

**Department of Physics, Chemistry and Biology**

**Master's Thesis**

**Protein-Protein Docking Using Starting Points Based On  
Structural Homology**

**Martin Hyvönen**

**150908**

**LITH-IFM-G-EX--15/3119—SE**



**Linköping University**  
**INSTITUTE OF TECHNOLOGY**

**Linköping University**

**Department of Physics, Chemistry and Biology**

**581 83 Linköping**

**Department of Physics, Chemistry and Biology**

# **Protein-Protein Docking Using Starting Points Based On Structural Homology**

**Martin Hyvönen**

**Thesis work was done at the Bioinformatics group,**

**Linköping university, Linköping**

**150908**

**Supervisor: Claudio Mirabello**

**Examiner: Björn Wallner**



**Linköping University**  
**INSTITUTE OF TECHNOLOGY**

**Linköping University**

**Department of Physics, Chemistry and Biology**

**581 83 Linköping**



**Avdelning, institution**

**Division, Department**

**Chemistry**

**Department of Physics, Chemistry and Biology**

**Linköping University**

**Datum**

**Date**

150908

**Språk**

Language

☐ Svenska/Swedish

☒ Engelska/English

☐ \_\_\_\_\_

**Rapporttyp**

Report category

☐ Licentiatavhandling

☒ Examensarbete

☐ C-uppsats

☐ D-uppsats

☐ Övrig rapport

☐ \_\_\_\_\_

**ISBN**

**ISRN: LITH-IFM-G-EX--15/3119--SE**

**Serietitel och serienummer ISSN**

**Title of series, numbering** \_\_\_\_\_

**URL för elektronisk version**

**Titel**

**Title**

**Protein-Protein Docking Using Starting Points Based On Structural Homology**

**Författare**

**Author**

**Martin Hyvönen**

**Sammanfattning**

**Abstract**

Protein-protein interactions build large networks which are essential in understanding complex diseases. Due to limitations of experimental methodology there are problems with large amounts of false negative and positive interactions; and a large gap in the amount of known interactions and structurally determined interactions. By using computational methods these problems can be alleviated.

In this thesis the quality of a newly developed pipeline (InterPred) were investigated for its ability to