAZ_PB37H	RIM	pathogenic	3.5	0.7	0.4	-0.2	Hetero-4-mer
HBB	P68871-HBAZ_PB37R	RIM	pathogenic	0.8	2	2.4	1.1	Hetero-4-mer

HBB	P68871-HBAZ_PB37S	RIM	pathogenic	3	0.7	3.5	1.2	Hetero-4-mer
HBB	P68871-HBAZ_PB37T	RIM	pathogenic	3.2	0.9	-0.3	-0.3	Hetero-4-mer
HBB	P68871-HBAZ_PB52R	RIM	pathogenic	2	1.5	-0.1	0.8	Hetero-4-mer
HBB	P68871-HBAZ_PD37A	RIM	pathogenic	2.3	0.9	-0.4	-0.4	Hetero-4-mer
HBB	P68871-HBAZ_PD37H	RIM	pathogenic	3.2	0.7	0	-0.4	Hetero-4-mer
HBB	P68871-HBAZ_PD37R	RIM	pathogenic	0.9	2	1.7	1.7	Hetero-4-mer
HBB	P68871-HBAZ_PD37S	RIM	pathogenic	2.9	0.6	3.7	0.7	Hetero-4-mer
HBB	P68871-HBAZ_PD37T	RIM	pathogenic	3.1	0.8	-0.2	-0.2	Hetero-4-mer
HBB	P68871-HBAZ_PD52R	RIM	pathogenic	1.8	1.5	1.1	0.8	Hetero-4-mer
HBB	P68871-HBAZ_RB105S	RIM	pathogenic	0	0.1	-4.9	-4.8	Hetero-4-mer
HBB	P68871-HBAZ_RB105T	RIM	pathogenic	-0.1	0.2	-4.9	-4.8	Hetero-4-mer
HBB	P68871-HBAZ_RB105W	RIM	pathogenic	-0.5	0.6	-8	-7.7	Hetero-4-mer
HBB	P68871-HBAZ_RB41K	RIM	pathogenic	0.7	0.2	0.1	1	Hetero-4-mer
HBB	P68871-HBAZ_RD105S	RIM	pathogenic	0.2	-0.2	-3.3	-3.4	Hetero-4-mer
HBB	P68871-HBAZ_RD105T	RIM	pathogenic	0.2	-0.1	-3.3	-3.4	Hetero-4-mer
HBB	P68871-HBAZ_RD105W	RIM	pathogenic	-0.3	0.4	-3.4	-3.5	Hetero-4-mer
HBB	P68871-HBAZ_RD41K	RIM	pathogenic	0	0.1	0.8	1	Hetero-4-mer
HBB	P68871-HBAZ_TB124I	RIM	pathogenic	-0.9	0.1	0.2	0.2	Hetero-4-mer
HBB	P68871-HBAZ_TD124I	RIM	pathogenic	-0.9	0.2	0.2	0.2	Hetero-4-mer
HBB	P68871-HP_EL44Q	RIM	pathogenic	0.2	1.2	2.4	2.5	Hetero-2-mer
HBB	P68871-HP_HL147D	RIM	pathogenic	0.1	0.6	-1.3	-1.3	Hetero-2-mer
HBB	P68871-HP_HL147L	RIM	pathogenic	0	0.9	0.6	0.6	Hetero-2-mer
HBB	P68871-HP_HL147P	RIM	pathogenic	0	0.8	0.7	0.7	Hetero-2-mer
HBB	P68871-HP_HL147Q	RIM	pathogenic	-0.4	0.5	-0.6	-0.6	Hetero-2-mer
HBB	P68871-HP_HL147Y	RIM	pathogenic	-1	0.3	4.5	0.4	Hetero-2-mer
HBB	P68871-HP_HL98L	RIM	pathogenic	1.9	1	1.6	1.6	Hetero-2-mer
HBB	P68871-HP_HL98N	RIM	pathogenic	1.5	0.1	2.2	2	Hetero-2-mer
HBB	P68871-HP_HL98P	RIM	pathogenic	6.6	1	1.3	0.8	Hetero-2-mer
HBB	P68871-HP_HL98Q	RIM	pathogenic	1.6	0.1	0.3	0.6	Hetero-2-mer
HBB	P68871-HP_HL98Y	RIM	pathogenic	0.5	0.5	-3.2	-3.2	Hetero-2-mer
HBB	P68871-HP_NL109D	RIM	pathogenic	0	-0.3	-0.3	-0.3	Hetero-2-mer
HBB	P68871-HP_NL109I	RIM	pathogenic	-1.1	0.3	0.5	0.5	Hetero-2-mer
HBB	P68871-HP_NL109K	RIM	pathogenic	-0.6	0.5	6.3	5.6	Hetero-2-mer
HBB	P68871-HP_RL105S	RIM	pathogenic	0.2	0	-8.6	-3.5	Hetero-2-mer
HBB	P68871-HP_RL105T	RIM	pathogenic	0.1	0.1	-8.5	-3.5	Hetero-2-mer
HBB	P68871-HP_RL105W	RIM	pathogenic	-0.5	0.2	-11	-6	Hetero-2-mer
HBB	P68871-HP_RL41K	RIM	pathogenic	-0.4	1.9	-0.3	-2.1	Hetero-2-mer
IVD	P26440_KA318Q	RIM	pathogenic	0.7	-0.1	0.8	0.8	Homo-4-mer
IVD	P26440_KB318Q	RIM	pathogenic	0.8	-0.3	0.6	0.6	Homo-4-mer
IVD	P26440_KC318Q	RIM	pathogenic	0.7	-0.1	0.4	0.4	Homo-4-mer
IVD	P26440_KD318Q	RIM	pathogenic	0.7	-0.2	0.9	0.9	Homo-4-mer
MLYCD	O95822_GB300V	RIM	pathogenic	3.4	0.1	-0.2	-0.2	Homo-4-mer
MLYCD	O95822_GC300V	RIM	pathogenic	1.9	0.2	0	0	Homo-4-mer
MLYCD	O95822_GD300V	RIM	pathogenic	2.5	0.1	-0.1	-0.1	Homo-4-mer
MMAB	Q96EY8_RA190C	RIM	pathogenic	0.9	0.3	0.3	0.2	Homo-3-mer
MMAB	Q96EY8_RA190H	RIM	pathogenic	0.4	-0.2	0.1	0.2	Homo-3-mer
MMAB	Q96EY8_RB190C	RIM	pathogenic	0.3	0.4	1	1.6	Homo-3-mer
MMAB	Q96EY8_RB190H	RIM	pathogenic	1.1	-0.1	0	0.7	Homo-3-mer
MMAB	Q96EY8_RC190C	RIM	pathogenic	0.4	0.2	0.5	0.4	Homo-3-mer
MMAB	Q96EY8_RC190H	RIM	pathogenic	0.4	-0.2	0	0.4	Homo-3-mer
MUT	P22033_YA364S	RIM	pathogenic	3.1	1.8	6.7	4.5	Homo-2-mer

PAH	P00439_EA422K	RIM	pathogenic	-0.1	0.9	18.9	15.1	Homo-2-mer
PAH	P00439_QA301H	RIM	pathogenic	0.7	-0.4	-1.3	-1.3	Homo-2-mer
PAH	P00439_QA301P	RIM	pathogenic	3.8	0.1	-0.3	-0.3	Homo-2-mer
PAH	P00439_RA297C	RIM	pathogenic	0.9	0.5	-5.1	-3.6	Homo-2-mer
PAH	P00439_RA297H	RIM	pathogenic	-0.1	-0.1	-6.7	-5.7	Homo-2-mer
PAH	P00439_RA297L	RIM	pathogenic	-0.7	0.6	-4.8	-4.8	Homo-2-mer
PAH	P00439_RA413C	RIM	pathogenic	1.1	0.1	-9.2	-6.7	Homo-2-mer
PAH	P00439_RA413G	RIM	pathogenic	1.2	0.3	-9.2	-6.8	Homo-2-mer
PAH	P00439_RA413P	RIM	pathogenic	3.4	0.5	-9.2	-6.7	Homo-2-mer
PAH	P00439_RA413S	RIM	pathogenic	0.6	-0.3	-9.1	-6.7	Homo-2-mer
PAH	P00439_RA420M	RIM	pathogenic	0.1	0.2	-0.9	-0.4	Homo-2-mer
PAH	P00439-D1_DA59Y	RIM	pathogenic	-0.7	0	-1.1	-1.1	Homo-2-mer
PAH	P00439-D1_EA56D	RIM	pathogenic	0.8	1.1	-1.4	0.5	Homo-2-mer
PAH	P00439-D1_EA66K	RIM	pathogenic	-0.9	1.6	-4.4	-4.6	Homo-2-mer
PAH	P00439-D1_PA69S	RIM	pathogenic	1.8	0.1	0.1	0.1	Homo-2-mer
PAH	P00439-D1_RA53C	RIM	pathogenic	0.3	0.5	-7.1	-7	Homo-2-mer
PAH	P00439-D1_RA53H	RIM	pathogenic	0.6	0.3	-3.1	-3.3	Homo-2-mer
PAH	P00439-D1_VA45A	RIM	pathogenic	0.5	2	2	2	Homo-2-mer
SLC24A13	Q9UJS0_EA252K	RIM	pathogenic	0.5	0.4	0.1	0.2	Homo-2-mer
TAT	P17735_LA312P	RIM	pathogenic	3.7	0.4	0.2	0.2	Homo-2-mer
TGFB1	P01137_CD225G	RIM	pathogenic	3.6	1.2	3.2	1.9	Homo-2-mer
TGFB1	P01137_CD225R	RIM	pathogenic	3.6	0.7	5.5	5.4	Homo-2-mer

8.3. Appendix 3. Supplementary material for Chapter 5

Table 8.3.1 Structural data of PPIs with known modulators. Extended information.

PPI	protein-protein complex BU[a]	receptor Uniprot/BU[b]	ligand Uniprot/BU[b]	Additional information[c]
Bcl-xL / Bak	Hetero-2mer	Q07817/Homo-2mer	Q16611/Monomer	Human / Human ligand: 16-163 residues
	receptor-PPI inhibitor complexes[d]: 1YSG,1YSI,2O1Y,2O2M,2O2N,*2YXJ*,3INQ,3QKD,3SP7,3WIZ,3ZK6,3ZLN,3ZLO,3ZLR,4C52,4C5D, 4EHR and 4TUH			
HPV E2 / E1	Hetero-2mer	P04015/Monomer	P03116/Homo-6mer	HPV type 11 / HPV type 11
	receptor-PPI inhibitor complexes[d]: *1R6N*			
IL-2 / IL-2R	Hetero-2mer	P60568/Monomer	P01589/Monomer	Human / Human
	receptor-PPI inhibitor complexes[d]: 1M48, 1M49, *1PY2*, 1PW6 and 1QVN			
Integrase / LEDGF	Hetero-4mer	Q72498/Homo-2mer	O75475/Monomer	HIV-1 / Human
	receptor-PPI inhibitor complexes[d]: 3LPT,*3LPU*,4E1M and 4E1N			
MDM2 / p53	Hetero-2mer	Q00987/Monomer	P04637/Monomer	Human / Human ligand: 14-39 residues
	receptor-PPI inhibitor complexes[d]: 1RV1,1T4E,1TTV,2LZG,3JZK,3LBK,3LBL,3TJ2,3TU1,3VZV,3W69,4DIJ,4ERE,*4ERF*,4HBM,4HG7, 4JV7,4JV9,4JVE,4JVR,4JWR,4MDN,4MDQ,4OAS,4OBA,4OCC,4ODE,4ODF,4OGN,4OGT,4OGV,4OQ3 and 4QO4			
XIAP BIR3 / Caspase	Hetero-2mer	P98170/Monomer	P55211/Homo-2mer	Human / Human
	receptor-PPI inhibitor complexes[d]: 1TFQ,*1TFT*,2JK7,2OPY,3CLX,3CM2,3EYL,3G76,3HL5,3OZ1,4HY4,4HY5,4KJU,4KJV,4KMN,4KMP, 4LGE,4LGU,4MTI and 4MU7			
XIAP BIR3 / Smac	Hetero-2mer	P98170/Monomer	Q9NR28/Homo-2mer	Human / Human
	receptor-PPI inhibitor complexes[d]: 1TFQ,1TFT,*2JK7*,2OPY,3CLX,3CM2,3EYL,3G76,3HL5,3OZ1,4HY4,4HY5,4KJU,4KJV,4KMN,4KMP, 4LGE,4LGU,4MTI and 4MU7			
TNFR1A / TNF-β	Hetero-2mer	P19438/Homo-2mer	P01374/Homo-3mer	Human / Human
	receptor-PPI inhibitor complexes[d]: *1FT4*			
ZipA / FtsZ	Hetero-2mer	P77173/Homo-2mer	P47204/Monomer	*E. coli* / *E. coli*
	receptor-PPI inhibitor complexes[d]: *1S1J*,1S1S, 1Y2F and 1Y2G			

[a] BU: biological unit of the protein-protein complex as annotated in the PDB

[b] Uniprot code of the receptor or ligand molecules: BU: biological unit of the receptor or ligand unbound structures as annotated in the PDB

[c] Additional information: organism of the interacting proteins. In case of using only part of the unbound structure, the residue range is shown. d PDB IDs of the complexes between the receptor protein and all known PPI inhibitors. In italics, the complex of the inhibitor with the highest IC50 experimentally reported.

[d] PDB IDs of the complexes between the receptor protein and all known PPI inhibitors. In italics, the complex of the inhibitor with the highest IC50 experimentally reported.

Figure 8.3.1 Some examples illustrating the difficulties in the assessment of predicted PPI inhibitor pockets.

Top image shows the Positive predicted value (PPV) and coverage (COV) of best-scoring predicted pockets on TIMBAL/2P2I database. The plot shows the best-scoring predicted pocket for each case, with those considered a hit (by PPC) represented as * symbol. Bottom images show two examples of PPI inhibitor site prediction (general view and detailed inset). The protein is shown in ribbon blue; the predicted pocket is represented as pink surface with the predicted pocket atoms as orange sticks; and the real inhibitor is shown in black sticks. Bottom left: PPI inhibitor bound to Keap1 (PDB 3VNG), an example of unsuccessful prediction, with good coverage but low PPV, due to the prediction of a large pocket (red residues, with correctly predicted residues in magenta) that includes the real one (green residues, with correctly predicted residues in magenta). Bottom right: PPI inhibitor bound to BRD4 (PDB 3U5L) an example of an apparently unsuccessful prediction by PPC, which is in fact correct according to PPV and coverage (both > 40%).

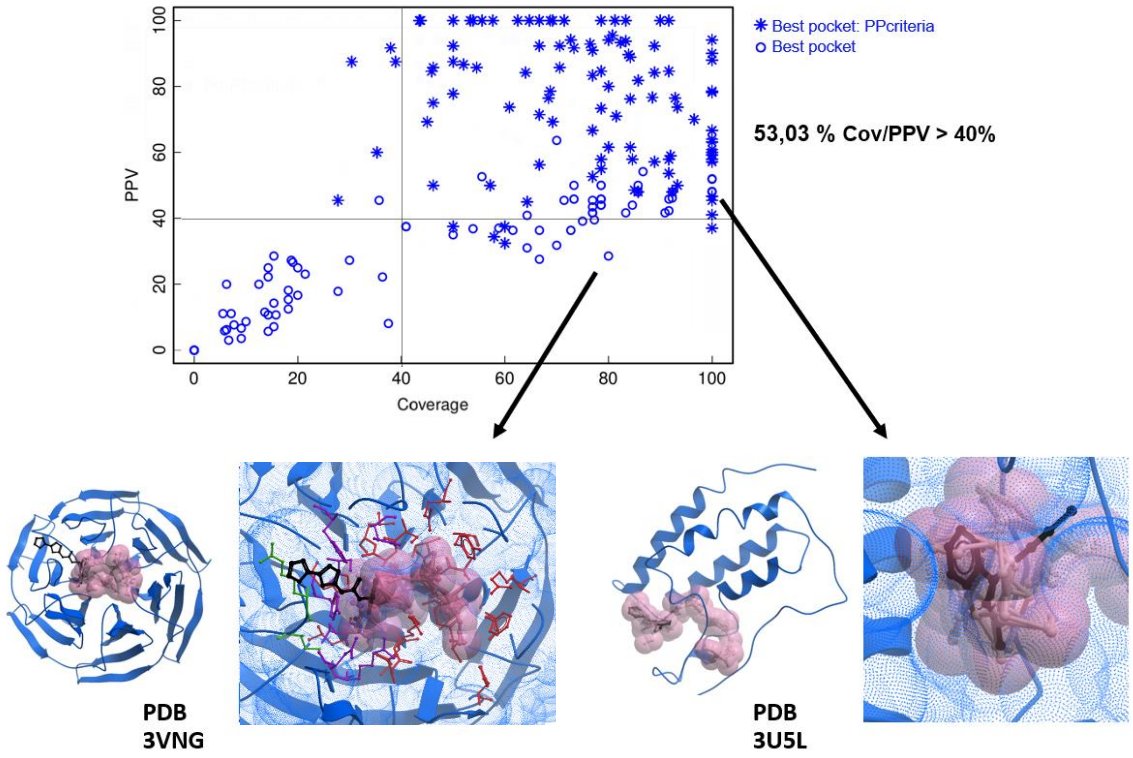


Figure 8.3.2 Best-scoring predicted pocket on the unbound proteins of benchmark set.

Unbound protein is represented in grey surface, with the best-scoring predicted pocket by *Fpocket* in orange. For comparison, the partner protein (or peptide) in the protein-protein complex is represented in blue ribbon, and a copy of each case in the same orientation is shown with the position of the known PPI inhibitor in green.

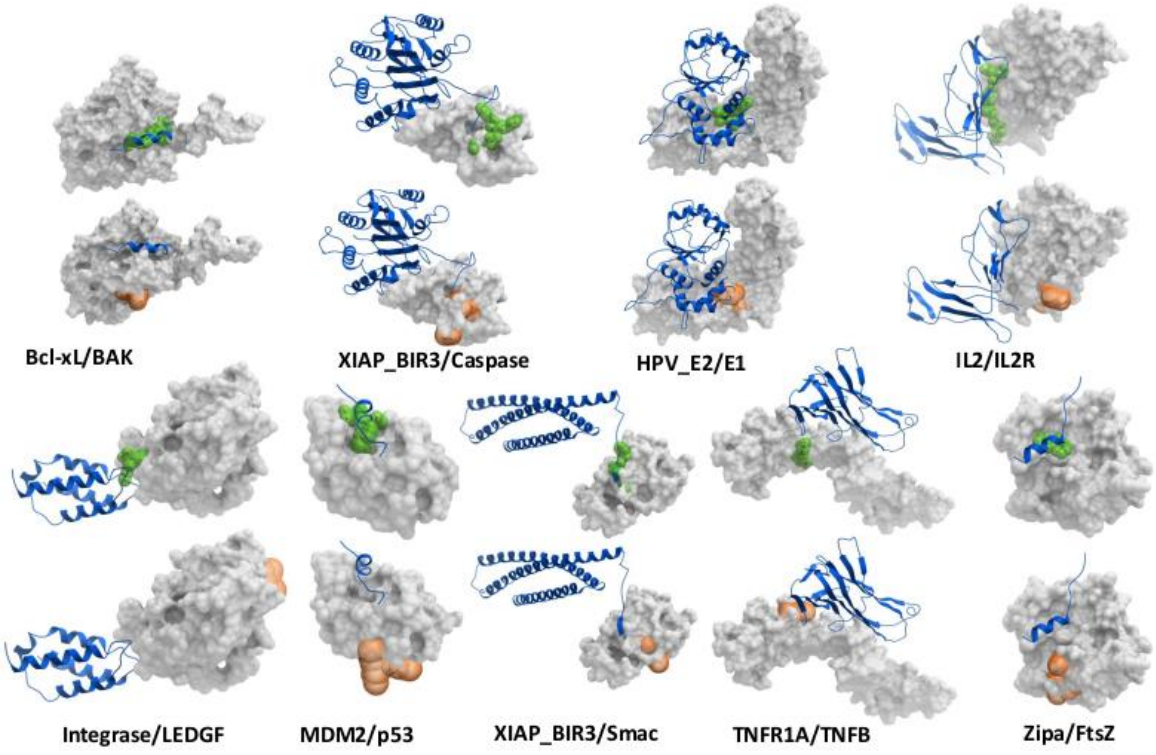


Figure 8.3.3 Assessment of the predicted pockets in different protein structures and conditions.

(A) PPV and coverage of the best-scoring pocket predicted on the unbound proteins, with respect to the reference pocket in the protein bound to the PPI inhibitor with the best IC50. (B) Best-scoring pocket predicted on the proteins taken from the structure of the protein bound to the PPI inhibitor with the best IC50. (C) Idem. as (A), but with respect to the reference pockets in the protein bound to all the different PPI inhibitors. (D) Best-scoring pocket predicted on the proteins taken from the protein-protein complex structure (but not forced to be located at the interface), with respect to the reference pockets in the protein bound to the different PPI inhibitors. (E) Best-scoring pocket predicted on the MD-based conformers generated from the unbound proteins, with respect to the reference pockets in the protein bound to the different PPI inhibitors. (F) Best-scoring pocket predicted on the MD-based conformers, considering only those located at the known protein-protein interface.

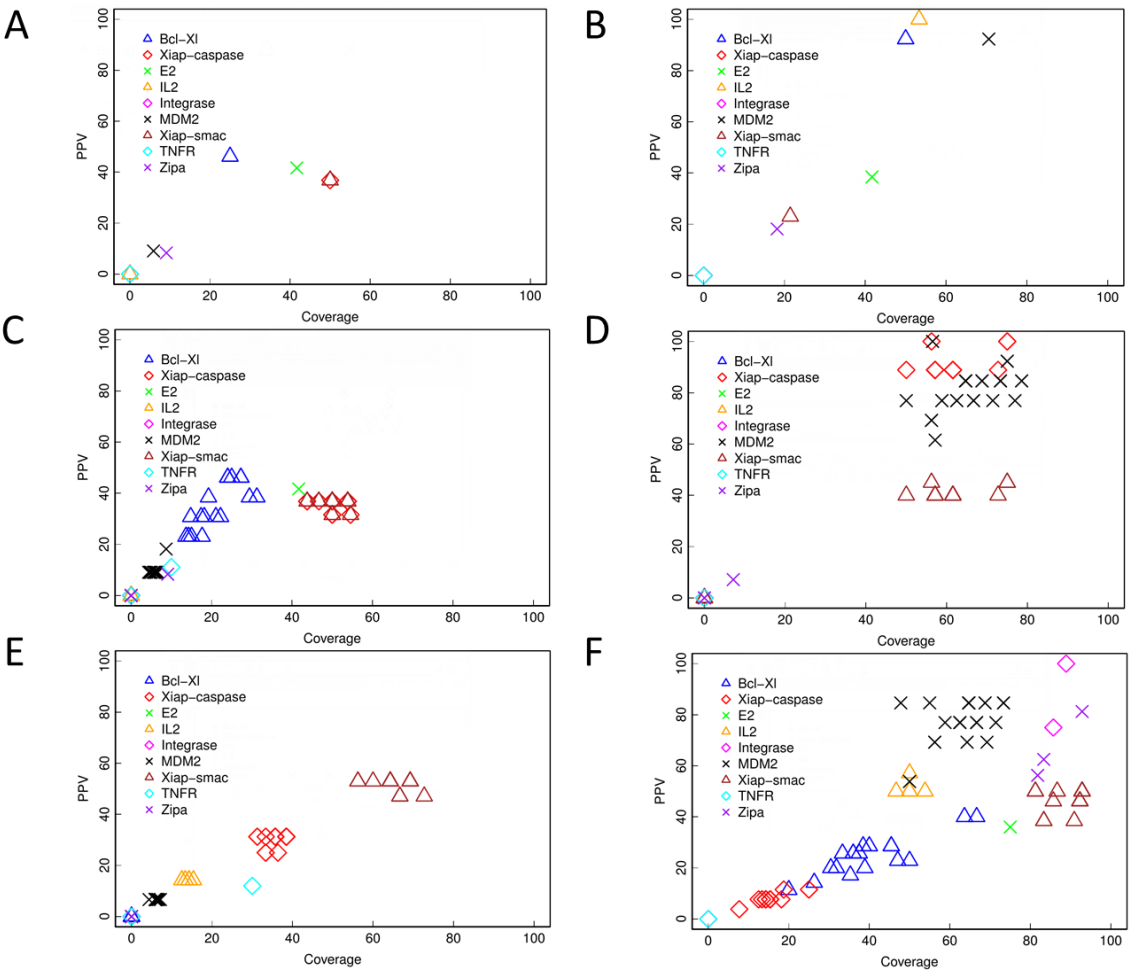


Figure 8.3.4 Evaluation of predicted pockets as a function of *Pscore*.
PPV values (with respect to the inhibitor with the best IC50) are shown for the best-scoring predicted pockets for each MD snapshot, as a function of their *Pscore*. Predicted pockets with coverage < 40% are shown in red, and those with coverage ≥ 40% are shown in blue. Filled circles are predicted pockets with ≥ 3 predicted hot-spots.

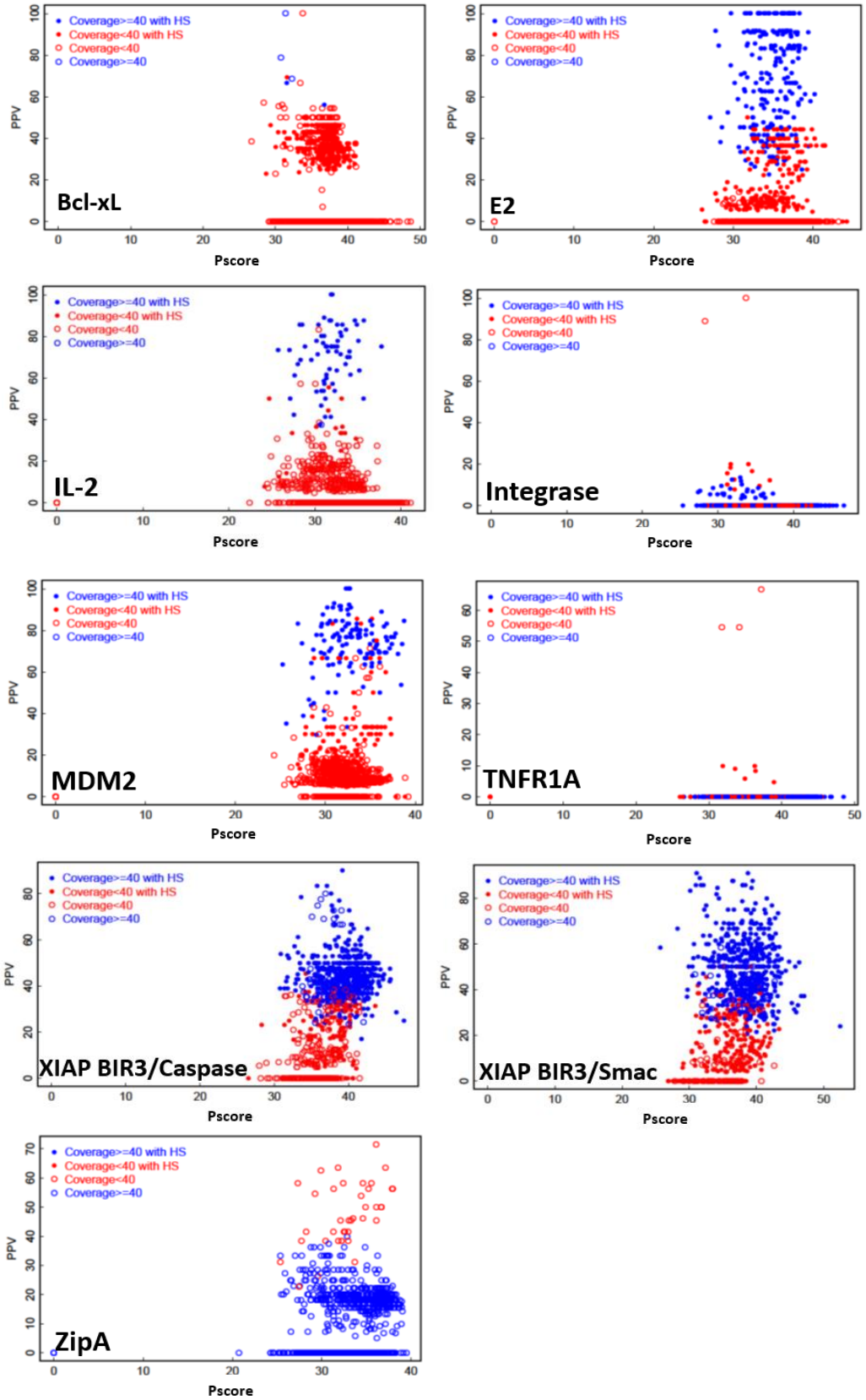


Figure 8.3.5. Evaluation of predicted pockets as a function of size.
0PPV values (with respect to the inhibitor with the best IC50) are shown for the best-scoring predicted pockets for each MD snapshot, as a function of their size (in number of residues). Predicted pockets with coverage < 40% are shown in red, and those with coverage ≥ 40% are shown in blue. Filled circles are predicted pockets with ≥ 3 predicted hot-spots.

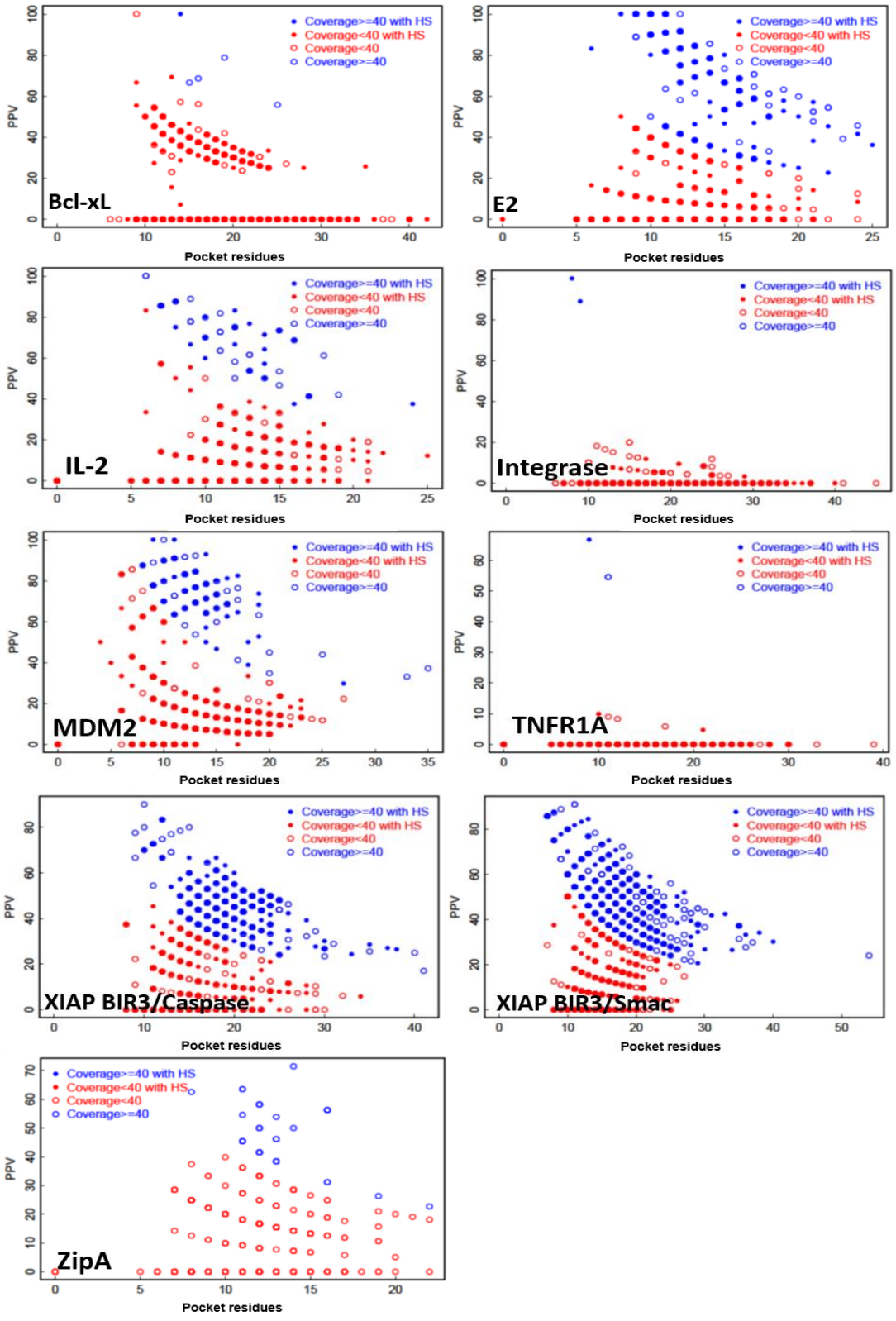
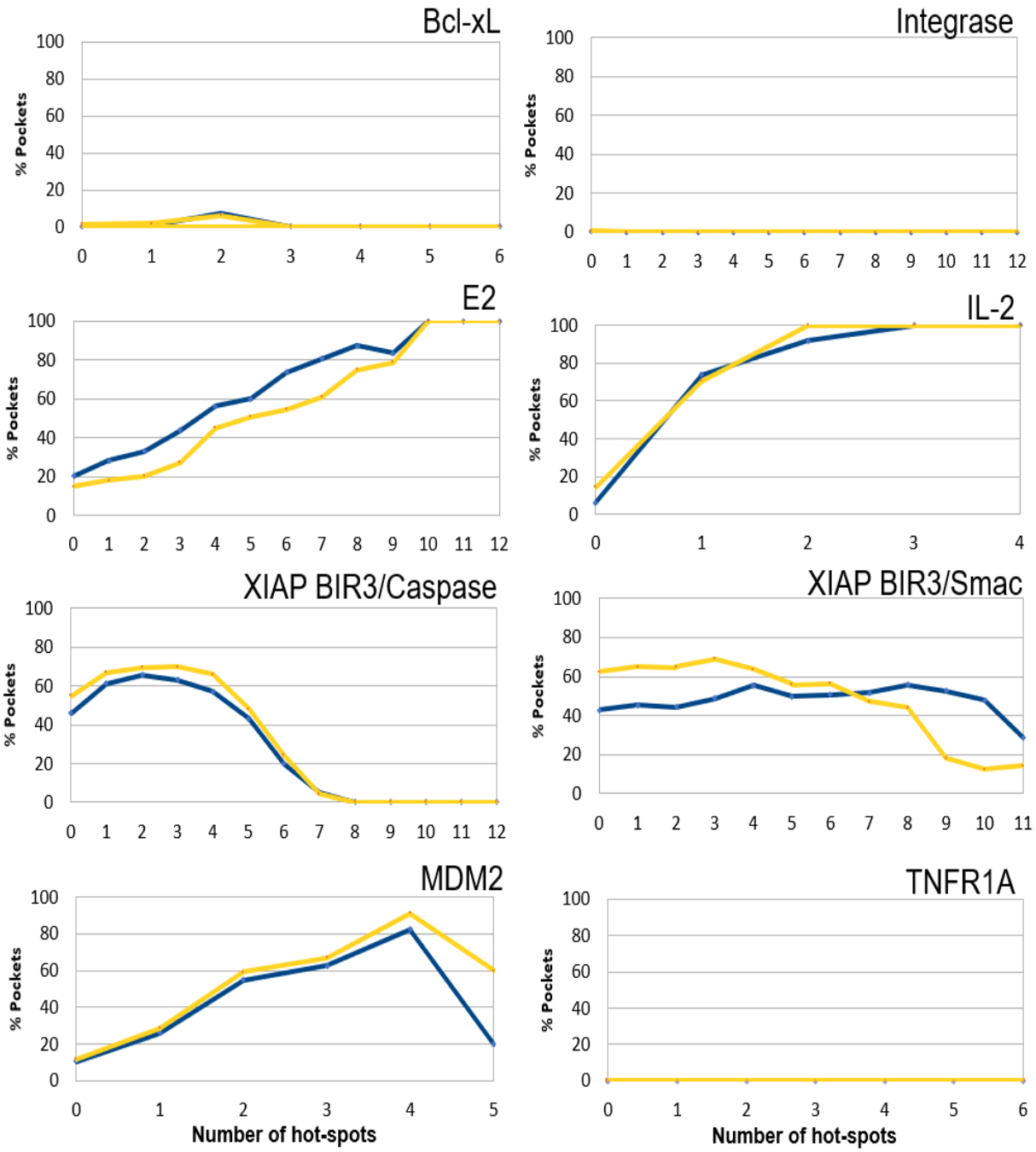


Figure 8.3.6 Predictive performance for best-scoring pockets from MD.
The percentage (success rate) of best-scoring pockets from MD snapshots that are ct (i.e. PPV and coverage ≥ 40%), as a function of the number of predicted hot-spots in the pocket is shown in blue. In yellow is shown the same success rates when pockets are filtered by druggability score >0.7%.



List of publications

Docking-based identification of small-molecule binding sites at protein-protein interfaces. Rosell M. and Fernández-Recio J. Comput Struct Biotechnol J. 2020 (sent)

Docking approaches for modeling multi-molecular assemblies. Rosell M. and Fernández-Recio J. Curr Opin Struct Biol. 2020, 64:59-65.

Structural consequences of disease-related mutations for protein-protein interactions. Rosell M. and Fernández-Recio J. Encyclopedia of Life Sciences. 2020: 1-9.

Integrative Modeling of Protein-Protein Interactions With pyDock for the New Docking Challenges. Rosell M., Rodríguez-Lumbreras L., Romero-Durana M., Jiménez-García B., Díaz L and Fernández-Recio J. Proteins 2020, 88(8):999-1008.

Blind Prediction of Homo- And Hetero-Protein Complexes: The CASP13-CAPRI Experiment. Lensink M. *et al*. Proteins 2019 Dec; 87(12), 1200-1221.

Modelling of protein complexes and molecular assemblies with pyDock. Rosell M, Rodríguez-Lumbreras L and Fernández-Recio J. Methods Mol Biol. 2020, 2165:175-198.

Structural and computational characterization of disease-related mutations involved in protein-protein interfaces. Navío D, Rosell M, Aguirre J, de la Cruz X and Fernández-Recio J. Int J Mol Sci. 2019 March; 20(7). 1583.

Hot-spot analysis for drug discovery targeting protein-protein interactions. Rosell M and Fernández-Recio J. Expert Opin Drug Discov. 2018 Apr;13(4):327-338.

Structural Prediction of Protein-Protein Interactions by Docking: Application to Biomedical Problems. Barradas-Bautista D, Rosell M, Pallara C and Fernández-Recio J. Adv Protein Chem Struct Biol. 2018;110:203-249.

Oral communications

Molecular characterization of genetic variants affecting protein-protein interactions. 28th November 2019. 2019 ICVV seminars, ICVV, Logroño.

Computational characterization of protein-protein interactions perturbed by pathological mutations can help to identify potential drug targets. 17th July 2019. 42nd Congress of SEBBM, Madrid.

Protein Interactions and Docking. 11th February 2019. Women in Science – Bioinfo4women, Barcelona Supercomputing Center, Barcelona.

Docking-based tools for discovery of protein-protein modulators. 15th May 2018. Severo Ochoa Research Seminars – Barcelona Supercomputing Center, Barcelona.

Characterization of pathological mutations affecting protein-protein interactions for drug discovery. 24th April 2018. 5th International BSC Doctoral Symposium – Barcelona Supercomputing Center, Barcelona.

List of posters

Mireia Rosell, Luis Rodríguez-Lumbreras, Miguel Romero, Brian Jiménez-García, Lucía Díaz, Juan Fernández-Recio. **Integrative modeling with pyDock for the new protein docking challenges in 7th CAPRI**. 7th CAPRI meeting 2019 (April). Hinxton, UK. [talk selected]

Mireia Rosell, Luis Rodríguez-Lumbreras, Miguel Romero, Lucía Díaz, Juan Fernández-Recio. **Combining pyDock *ab initio* docking and template-based modeling for the CASP13-CAPRI Challenge**. CASP13 meeting 2018 (December). Riviera Maya, Mexico.

Mireia Rosell, Juan Fernández-Recio. **Characterizing pathological mutations affecting protein-protein**

interactions for drug discovery. XL SEBBM Congress 2017 (October). Barcelona, Spain.

Mireia Rosell, Juan Fernández-Recio. **Characterization of Protein-Protein Interfaces and Identification of Transient Cavities for its Modulation.** 3rd International BSC Doctoral Symposium 2016 (May). Barcelona, Spain.

Mireia Rosell, Juan Fernández-Recio. **Docking-based tools for discovery of protein-protein modulators.** PROTEIN SOCIETY, 29th Annual Symposium 2015 (July). Barcelona, Spain.

Mireia Rosell, Juan Fernández-Recio. **Docking-based strategies for the design of small-molecule modulators of protein-protein interactions.** IRB Biomed 2016 (November). Barcelona, Spain.

Mireia Rosell-Oliveras, Solène Grosdidier, Juan Fernández-Recio. **Protein-Protein Docking for discovery of protein interaction inhibitors.** ICM User Group Meeting 2012 (April), San Diego, USA.

Bibliography

1. Barradas-Bautista D, Rosell M, Pallara C, Fernández-Recio J: **Structural Prediction of Protein-Protein Interactions by Docking: Application to Biomedical Problems**. *Adv Protein Chem Struct Biol* 2018, **110**:203-249.
2. Rosell M, Fernández-Recio J: **Hot-spot analysis for drug discovery targeting protein-protein interactions**. *Expert opinion on drug discovery* 2018, **13**(4):327-338.
3. Rosell M, Fernandez-Recio J: **Structural Consequences of Disease-Related Mutations for Protein–Protein Interactions**. In: *Encyclopedia of Life Sciences.* 2020: 1-9.
4. Rosell M, Rodríguez-Lumbreras LA, Fernández-Recio J: **Modeling of Protein Complexes and Molecular Assemblies with pyDock**. *Methods in molecular biology (Clifton, NJ)* 2020, **2165**:175-198.
5. Li X-Q, Du D: **Gene direction in living organisms**. *Scientific Reports* 2012, **2**(1):982.
6. Gilbert W: **Why genes in pieces?** *Nature* 1978, **271**(5645):501-501.
7. Butterworth PJ: **Lehninger: principles of biochemistry (4th edn) D. L. Nelson and M. C. Cox, W. H. Freeman & Co., New York, 1119 pp (plus 17 pp glossary), ISBN 0-7167-4339-6 (2004)**. *Cell Biochemistry and Function* 2005, **23**(4):293-294.
8. Levy ED, Erba EB, Robinson CV, Teichmann SA: **Assembly reflects evolution of protein complexes**. *Nature* 2008, **453**(7199):1262-1265.
9. Rolland T, Tasan M, Charloteaux B, Pevzner SJ, Zhong Q, Sahni N, Yi S, Lemmens I, Fontanillo C, Mosca R *et al*: **A proteome-scale map of the human interactome network**. *Cell* 2014, **159**(5):1212-1226.
10. Muir P, Li S, Lou S, Wang D, Spakowicz DJ, Salichos L, Zhang J, Weinstock GM, Isaacs F, Rozowsky J *et al*: **The real cost of sequencing: scaling computation to keep pace with data generation**. *Genome Biology* 2016, **17**(1):53.
11. Stenson PD, Ball EV, Mort M, Phillips AD, Shaw K, Cooper DN: **The Human Gene Mutation Database (HGMD) and Its Exploitation in the Fields of Personalized Genomics and Molecular Evolution**. *Current Protocols in Bioinformatics* 2012, **39**(1):1.13.11-11.13.20.
12. Hamosh A, Scott AF, Amberger J, Bocchini C, Valle D, McKusick VA: **Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders**. *Nucleic acids research* 2002, **30**(1):52-55.
13. Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, Maglott DR: **ClinVar: public archive of relationships among sequence variation and human phenotype**. *Nucleic acids research* 2014, **42**(Database issue):D980-D985.
14. Khera AV, Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH, Natarajan P, Lander ES, Lubitz SA, Ellinor PT *et al*: **Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations**. *Nature genetics* 2018, **50**(9):1219-1224.
15. Garralda E, Paz K, López-Casas PP, Jones S, Katz A, Kann LM, López-Rios F, Sarno F, Al-Shahrour F, Vasquez D *et al*: **Integrated next-generation sequencing and avatar mouse models for personalized cancer treatment**. *Clin Cancer Res* 2014, **20**(9):2476-2484.
16. Tennessen JA, Bigham AW, O'Connor TD, Fu W, Kenny EE, Gravel S, McGee S, Do R, Liu X, Jun G *et al*: **Evolution and functional impact of rare coding variation from deep sequencing of human exomes**. *Science (New York, NY)* 2012, **337**(6090):64-69.
17. Sahni N, Yi S, Taipale M, Fuxman Bass JI, Coulombe-Huntington J, Yang F, Peng J, Weile J, Karras GI, Wang Y *et al*: **Widespread macromolecular interaction perturbations in human genetic disorders**. *Cell* 2015, **161**(3):647-660.
18. Yang H, Wu J-J, Tang T, Liu K-D, Dai C: **CRISPR/Cas9-mediated genome editing efficiently creates specific mutations at multiple loci using one sgRNA in Brassica napus**. *Scientific reports* 2017, **7**(1):7489-7489.

19. Xue Y, Ankala A, Wilcox WR, Hegde MR: **Solving the molecular diagnostic testing conundrum for Mendelian disorders in the era of next-generation sequencing: single-gene, gene panel, or exome/genome sequencing**. *Genetics In Medicine* 2015, **17**:444.
20. Kiel C, Serrano L: **Structure-energy-based predictions and network modelling of RASopathy and cancer missense mutations**. *Molecular systems biology* 2014, **10**:727.
21. Wu X, Hasan MA, Chen JY: **Pathway and network analysis in proteomics**. *Journal of theoretical biology* 2014, **362**:44-52.
22. Hartwell LH, Hopfield JJ, Leibler S, Murray AW: **From molecular to modular cell biology**. *Nature* 1999, **402**(6761 Suppl):C47-52.
23. Pizzuti C, Rombo SE: **Algorithms and tools for protein-protein interaction networks clustering, with a special focus on population-based stochastic methods**. *Bioinformatics (Oxford, England)* 2014, **30**(10):1343-1352.
24. Goh KI, Cusick ME, Valle D, Childs B, Vidal M, Barabasi AL: **The human disease network**. *Proceedings of the National Academy of Sciences of the United States of America* 2007, **104**(21):8685-8690.
25. Mosca R, Ceol A, Aloy P: **Interactome3D: adding structural details to protein networks**. *Nature methods* 2013, **10**(1):47-53.
26. Orchard S, Ammari M, Aranda B, Breuza L, Briganti L, Broackes-Carter F, Campbell NH, Chavali G, Chen C, del-Toro N *et al*: **The MIntAct project--IntAct as a common curation platform for 11 molecular interaction databases**. *Nucleic acids research* 2014, **42**(Database issue):D358-D363.
27. Licata L, Briganti L, Peluso D, Perfetto L, Iannuccelli M, Galeota E, Sacco F, Palma A, Nardozza AP, Santonico E *et al*: **MINT, the molecular interaction database: 2012 update**. *Nucleic acids research* 2012, **40**(Database issue):D857-D861.
28. Salwinski L, Miller CS, Smith AJ, Pettit FK, Bowie JU, Eisenberg D: **The Database of Interacting Proteins: 2004 update**. *Nucleic acids research* 2004, **32**(Database issue):D449-D451.
29. Goll J, Rajagopala SV, Shiau SC, Wu H, Lamb BT, Uetz P: **MPIDB: the microbial protein interaction database**. *Bioinformatics (Oxford, England)* 2008, **24**(15):1743-1744.
30. Launay G, Salza R, Multedo D, Thierry-Mieg N, Ricard-Blum S: **MatrixDB, the extracellular matrix interaction database: updated content, a new navigator and expanded functionalities**. *Nucleic acids research* 2015, **43**(Database issue):D321-D327.
31. Breuer K, Foroushani AK, Laird MR, Chen C, Sribnaia A, Lo R, Winsor GL, Hancock REW, Brinkman FSL, Lynn DJ: **InnateDB: systems biology of innate immunity and beyond--recent updates and continuing curation**. *Nucleic acids research* 2013, **41**(Database issue):D1228-D1233.
32. Chatr-Aryamontri A, Breitkreutz BJ, Oughtred R, Boucher L, Heinicke S, Chen D, Stark C, Breitkreutz A, Kolas N, O'Donnell L *et al*: **The BioGRID interaction database: 2015 update**. *Nucleic acids research* 2015, **43**(Database issue):D470-478.
33. Isserlin R, El-Badrawi RA, Bader GD: **The Biomolecular Interaction Network Database in PSI-MI 2.5**. *Database : the journal of biological databases and curation* 2011, **2011**:baq037-baq037.
34. Keshava Prasad TS, Goel R, Kandasamy K, Keerthikumar S, Kumar S, Mathivanan S, Telikicherla D, Raju R, Shafreen B, Venugopal A *et al*: **Human Protein Reference Database--2009 update**. *Nucleic acids research* 2009, **37**(Database issue):D767-D772.
35. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP *et al*: **STRING v10: protein-protein interaction networks, integrated over the tree of life**. *Nucleic acids research* 2015, **43**(Database issue):D447-452.
36. Mosca R, Tenorio-Laranga J, Olivella R, Alcalde V, Ceol A, Soler-Lopez M, Aloy P: **dSysMap: exploring the edgetic role of disease mutations**. *Nature methods* 2015, **12**(3):167-168.

37. Jassal B, Matthews L, Viteri G, Gong C, Lorente P, Fabregat A, Sidiropoulos K, Cook J, Gillespie M, Haw R *et al*: **The reactome pathway knowledgebase**. *Nucleic acids research* 2020, **48**(D1):D498-D503.
38. Ghiassian SD, Menche J, Chasman DI, Giulianini F, Wang R, Ricchiuto P, Aikawa M, Iwata H, Müller C, Zeller T *et al*: **Endophenotype Network Models: Common Core of Complex Diseases**. *Scientific Reports* 2016, **6**(1):27414.
39. Wang X, Wei X, Thijssen B, Das J, Lipkin SM, Yu H: **Three-dimensional reconstruction of protein networks provides insight into human genetic disease**. *Nature biotechnology* 2012, **30**(2):159-164.
40. David A, Razali R, Wass MN, Sternberg MJ: **Protein-protein interaction sites are hot spots for disease-associated nonsynonymous SNPs**. *Human mutation* 2012, **33**(2):359-363.
41. David A, Sternberg MJ: **The Contribution of Missense Mutations in Core and Rim Residues of Protein-Protein Interfaces to Human Disease**. *Journal of molecular biology* 2015, **427**(17):2886-2898.
42. Mering Cv, Huynen M, Jaeggi D, Schmidt S, Bork P, Snel B: **STRING: a database of predicted functional associations between proteins**. *Nucleic acids research* 2003, **31**(1):258-261.
43. Aebersold R, Agar JN, Amster IJ, Baker MS, Bertozzi CR, Boja ES, Costello CE, Cravatt BF, Fenselau C, Garcia BA *et al*: **How many human proteoforms are there?** *Nat Chem Biol* 2018, **14**(3):206-214.
44. Venkatesan K, Rual J-F, Vazquez A, Stelzl U, Lemmens I, Hirozane-Kishikawa T, Hao T, Zenkner M, Xin X, Goh K-I *et al*: **An empirical framework for binary interactome mapping**. *Nature methods* 2009, **6**(1):83-90.
45. Stumpf MP, Thorne T, de Silva E, Stewart R, An HJ, Lappe M, Wiuf C: **Estimating the size of the human interactome**. *Proceedings of the National Academy of Sciences of the United States of America* 2008, **105**(19):6959-6964.
46. Szilagyi A, Zhang Y: **Template-based structure modeling of protein–protein interactions**. *Current Opinion in Structural Biology* 2014, **24**:10-23.
47. Negroni J, Mosca R, Aloy P: **Assessing the applicability of template-based protein docking in the twilight zone**. *Structure (London, England : 1993)* 2014, **22**(9):1356-1362.
48. Kundrotas PJ, Zhu Z, Janin J, Vakser IA: **Templates are available to model nearly all complexes of structurally characterized proteins**. *Proceedings of the National Academy of Sciences of the United States of America* 2012, **109**(24):9438-9441.
49. Porter KA, Desta I, Kozakov D, Vajda S: **What method to use for protein–protein docking?** *Current Opinion in Structural Biology* 2019, **55**:1-7.
50. Sievers F, Higgins DG: **Clustal Omega, accurate alignment of very large numbers of sequences**. *Methods in molecular biology (Clifton, NJ)* 2014, **1079**:105-116.
51. Notredame C, Higgins DG, Heringa J: **T-Coffee: A novel method for fast and accurate multiple sequence alignment**. *Journal of molecular biology* 2000, **302**(1):205-217.
52. Edgar RC: **MUSCLE: multiple sequence alignment with high accuracy and high throughput**. *Nucleic acids research* 2004, **32**(5):1792-1797.
53. Holm L: **DALI and the persistence of protein shape**. *Protein Science* 2020, **29**(1):128-140.
54. Russell RB, Barton GJ: **Multiple protein sequence alignment from tertiary structure comparison: assignment of global and residue confidence levels**. *Proteins* 1992, **14**(2):309-323.
55. Zhang Y, Skolnick J: **TM-align: a protein structure alignment algorithm based on the TM-score**. *Nucleic acids research* 2005, **33**(7):2302-2309.
56. Maiti R, Van Domselaar GH, Zhang H, Wishart DS: **SuperPose: a simple server for sophisticated structural superposition**. *Nucleic acids research* 2004, **32**(Web Server issue):W590-W594.

57. Gibrat JF, Madej T, Bryant SH: **Surprising similarities in structure comparison**. *Current opinion in structural biology* 1996, **6**(3):377-385.
58. Zhang Y: **I-TASSER server for protein 3D structure prediction**. *BMC bioinformatics* 2008, **9**(1):40.
59. Söding J, Biegert A, Lupas AN: **The HHpred interactive server for protein homology detection and structure prediction**. *Nucleic acids research* 2005, **33**(Web Server issue):W244-W248.
60. Šali A, Potterton L, Yuan F, van Vlijmen H, Karplus M: **Evaluation of comparative protein modeling by MODELLER**. *Proteins: Structure, Function, and Bioinformatics* 1995, **23**(3):318-326.
61. André I, Bradley P, Wang C, Baker D: **Prediction of the structure of symmetrical protein assemblies**. *Proceedings of the National Academy of Sciences of the United States of America* 2007, **104**(45):17656-17661.
62. Lyskov S, Gray JJ: **The RosettaDock server for local protein-protein docking**. *Nucleic acids research* 2008, **36**(Web Server issue):W233-W238.
63. Moal IH, Barradas-Bautista D, Jiménez-García B, Torchala M, van der Velde A, Vreven T, Weng Z, Bates PA, Fernández-Recio J: **IRaPPA: information retrieval based integration of biophysical models for protein assembly selection**. *Bioinformatics (Oxford, England)* 2017, **33**(12):1806-1813.
64. Vreven T, Moal IH, Vangone A, Pierce BG, Kastritis PL, Torchala M, Chaleil R, Jiménez-García B, Bates PA, Fernandez-Recio J *et al*: **Updates to the Integrated Protein–Protein Interaction Benchmarks: Docking Benchmark Version 5 and Affinity Benchmark Version 2**. *Journal of molecular biology* 2015, **427**(19):3031-3041.
65. Kuntz ID, Blaney JM, Oatley SJ, Langridge R, Ferrin TE: **A geometric approach to macromolecule-ligand interactions**. *Journal of molecular biology* 1982, **161**(2):269-288.
66. Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ: **PatchDock and SymmDock: servers for rigid and symmetric docking**. *Nucleic acids research* 2005, **33**:W363-W367.
67. Katchalski-Katzir E, Shariv I, Eisenstein M, Friesem AA, Aflalo C, Vakser IA: **Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques**. *Proceedings of the National Academy of Sciences of the United States of America* 1992, **89**(6):2195-2199.
68. Gabb HA, Jackson RM, Sternberg MJ: **Modelling protein docking using shape complementarity, electrostatics and biochemical information**. *Journal of molecular biology* 1997, **272**(1):106-120.
69. Chen R, Li L, Weng Z: **ZDOCK: An initial-stage protein-docking algorithm**. *Proteins* 2003, **52**(1):80-87.
70. Redington PK: **MOLFIT: A computer program for molecular superposition**. *Computers & Chemistry* 1992, **16**(3):217-222.
71. Ritchie DW, Kemp GJL: **Protein docking using spherical polar Fourier correlations**. *Proteins: Structure, Function, and Bioinformatics* 2000, **39**(2):178-194.
72. Garzon JI, Lopéz-Blanco JR, Pons C, Kovacs J, Abagyan R, Fernandez-Recio J, Chacon P: **FRODOCK: a new approach for fast rotational protein-protein docking**. *Bioinformatics (Oxford, England)* 2009, **25**(19):2544-2551.
73. Fernández-Recio J, Totrov M, Abagyan R: **Soft protein-protein docking in internal coordinates**. *Protein science : a publication of the Protein Society* 2002, **11**(2):280-291.
74. Fernandez-Recio J, Totrov M, Abagyan R: **ICM-DISCO docking by global energy optimization with fully flexible side-chains**. *Proteins* 2003, **52**.
75. Gray JJ, Moughon S, Wang C, Schueler-Furman O, Kuhlman B, Rohl CA, Baker D: **Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations**. *Journal of molecular biology* 2003, **331**.

76. Dominguez C, Boelens R, Bonvin AM: **HADDOCK: a protein-protein docking approach based on biochemical or biophysical information**. *Journal of the American Chemical Society* 2003, **125**(7):1731-1737.
77. Moal IH, Bates PA: **SwarmDock and the use of normal modes in protein-protein docking**. *International journal of molecular sciences* 2010, **11**(10):3623-3648.
78. Jiménez-García B, Roel-Touris J, Romero-Durana M, Vidal M, Jiménez-González D, Fernández-Recio J: **LightDock: a new multi-scale approach to protein–protein docking**. *Bioinformatics (Oxford, England)* 2018, **34**(1):49-55.
79. Nadaradjane AA, Quignot C, Traoré S, Andreani J, Guerois R: **Docking proteins and peptides under evolutionary constraints in Critical Assessment of PRediction of Interactions rounds 38 to 45**. *Proteins: Structure, Function, and Bioinformatics* 2019, **(*in press*)**.
80. Cheng TM, Blundell TL, Fernandez-Recio J: **pyDock: electrostatics and desolvation for effective scoring of rigid-body protein-protein docking**. *Proteins* 2007, **68**(2):503-515.
81. Pierce B, Weng Z: **ZRANK: reranking protein docking predictions with an optimized energy function**. *Proteins* 2007, **67**(4):1078-1086.
82. Pons C, Talavera D, de la Cruz X, Orozco M, Fernandez-Recio J: **Scoring by intermolecular pairwise propensities of exposed residues (SIPPER): a new efficient potential for protein-protein docking**. *Journal of chemical information and modeling* 2010, **51**(2):370-377.
83. Jiménez-García B, Pons C, Fernández-Recio J: **pyDockWEB: a web server for rigid-body protein–protein docking using electrostatics and desolvation scoring**. *Bioinformatics (Oxford, England)* 2013, **29**(13):1698-1699.
84. Andrusier N, Nussinov R, Wolfson HJ: **FireDock: fast interaction refinement in molecular docking**. *Proteins* 2007, **69**(1):139-159.
85. Lensink MF, Wodak SJ: **Docking and scoring protein interactions: CAPRI 2009**. *Proteins* 2010, **78**(15):3073-3084.
86. Lensink MF, Velankar S, Kryshtafovych A, Huang S-Y, Schneidman-Duhovny D, Sali A, Segura J, Fernandez-Fuentes N, Viswanath S, Elber R *et al*: **Prediction of homoprotein and heteroprotein complexes by protein docking and template-based modeling: A CASP-CAPRI experiment**. *Proteins* 2016, **84 Suppl 1**(Suppl Suppl 1):323-348.
87. Moal IH, Fernandez-Recio J: **SKEMPI: a Structural Kinetic and Energetic database of Mutant Protein Interactions and its use in empirical models**. *Bioinformatics (Oxford, England)* 2012, **28**(20):2600-2607.
88. Jankauskaite J, Jiménez-García B, Dapkunas J, Fernández-Recio J, Moal IH: **SKEMPI 2.0: an updated benchmark of changes in protein-protein binding energy, kinetics and thermodynamics upon mutation**. *Bioinformatics (Oxford, England)* 2018, **35**(3):462-469.
89. Moal IH, Jimenez-Garcia B, Fernandez-Recio J: **CCharPPI web server: computational characterization of protein-protein interactions from structure**. *Bioinformatics (Oxford, England)* 2015, **31**(1):123-125.
90. Kortemme T, Baker D: **A simple physical model for binding energy hot spots in protein-protein complexes**. *Proceedings of the National Academy of Sciences of the United States of America* 2002, **99**(22):14116-14121.
91. Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L: **The FoldX web server: an online force field**. *Nucleic acids research* 2005, **33**(Web Server issue):W382-388.
92. Neuvirth H, Raz R, Schreiber G: **ProMate: A Structure Based Prediction Program to Identify the Location of Protein–Protein Binding Sites**. *Journal of molecular biology* 2004, **338**(1):181-199.
93. Liang S, Zhang C, Liu S, Zhou Y: **Protein binding site prediction using an empirical scoring function**. *Nucleic acids research* 2006, **34**(13):3698-3707.

94. Chen H, Zhou H-X: **Prediction of interface residues in protein–protein complexes by a consensus neural network method: Test against NMR data**. *Proteins: Structure, Function, and Bioinformatics* 2005, **61**(1):21-35.
95. Fernández-Recio J, Totrov M, Abagyan R: **Identification of Protein–Protein Interaction Sites from Docking Energy Landscapes**. *Journal of molecular biology* 2004, **335**(3):843-865.
96. Grosdidier S, Fernandez-Recio J: **Identification of hot-spot residues in protein-protein interactions by computational docking**. *BMC bioinformatics* 2008, **9**:447.
97. Jiménez-García B, Bernadó P, Fernández-Recio J: **Structural Characterization of Protein-Protein Interactions with pyDockSAXS**. *Methods in molecular biology (Clifton, NJ)* 2020, **2112**:131-144.
98. Jimenez-Garcia B, Pons C, Svergun DI, Bernado P, Fernandez-Recio J: **pyDockSAXS: protein-protein complex structure by SAXS and computational docking**. *Nucleic acids research* 2015, **43**(W1):W356-361.
99. Pons C, D'Abramo M, Svergun DI, Orozco M, Bernadó P, Fernández-Recio J: **Structural Characterization of Protein–Protein Complexes by Integrating Computational Docking with Small-angle Scattering Data**. *Journal of molecular biology* 2010, **403**(2):217-230.
100. Chelliah V, Blundell TL, Fernandez-Recio J: **Efficient Restraints for Protein–Protein Docking by Comparison of Observed Amino Acid Substitution Patterns with those Predicted from Local Environment**. *Journal of molecular biology* 2006, **357**(5):1669-1682.
101. Pérez-Cano L, Romero-Durana M, Fernández-Recio J: **Structural and energy determinants in protein-RNA docking**. *Methods (San Diego, Calif)* 2017, **118-119**:163-170.
102. Rosell A, Meury M, Álvarez-Marimon E, Costa M, Pérez-Cano L, Zorzano A, Fernández-Recio J, Palacín M, Fotiadis D: **Structural bases for the interaction and stabilization of the human amino acid transporter LAT2 with its ancillary protein 4F2hc**. *Proceedings of the National Academy of Sciences of the United States of America* 2014, **111**(8):2966-2971.
103. Lucas M, Gaspar AH, Pallara C, Rojas AL, Fernández-Recio J, Machner MP, Hierro A: **Structural basis for the recruitment and activation of the Legionella phospholipase VipD by the host GTPase Rab5**. *Proceedings of the National Academy of Sciences* 2014, **111**(34):E3514.
104. Krissinel E, Henrick K: **Inference of macromolecular assemblies from crystalline state**. *Journal of molecular biology* 2007, **372**(3):774-797.
105. Bliven S, Lafita A, Parker A, Capitani G, Duarte JM: **Automated evaluation of quaternary structures from protein crystals**. *PLOS Computational Biology* 2018, **14**(4):e1006104.
106. Xu Q, Dunbrack RL, Jr: **The protein common interface database (ProtCID)—a comprehensive database of interactions of homologous proteins in multiple crystal forms**. *Nucleic acids research* 2010, **39**(suppl_1):D761-D770.
107. Dey S, Ritchie DW, Levy ED: **PDB-wide identification of biological assemblies from conserved quaternary structure geometry**. *Nature methods* 2018, **15**(1):67-72.
108. Dapkūnas J, Timinskas A, Olechnovič K, Margelevičius M, Dičiūnas R, Venclovas Č: **The PPI3D web server for searching, analyzing and modeling protein–protein interactions in the context of 3D structures**. *Bioinformatics (Oxford, England)* 2017, **33**(6):935-937.
109. Clackson T, Wells JA: **A hot spot of binding energy in a hormone-receptor interface**. *Science (New York, NY)* 1995, **267**(5196):383-386.
110. Bogan AA, Thorn KS: **Anatomy of hot spots in protein interfaces**. *Journal of molecular biology* 1998, **280**(1):1-9.
111. Guerois R, Nielsen JE, Serrano L: **Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations**. *Journal of molecular biology* 2002, **320**(2):369-387.

112. Assi SA, Tanaka T, Rabbitts TH, Fernandez-Fuentes N: **PCRPi: Presaging Critical Residues in Protein interfaces, a new computational tool to chart hot spots in protein interfaces**. *Nucleic acids research* 2010, **38**(6):e86.
113. Bradford JR, Westhead DR: **Improved prediction of protein-protein binding sites using a support vector machines approach**. *Bioinformatics (Oxford, England)* 2005, **21**(8):1487-1494.
114. Koes DR, Camacho CJ: **PocketQuery: protein-protein interaction inhibitor starting points from protein-protein interaction structure**. *Nucleic acids research* 2012, **40**(Web Server issue):W387-392.
115. Pavelka A, Chovancova E, Damborsky J: **HotSpot Wizard: a web server for identification of hot spots in protein engineering**. *Nucleic acids research* 2009, **37**(Web Server issue):W376-383.
116. Kruger DM, Ignacio Garzon J, Chacon P, Gohlke H: **DrugScorePPI knowledge-based potentials used as scoring and objective function in protein-protein docking**. *PloS one* 2014, **9**(2):e89466.
117. Geppert T, Hoy B, Wessler S, Schneider G: **Context-based identification of protein-protein interfaces and "hot-spot" residues**. *Chem Biol* 2011, **18**(3):344-353.
118. Shingate P, Manoharan M, Sukhwal A, Sowdhamini R: **ECMIS: computational approach for the identification of hotspots at protein-protein interfaces**. In: *BMC bioinformatics.* vol. 15; 2014: 303.
119. Segura J, Fernandez-Fuentes N: **PCRPi-DB: a database of computationally annotated hot spots in protein interfaces**. *Nucleic acids research* 2011, **39**(Database issue):D755-760.
120. Levy ED: **A simple definition of structural regions in proteins and its use in analyzing interface evolution**. *Journal of molecular biology* 2010, **403**(4):660-670.
121. Wells JA, McClendon CL: **Reaching for high-hanging fruit in drug discovery at protein-protein interfaces**. *Nature* 2007, **450**(7172):1001-1009.
122. Cukuroglu E, Gursoy A, Nussinov R, Keskin O: **Non-redundant unique interface structures as templates for modeling protein interactions**. *PloS one* 2014, **9**(1):e86738.
123. Engin HB, Guney E, Keskin O, Oliva B, Gursoy A: **Integrating structure to protein-protein interaction networks that drive metastasis to brain and lung in breast cancer**. *PloS one* 2013, **8**(11):e81035-e81035.
124. Garcia-Seisdedos H, Empereur-Mot C, Elad N, Levy ED: **Proteins evolve on the edge of supramolecular self-assembly**. *Nature* 2017, **548**(7666):244-247.
125. Empereur-Mot C, Garcia-Seisdedos H, Elad N, Dey S, Levy ED: **Geometric description of self-interaction potential in symmetric protein complexes**. *Scientific Data* 2019, **6**(1):64.
126. Halakou F, Kilic ES, Cukuroglu E, Keskin O, Gursoy A: **Enriching Traditional Protein-protein Interaction Networks with Alternative Conformations of Proteins**. *Scientific Reports* 2017, **7**(1):7180.
127. Kulkarni MV, Tettamanzi MC, Murphy JW, Keeler C, Myszka DG, Chayen NE, Lolis EJ, Hodsdon ME: **Two independent histidines, one in human prolactin and one in its receptor, are critical for pH-dependent receptor recognition and activation**. *J Biol Chem* 2010, **285**(49):38524-38533.
128. Romero-Durana M, Jiménez-García B, Fernández-Recio J: **pyDockEneRes: per-residue decomposition of protein–protein docking energy**. *Bioinformatics (Oxford, England)* 2020.
129. Pires DEV, Ascher DB, Blundell TL: **mCSM: predicting the effects of mutations in proteins using graph-based signatures**. *Bioinformatics (Oxford, England)* 2014, **30**(3):335-342.
130. Geng C, Vangone A, Folkers GE, Xue LC, Bonvin AMJJ: **iSEE: Interface structure, evolution, and energy-based machine learning predictor of binding affinity changes upon mutations**. *Proteins* 2019, **87**(2):110-119.

131. Li M, Simonetti FL, Goncearenco A, Panchenko AR: **MutaBind estimates and interprets the effects of sequence variants on protein-protein interactions**. *Nucleic acids research* 2016, **44**(W1):W494-W501.
132. Sunaga N, Kaira K, Tomizawa Y, Shimizu K, Imai H, Takahashi G, Kakegawa S, Ohtaki Y, Nagashima T, Kasahara N *et al*: **Clinicopathological and prognostic significance of interleukin-8 expression and its relationship to KRAS mutation in lung adenocarcinoma**. *Br J Cancer* 2014, **110**(8):2047-2053.
133. Wilson CGM, Arkin MR: **Small-molecule inhibitors of IL-2/IL-2R: lessons learned and applied**. *Curr Top Microbiol Immunol* 2011, **348**:25-59.
134. Gershell LJ, Atkins JH: **A brief history of novel drug discovery technologies**. *Nat Rev Drug Discov* 2003, **2**(4):321-327.
135. Morrison C: **Fresh from the biotech pipeline-2016**. *Nature biotechnology* 2017, **35**(2):108-112.
136. Santos R, Ursu O, Gaulton A, Bento AP, Donadi RS, Bologa CG, Karlsson A, Al-Lazikani B, Hersey A, Oprea TI *et al*: **A comprehensive map of molecular drug targets**. *Nat Rev Drug Discov* 2017, **16**(1):19-34.
137. Taboureau O, Baell JB, Fernández-Recio J, al. e: **Established and emerging trends in computational drug discovery in the structural genomics era.** *Chem Biol* 2012, **19**:29-41.
138. Zanzoni A, Soler-Lopez M, Aloy P: **A network medicine approach to human disease**. *FEBS Lett* 2009, **583**(11):1759-1765.
139. Vidal M, Cusick ME, Barabasi AL: **Interactome networks and human disease**. *Cell* 2011, **144**(6):986-998.
140. Scott DE, Bayly AR, Abell C, Skidmore J: **Small molecules, big targets: drug discovery faces the protein-protein interaction challenge**. *Nat Rev Drug Discov* 2016, **15**(8):533-550.
141. Pérot S, Regad L, Reynès C, Spérandio O, Miteva MA, Villoutreix BO, Camproux A-C: **Insights into an original pocket-ligand pair classification: a promising tool for ligand profile prediction**. *PloS one* 2013, **8**(6):e63730-e63730.
142. Zhang M-Q, Wilkinson B: **Drug discovery beyond the 'rule-of-five'**. *Curr Opin Biotechnol* 2007, **18**(6):478-488.
143. Eyrisch S, Helms V: **Transient pockets on protein surfaces involved in protein-protein interaction**. *J Med Chem* 2007, **50**(15):3457-3464.
144. Le Guilloux V, Schmidtke P, Tuffery P: **Fpocket: an open source platform for ligand pocket detection**. *BMC bioinformatics* 2009, **10**:168.
145. Brady GP, Jr., Stouten PF: **Fast prediction and visualization of protein binding pockets with PASS**. *Journal of computer-aided molecular design* 2000, **14**(4):383-401.
146. Laurie AT, Jackson RM: **Q-SiteFinder: an energy-based method for the prediction of protein-ligand binding sites**. *Bioinformatics (Oxford, England)* 2005, **21**(9):1908-1916.
147. Degac J, Winter U, Helms V: **Graph-Based Clustering of Predicted Ligand-Binding Pockets on Protein Surfaces**. *Journal of chemical information and modeling* 2015, **55**(9):1944-1952.
148. Li H, Liu A, Zhao Z, Xu Y, Lin J, Jou D, Li C: **Fragment-based drug design and drug repositioning using multiple ligand simultaneous docking (MLSD): identifying celecoxib and template compounds as novel inhibitors of signal transducer and activator of transcription 3 (STAT3)**. *Journal of medicinal chemistry* 2011, **54**(15):5592-5596.
149. Fuller JC, Burgoyne NJ, Jackson RM: **Predicting druggable binding sites at the protein–protein interface**. *Drug Discovery Today* 2009, **14**(3):155-161.
150. Arkin MR, Tang Y, Wells JA: **Small-molecule inhibitors of protein-protein interactions: progressing toward the reality**. *Chemistry & biology* 2014, **21**(9):1102-1114.
151. Kozakov D, Hall DR, Chuang G-Y, Cencic R, Brenke R, Grove LE, Beglov D, Pelletier J, Whitty A, Vajda S: **Structural conservation of druggable hot spots in protein-protein interfaces**.

Proceedings of the National Academy of Sciences of the United States of America 2011, **108**(33):13528-13533.

152. Jubb H, Blundell TL, Ascher DB: **Flexibility and small pockets at protein-protein interfaces: New insights into druggability**. *Progress in biophysics and molecular biology* 2015, **119**(1):2-9.
153. London N, Raveh B, Schueler-Furman O: **Druggable protein-protein interactions--from hot spots to hot segments**. *Current opinion in chemical biology* 2013, **17**(6):952-959.
154. Guo W, Wisniewski JA, Ji H: **Hot spot-based design of small-molecule inhibitors for protein–protein interactions**. *Bioorganic & Medicinal Chemistry Letters* 2014, **24**(11):2546-2554.
155. Metz A, Pfleger C, Kopitz H, Pfeiffer-Marek S, Baringhaus K-H, Gohlke H: **Hot spots and transient pockets: predicting the determinants of small-molecule binding to a protein-protein interface**. *Journal of chemical information and modeling* 2012, **52**(1):120-133.
156. Geppert T, Bauer S, Hiss JA, Conrad E, Reutlinger M, Schneider P, Weisel M, Pfeiffer B, Altmann K-H, Waibler Z *et al*: **Immunosuppressive small molecule discovered by structure-based virtual screening for inhibitors of protein-protein interactions**. *Angew Chem Int Ed Engl* 2012, **51**(1):258-261.
157. Winter A, Higueruelo AP, Marsh M, Sigurdardottir A, Pitt WR, Blundell TL: **Biophysical and computational fragment-based approaches to targeting protein-protein interactions: applications in structure-guided drug discovery**. *Q Rev Biophys* 2012, **45**(4):383-426.
158. Bai F, Morcos F, Cheng RR, Jiang H, Onuchic JN: **Elucidating the druggable interface of protein-protein interactions using fragment docking and coevolutionary analysis**. *Proceedings of the National Academy of Sciences of the United States of America* 2016, **113**(50):E8051-E8058.
159. Li H, Kasam V, Tautermann CS, Seeliger D, Vaidehi N: **Computational method to identify druggable binding sites that target protein-protein interactions**. *Journal of chemical information and modeling* 2014, **54**(5):1391-1400.
160. Cimermancic P, Weinkam P, Rettenmaier TJ, Bichmann L, Keedy DA, Woldeyes RA, Schneidman-Duhovny D, Demerdash ON, Mitchell JC, Wells JA *et al*: **CryptoSite: Expanding the Druggable Proteome by Characterization and Prediction of Cryptic Binding Sites**. *Journal of molecular biology* 2016, **428**(4):709-719.
161. Schmidtke P, Bidon-Chanal A, Luque FJ, Barril X: **MDpocket: open-source cavity detection and characterization on molecular dynamics trajectories**. *Bioinformatics (Oxford, England)* 2011, **27**(23):3276-3285.
162. Petrek M, Otyepka M, Banás P, Kosinová P, Koca J, Damborský J: **CAVER: a new tool to explore routes from protein clefts, pockets and cavities**. *BMC bioinformatics* 2006, **7**:316-316.
163. Makley LN, Gestwicki JE: **Expanding the number of 'druggable' targets: non-enzymes and protein-protein interactions**. *Chem Biol Drug Des* 2013, **81**(1):22-32.
164. Higueruelo AP, Schreyer A, Bickerton GRJ, Pitt WR, Groom CR, Blundell TL: **Atomic interactions and profile of small molecules disrupting protein-protein interfaces: the TIMBAL database**. *Chem Biol Drug Des* 2009, **74**(5):457-467.
165. Basse MJ, Betzi S, Bourgeas R, Bouzidi S, Chetrit B, Hamon V, Morelli X, Roche P: **2P2Idb: a structural database dedicated to orthosteric modulation of protein-protein interactions**. *Nucleic acids research* 2013, **41**(Database issue):D824-D827.
166. Labbé CM, Laconde G, Kuenemann MA, Villoutreix BO, Sperandio O: **iPPI-DB: a manually curated and interactive database of small non-peptide inhibitors of protein-protein interactions**. *Drug discovery today* 2013, **18**(19-20):958-968.
167. Labbé CM, Kuenemann MA, Zarzycka B, Vriend G, Nicolaes GAF, Lagorce D, Miteva MA, Villoutreix BO, Sperandio O: **iPPI-DB: an online database of modulators of protein-protein interactions**. *Nucleic acids research* 2016, **44**(D1):D542-D547.

168. Shin W-H, Christoffer CW, Kihara D: **In silico structure-based approaches to discover protein-protein interaction-targeting drugs**. *Methods (San Diego, Calif)* 2017, **131**:22-32.
169. Vellani V, Giacomoni C: **Gabapentin Inhibits Protein Kinase C Epsilon Translocation in Cultured Sensory Neurons with Additive Effects When Coapplied with Paracetamol (Acetaminophen)**. *ScientificWorldJournal* 2017, **2017**:3595903-3595903.
170. Lensink MF, Brysbaert G, Nadzirin N, Velankar S, Chaleil RAG, Gerguri T, Bates PA, Laine E, Carbone A, Grudinin S *et al*: **Blind prediction of homo- and hetero-protein complexes: The CASP13-CAPRI experiment**. *Proteins: Structure, Function, and Bioinformatics* 2019, **87**(12):1200-1221.
171. Rosell M, Rodríguez-Lumbreras LA, Romero-Durana M, Jiménez-García B, Díaz L, Fernández-Recio J: **Integrative modeling of protein-protein interactions with pyDock for the new docking challenges**. *Proteins: Structure, Function, and Bioinformatics* 2020, **88**(8):999-1008.
172. Rosell M, Fernández-Recio J: **Docking approaches for modeling multi-molecular assemblies**. *Current opinion in structural biology* 2020, **64**:59-65.
173. Méndez R, Leplae R, De Maria L, Wodak SJ: **Assessment of blind predictions of protein–protein interactions: Current status of docking methods**. *Proteins: Structure, Function, and Bioinformatics* 2003, **52**(1):51-67.
174. Lensink MF, Méndez R, Wodak SJ: **Docking and scoring protein complexes: CAPRI 3rd Edition**. *Proteins: Structure, Function, and Bioinformatics* 2007, **69**(4):704-718.
175. Lensink MF, Velankar S, Baek M, Heo L, Seok C, Wodak SJ: **The challenge of modeling protein assemblies: the CASP12-CAPRI experiment**. *Proteins: Structure, Function, and Bioinformatics* 2018, **86**(S1):257-273.
176. Lensink MF, Nadzirin N, Velankar S, Wodak SJ: **Modeling protein-protein, protein-peptide, and protein-oligosaccharide complexes: CAPRI 7th edition**. *Proteins: Structure, Function, and Bioinformatics* 2020, **88**(8):916-938.
177. Sali A, Blundell TL: **Comparative protein modelling by satisfaction of spatial restraints**. *J Mol Biol* 1993, **234**(3):779-815.
178. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool**. *Journal of molecular biology* 1990, **215**(3):403-410.
179. Shen MY, Sali A: **Statistical potential for assessment and prediction of protein structures**. *Protein Sci* 2006, **15**(11):2507-2524.
180. Roy A, Kucukural A, Zhang Y: **I-TASSER: a unified platform for automated protein structure and function prediction**. *Nature Protocols* 2010, **5**:725.
181. Chen R, Weng Z: **A novel shape complementarity scoring function for protein-protein docking**. *Proteins* 2003, **51**(3):397-408.
182. Grosdidier S, Pons C, Solernou A, Fernandez-Recio J: **Prediction and scoring of docking poses with pyDock**. *Proteins* 2007, **69**(4):852-858.
183. Doruker P, Atilgan AR, Bahar I: **Dynamics of proteins predicted by molecular dynamics simulations and analytical approaches: application to alpha-amylase inhibitor**. *Proteins* 2000, **40**(3):512-524.
184. Liu S, Zhang C, Zhou H, Zhou Y: **A physical reference state unifies the structure-derived potential of mean force for protein folding and binding**. *Proteins* 2004, **56**(1):93-101.
185. Desmet J, Verstraete K, Bloch Y, Lorent E, Wen Y, Devreese B, Vandenbroucke K, Loverix S, Hettmann T, Deroo S *et al*: **Structural basis of IL-23 antagonism by an Alphabody protein scaffold**. *Nature Communications* 2014, **5**:5237.
186. Kamishikiryo J, Fukuhara H, Okabe Y, Kuroki K, Maenaka K: **Molecular basis for LLT1 protein recognition by human CD161 protein (NKRP1A/KLRB1)**. *J Biol Chem* 2011, **286**(27):23823-23830.
187. Kita S, Matsubara H, Kasai Y, Tamaoki T, Okabe Y, Fukuhara H, Kamishikiryo J, Krayukhina E, Uchiyama S, Ose T *et al*: **Crystal structure of extracellular domain of human lectin-like**

transcript 1 (LLT1), the ligand for natural killer receptor-P1A. *European Journal of Immunology* 2015, **45**(6):1605-1613.
188. Javaheri A, Kruse T, Moonens K, Mejías-Luque R, Debraekeleer A, Asche CI, Tegtmeyer N, Kalali B, Bach NC, Sieber SA *et al*: **Helicobacter pylori adhesin HopQ engages in a virulence-enhancing interaction with human CEACAMs**. *Nature Microbiology* 2016, **2**:16189.
189. Cheng TM, Blundell TL, Fernandez-Recio J: **Structural assembly of two-domain proteins by rigid-body docking**. *BMC bioinformatics* 2008, **9**:441.
190. Ekiert DC, Bhabha G, Isom GL, Greenan G, Ovchinnikov S, Henderson IR, Cox JS, Vale RD: **Architectures of Lipid Transport Systems for the Bacterial Outer Membrane**. *Cell* 2017, **169**(2):273-285.e217.
191. Theodoridis SK, Konstantinos: **Pattern Recognition**. London (England): Academic Press; 1999.
192. Pons C, Solernou A, Perez-Cano L, Grosdidier S, Fernandez-Recio J: **Optimization of pyDock for the new CAPRI challenges: Docking of homology-based models, domain-domain assembly and protein-RNA binding**. *Proteins* 2010, **78**(15):3182-3188.
193. Case D, Darden T, Cheatham III T, Simmerling C, Wang J, Duke R, Luo R, Walker R, Zhang W, Merz K: **AMBER 12; University of California: San Francisco, 2012**. *There is no corresponding record for this reference[Google Scholar]* 2010:1-826.
194. Case D, Cerutti DS, Cheatham T, Darden T, Duke R, Giese TJ, Gohlke H, Götz A, Greene D, Homeyer N *et al*: **Amber 2017, University of California, San Francisco**; 2017.
195. Cheatham TE, 3rd, Cieplak P, Kollman PA: **A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat**. *J Biomol Struct Dyn* 1999, **16**(4):845-862.
196. Pallara C, Jimenez-Garcia B, Perez-Cano L, Romero-Durana M, Solernou A, Grosdidier S, Pons C, Moal IH, Fernandez-Recio J: **Expanding the frontiers of protein-protein modeling: from docking and scoring to binding affinity predictions and other challenges**. *Proteins* 2013, **81**(12):2192-2200.
197. Gasteiger J, Marsili M: **A new model for calculating atomic charges in molecules**. *Tetrahedron Letters* 1978, **19**(34):3181-3184.
198. Ruiz-Carmona S, Alvarez-Garcia D, Foloppe N, Garmendia-Doval AB, Juhos S, Schmidtke P, Barril X, Hubbard RE, Morley SD: **rDock: A Fast, Versatile and Open Source Program for Docking Ligands to Proteins and Nucleic Acids**. *PLOS Computational Biology* 2014, **10**(4):e1003571.
199. Jakalian A, Jack DB, Bayly CI: **Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation**. *Journal of Computational Chemistry* 2002, **23**(16):1623-1641.
200. Alhassid A, Ben-David A, Tabachnikov O, Libster D, Naveh E, Zolotnitsky G, Shoham Y, Shoham G: **Crystal structure of an inverting GH 43 1,5-α-L-arabinanase from Geobacillus stearothermophilus complexed with its substrate**. *Biochemical Journal* 2009, **422**(1):73-82.
201. Quignot C, Rey J, Yu J, Tufféry P, Guerois R, Andreani J: **InterEvDock2: an expanded server for protein docking using evolutionary and biological information from homology models and multimeric inputs**. *Nucleic acids research* 2018, **46**(W1):W408-W416.
202. Olechnovič K, Venclovas Č: **VoroMQA: Assessment of protein structure quality using interatomic contact areas**. *Proteins: Structure, Function, and Bioinformatics* 2017, **85**(6):1131-1145.
203. Oliva R, Vangone A, Cavallo L: **Ranking multiple docking solutions based on the conservation of inter-residue contacts**. *Proteins: Structure, Function, and Bioinformatics* 2013, **81**(9):1571-1584.

204. Chermak E, Petta A, Serra L, Vangone A, Scarano V, Cavallo L, Oliva R: **CONSRANK: a server for the analysis, comparison and ranking of docking models based on inter-residue contacts**. *Bioinformatics (Oxford, England)* 2015, **31**(9):1481-1483.
205. Chermak E, De Donato R, Lensink MF, Petta A, Serra L, Scarano V, Cavallo L, Oliva R: **Introducing a Clustering Step in a Consensus Approach for the Scoring of Protein-Protein Docking Models**. *PloS one* 2016, **11**(11):e0166460-e0166460.
206. Yan Y, Zhang D, Zhou P, Li B, Huang S-Y: **HDOCK: a web server for protein-protein and protein-DNA/RNA docking based on a hybrid strategy**. *Nucleic acids research* 2017, **45**(W1):W365-W373.
207. Yan Y, Huang S-Y: **Pushing the accuracy limit of shape complementarity for protein-protein docking**. *BMC bioinformatics* 2018, **20**(25):696.
208. Marze NA, Roy Burman SS, Sheffler W, Gray JJ: **Efficient flexible backbone protein–protein docking for challenging targets**. *Bioinformatics (Oxford, England)* 2018, **34**(20):3461-3469.
209. Nadalin F, Carbone A: **Protein-protein interaction specificity is captured by contact preferences and interface composition**. *Bioinformatics (Oxford, England)* 2018, **34**(3):459-468.
210. Dequeker C, Laine E, Carbone A: **Decrypting protein surfaces by combining evolution, geometry, and molecular docking**. *Proteins* 2019, **87**(11):952-965.
211. Mintseris J, Pierce B, Wiehe K, Anderson R, Chen R, Weng Z: **Integrating statistical pair potentials into protein complex prediction**. *Proteins: Structure, Function, and Bioinformatics* 2007, **69**(3):511-520.
212. Roel-Touris J, Bonvin AMJJ, Jiménez-García B: **LightDock goes information-driven**. *Bioinformatics (Oxford, England)* 2019 (*in press*).
213. Andreani J, Faure G, Guerois R: **InterEvScore: a novel coarse-grained interface scoring function using a multi-body statistical potential coupled to evolution**. *Bioinformatics (Oxford, England)* 2013, **29**(14):1742-1749.
214. Song Y, DiMaio F, Wang Ray Y-R, Kim D, Miles C, Brunette TJ, Thompson J, Baker D: **High-Resolution Comparative Modeling with RosettaCM**. *Structure (London, England : 1993)* 2013, **21**(10):1735-1742.
215. Koukos PI, Bonvin A: **Integrative Modelling of Biomolecular Complexes**. *Journal of molecular biology* 2020, **432**(9):2861-2881.
216. Karaca E, Bonvin AMJJ: **On the usefulness of ion-mobility mass spectrometry and SAXS data in scoring docking decoys**. *Acta Crystallographica Section D* 2013, **69**(5):683-694.
217. Schneidman-Duhovny D, Hammel M: **Modeling Structure and Dynamics of Protein Complexes with SAXS Profiles**. *Methods in molecular biology (Clifton, NJ)* 2018, **1764**:449-473.
218. Schneidman-Duhovny D, Hammel M, Sali A: **Macromolecular docking restrained by a small angle X-ray scattering profile**. *Journal of structural biology* 2011, **173**(3):461-471.
219. Schindler Christina EM, de Vries Sjoerd J, Sasse A, Zacharias M: **SAXS Data Alone can Generate High-Quality Models of Protein-Protein Complexes**. *Structure (London, England : 1993)* 2016, **24**(8):1387-1397.
220. Ignatov M, Kazennov A, Kozakov D: **ClusPro FMFT-SAXS: Ultra-fast Filtering Using Small-Angle X-ray Scattering Data in Protein Docking**. *Journal of molecular biology* 2018, **430**(15):2249-2255.
221. Vreven T, Schweppe DK, Chavez JD, Weisbrod CR, Shibata S, Zheng C, Bruce JE, Weng Z: **Integrating Cross-Linking Experiments with Ab Initio Protein-Protein Docking**. *Journal of molecular biology* 2018, **430**(12):1814-1828.
222. Navío D, Rosell M, Aguirre J, de la Cruz X, Fernández-Recio J: **Structural and Computational Characterization of Disease-Related Mutations Involved in Protein-Protein Interfaces**. *International journal of molecular sciences* 2019, **20**(7):1583.

223. Gao M, Zhou H, Skolnick J: **Insights into Disease-Associated Mutations in the Human Proteome through Protein Structural Analysis**. *Structure (London, England : 1993)* 2015, **23**(7):1362-1369.
224. Jubb HC, Pandurangan AP, Turner MA, Ochoa-Montaño B, Blundell TL, Ascher DB: **Mutations at protein-protein interfaces: Small changes over big surfaces have large impacts on human health**. *Progress in Biophysics and Molecular Biology* 2017, **128**:3-13.
225. Gregersen N, Bross P, Vang S, Christensen JH: **Protein Misfolding and Human Disease**. 2006, **7**(1):103-124.
226. Aguzzi A, O'Connor T: **Protein aggregation diseases: pathogenicity and therapeutic perspectives**. *Nature Reviews Drug Discovery* 2010, **9**:237.
227. Vaser R, Adusumalli S, Leng SN, Sikic M, Ng PC: **SIFT missense predictions for genomes**. *Nature Protocols* 2015, **11**:1.
228. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J: **A general framework for estimating the relative pathogenicity of human genetic variants**. *Nature genetics* 2014, **46**:310.
229. Adzhubei I, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR: **A method and server for predicting damaging missense mutations**. *Nature methods* 2010, **7**(4):248-249.
230. Niroula A, Urolagin S, Vihinen M: **PON-P2: prediction method for fast and reliable identification of harmful variants**. *PloS one* 2015, **10**(2):e0117380.
231. López-Ferrando V, Gazzo A, de la Cruz X, Orozco M, Gelpí JL: **PMut: a web-based tool for the annotation of pathological variants on proteins, 2017 update**. *Nucleic acids research* 2017, **45**(W1):W222-W228.
232. Riera C, Lois S, de la Cruz X: **Prediction of pathological mutations in proteins: The challenge of integrating sequence conservation and structure stability principles**; 2014.
233. Sunyaev SR: **Inferring causality and functional significance of human coding DNA variants**. *Human molecular genetics* 2012, **21**(R1):R10-R17.
234. Yates CM, Sternberg MJE: **The Effects of Non-Synonymous Single Nucleotide Polymorphisms (nsSNPs) on Protein–Protein Interactions**. *Journal of molecular biology* 2013, **425**(21):3949-3963.
235. Kucukkal TG, Petukh M, Li L, Alexov E: **Structural and physico-chemical effects of disease and non-disease nsSNPs on proteins**. *Current opinion in structural biology* 2015, **32**:18-24.
236. Berman HM, Kleywegt GJ, Nakamura H, Markley JL: **The Protein Data Bank archive as an open data resource**. *Journal of computer-aided molecular design* 2014, **28**(10):1009-1014.
237. The UniProt Consortium: **UniProt: the universal protein knowledgebase**. *Nucleic acids research* 2016, **45**(D1):D158-D169.
238. Riera C, Lois S, Domínguez C, Fernandez-Cadenas I, Montaner J, Rodríguez-Sureda V, de la Cruz X: **Molecular damage in Fabry disease: Characterization and prediction of alpha-galactosidase A pathological mutations**. 2015, **83**(1):91-104.
239. Riera C, Padilla N, de la Cruz X: **The Complementarity Between Protein-Specific and General Pathogenicity Predictors for Amino Acid Substitutions**. *Human mutation* 2016, **37**(10):1013-1024.
240. Mistry J, Finn RD, Eddy SR, Bateman A, Punta M: **Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions**. *Nucleic acids research* 2013, **41**(12):e121-e121.
241. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J *et al*: **Pfam: the protein families database**. *Nucleic acids research* 2013, **42**(D1):D222-D230.

242. Jackson RM, Gabb HA, Sternberg MJ: **Rapid refinement of protein interfaces incorporating solvation: application to the docking problem**. *Journal of molecular biology* 1998, **276**(1):265-285.
243. Barradas-Bautista D, Fernández-Recio J: **Docking-based modeling of protein-protein interfaces for extensive structural and functional characterization of missense mutations**. *PloS one* 2017, **12**(8):e0183643.
244. Development Core Team R: **R: A Language and Environment for Statistical Computing**, vol. 1; 2011.
245. Agius R, Torchala M, Moal IH, Fernández-Recio J, Bates PA: **Characterizing Changes in the Rate of Protein-Protein Dissociation upon Interface Mutation Using Hotspot Energy and Organization**. *PLOS Computational Biology* 2013, **9**(9):e1003216.
246. Guharoy M, Chakrabarti P: **Conservation and relative importance of residues across protein-protein interfaces**. *Proceedings of the National Academy of Sciences of the United States of America* 2005, **102**(43):15447-15452.
247. de Beer TAP, Laskowski RA, Parks SL, Sipos B, Goldman N, Thornton JM: **Amino Acid Changes in Disease-Associated Variants Differ Radically from Variants Observed in the 1000 Genomes Project Dataset**. *PLOS Computational Biology* 2013, **9**(12):e1003382.
248. Thusberg J, Olatubosun A, Vihinen M: **Performance of mutation pathogenicity prediction methods on missense variants**. 2011, **32**(4):358-368.
249. Kammenga JE: **The background puzzle: how identical mutations in the same gene lead to different disease symptoms**. *The FEBS Journal* 2017, **284**(20):3362-3373.
250. Keskin O, Ma B, Nussinov R: **Hot Regions in Protein–Protein Interactions: The Organization and Contribution of Structurally Conserved Hot Spot Residues**. *Journal of molecular biology* 2005, **345**(5):1281-1294.
251. Teppa E, Zea DJ, Marino-Buslje C: **Protein-protein interactions leave evolutionary footprints: High molecular coevolution at the core of interfaces**. *Protein science : a publication of the Protein Society* 2017, **26**(12):2438-2444.
252. Moreira IS, Fernandes PA, Ramos MJ: **Hot spots—A review of the protein–protein interface determinant amino-acid residues**. 2007, **68**(4):803-812.
253. Kenneth Morrow J, Zhang S: **Computational prediction of protein hot spot residues**. *Current pharmaceutical design* 2012, **18**(9):1255-1265.
254. Krivov GG, Shapovalov MV, Dunbrack RL, Jr.: **Improved prediction of protein side-chain conformations with SCWRL4**. *Proteins* 2009, **77**(4):778-795.
255. Fernandez-Recio J, Totrov M, Skorodumov C, Abagyan R: **Optimal docking area: A new method for predicting protein–protein interaction sites**. *Proteins: Structure, Function, and Bioinformatics* 2005, **58**(1):134-143.
256. Amengual-Rigo P, Fernández-Recio J, Guallar V: **UEP: an open-source and fast classifier for predicting the impact of mutations in protein-protein complexes**. *Bioinformatics (Oxford, England)* 2020.
257. Hwang H, Vreven T, Janin J, Weng Z: **Protein–protein docking benchmark version 4.0**. *Proteins: Structure, Function, and Bioinformatics* 2010, **78**(15):3111-3114.
258. Smith EH, Thomas C, McHugh D, Gavrilov D, Raymond K, Rinaldo P, Tortorelli S, Matern D, Highsmith WE, Oglesbee D: **Allelic diversity in MCAD deficiency: The biochemical classification of 54 variants identified during 5years of ACADM sequencing**. *Molecular Genetics and Metabolism* 2010, **100**(3):241-250.
259. Grantham R: **Amino Acid Difference Formula to Help Explain Protein Evolution**. *Science (New York, NY)* 1974, **185**(4154):862.
260. Martinelli S, Krumbach OHF, Pantaleoni F, Coppola S, Amin E, Pannone L, Nouri K, Farina L, Dvorsky R, Lepri F *et al*: **Functional Dysregulation of CDC42 Causes Diverse Developmental Phenotypes**. *Am J Hum Genet* 2018, **102**(2):309-320.

261. Salazar D, Zhang L, deGala GD, Frerman FE: **Expression and characterization of two pathogenic mutations in human electron transfer flavoprotein**. *J Biol Chem* 1997, **272**(42):26425-26433.
262. Olsen RKJ, Andresen BS, Christensen E, Bross P, Skovby F, Gregersen N: **Clear relationship between ETF/ETFDH genotype and phenotype in patients with multiple acyl-CoA dehydrogenation deficiency**. *Human mutation* 2003, **22**(1):12-23.
263. Magalhães Rebelo AP, Dal Bello F, Knedlik T, Kaar N, Volpin F, Shin SH, Giacomello M: **Chemical Modulation of Mitochondria-Endoplasmic Reticulum Contact Sites**. *Cells* 2020, **9**(7).
264. Bongini P, Trezza A, Bianchini M, Spiga O, Niccolai N: **A possible strategy to fight COVID-19: Interfering with spike glycoprotein trimerization**. *Biochem Biophys Res Commun* 2020, **528**(1):35-38.
265. Grosdidier S, Fernandez-Recio J: **Docking and scoring: applications to drug discovery in the interactomics era**. *Expert opinion on drug discovery* 2009, **4**(6):673-686.
266. Ran X, Gestwicki JE: **Inhibitors of protein-protein interactions (PPIs): an analysis of scaffold choices and buried surface area**. *Current opinion in chemical biology* 2018, **44**:75-86.
267. Mabonga L, Kappo AP: **Protein-protein interaction modulators: advances, successes and remaining challenges**. *Biophys Rev* 2019, **11**(4):559-581.
268. Pérot S, Sperandio O, Miteva MA, Camproux AC, Villoutreix BO: **Druggable pockets and binding site centric chemical space: a paradigm shift in drug discovery**. *Drug Discov Today* 2010, **15**(15-16):656-667.
269. Lee GM, Craik CS: **Trapping Moving Targets with Small Molecules**. *Science (New York, NY)* 2009, **324**(5924):213-215.
270. Grosdidier S, Fernandez-Recio J: **Protein-protein docking and hot-spot prediction for drug discovery**. *Current pharmaceutical design* 2012, **18**(30):4607-4618.
271. Westbrook J, Feng Z, Jain S, Bhat TN, Thanki N, Ravichandran V, Gilliland GL, Bluhm W, Weissig H, Greer DS *et al*: **The Protein Data Bank: unifying the archive**. *Nucleic acids research* 2002, **30**(1):245-248.
272. Chittenden T, Flemington C, Houghton AB, Ebb RG, Gallo GJ, Elangovan B, Chinnadurai G, Lutz RJ: **A conserved domain in Bak, distinct from BH1 and BH2, mediates cell death and protein binding functions**. *Embo j* 1995, **14**(22):5589-5596.
273. Oltersdorf T, Elmore SW, Shoemaker AR, Armstrong RC, Augeri DJ, Belli BA, Bruncko M, Deckwerth TL, Dinges J, Hajduk PJ *et al*: **An inhibitor of Bcl-2 family proteins induces regression of solid tumours**. *Nature* 2005, **435**(7042):677-681.
274. Lee EF, Czabotar PE, Smith BJ, Deshayes K, Zobel K, Colman PM, Fairlie WD: **Crystal structure of ABT-737 complexed with Bcl-xL: implications for selectivity of antagonists of the Bcl-2 family**. *Cell Death & Differentiation* 2007, **14**(9):1711-1713.
275. Shiozaki EN, Chai J, Rigotti DJ, Riedl SJ, Li P, Srinivasula SM, Alnemri ES, Fairman R, Shi Y: **Mechanism of XIAP-Mediated Inhibition of Caspase-9**. *Mol Cell* 2003, **11**(2):519-527.
276. Wu G, Chai J, Suber TL, Wu J-W, Du C, Wang X, Shi Y: **Structural basis of IAP recognition by Smac/DIABLO**. *Nature* 2000, **408**(6815):1008-1012.
277. Brady M, Vlatković N, Boyd MT: **Regulation of p53 and MDM2 Activity by MTBP**. *Molecular and Cellular Biology* 2005, **25**(2):545-553.
278. Mauri DN, Ebner R, Montgomery RI, Kochel KD, Cheung TC, Yu GL, Ruben S, Murphy M, Eisenberg RJ, Cohen GH *et al*: **LIGHT, a new member of the TNF superfamily, and lymphotoxin alpha are ligands for herpesvirus entry mediator**. *Immunity* 1998, **8**(1):21-30.
279. Mosyak L, Zhang Y, Glasfeld E, Haney S, Stahl M, Seehra J, Somers WS: **The bacterial cell-division protein ZipA and its interaction with an FtsZ fragment revealed by X-ray crystallography**. *The EMBO Journal* 2000, **19**(13):3179-3191.

280. Chylack LT, Jr., Fu L, Mancini R, Martin-Rehrmann MD, Saunders AJ, Konopka G, Tian D, Hedley-Whyte ET, Folkerth RD, Goldstein LE: **Lens epithelium-derived growth factor (LEDGF/p75) expression in fetal and adult human brain**. *Exp Eye Res* 2004, **79**(6):941-948.
281. Abbate EA, Berger JM, Botchan MR: **The X-ray structure of the papillomavirus helicase in complex with its molecular matchmaker E2**. *Genes Dev* 2004, **18**(16):1981-1996.
282. Goudy K, Aydin D, Barzaghi F, Gambineri E, Vignoli M, Ciullini Mannurita S, Doglioni C, Ponzoni M, Cicalese MP, Assanelli A *et al*: **Human IL2RA null mutation mediates immunodeficiency with lymphoproliferation and autoimmunity**. *Clin Immunol* 2013, **146**(3):248-261.
283. Friesner RA, Murphy RB, Repasky MP, Frye LL, Greenwood JR, Halgren TA, Sanschagrin PC, Mainz DT: **Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein–Ligand Complexes**. *Journal of Medicinal Chemistry* 2006, **49**(21):6177-6196.
284. Mysinger MM, Carchia M, Irwin JJ, Shoichet BK: **Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking**. *Journal of medicinal chemistry* 2012, **55**(14):6582-6594.
285. Moroy G, Martin E, Dejaegere A, Stote RH: **Molecular basis for Bcl-2 homology 3 domain recognition in the Bcl-2 protein family: identification of conserved hot spot interactions**. *J Biol Chem* 2009, **284**(26):17499-17511.
286. Lee EF, Fairlie WD: **The Structural Biology of Bcl-x(L)**. *International journal of molecular sciences* 2019, **20**(9):2234.
287. Campbell ST, Carlson KJ, Buchholz CJ, Helmers MR, Ghosh I: **Mapping the BH3 Binding Interface of Bcl-xL, Bcl-2, and Mcl-1 Using Split-Luciferase Reassembly**. *Biochemistry* 2015, **54**(16):2632-2643.
288. Wang Y, Coulombe R, Cameron DR, Thauvette L, Massariol M-J, Amon LM, Fink D, Titolo S, Welchner E, Yoakim C *et al*: **Crystal structure of the E2 transactivation domain of human papillomavirus type 11 bound to a protein interaction inhibitor**. *J Biol Chem* 2004, **279**(8):6976-6985.
289. O'Neill JW, Manion MK, Maguire B, Hockenbery DM: **BCL-XL Dimerization by Three-dimensional Domain Swapping**. *Journal of molecular biology* 2006, **356**(2):367-381.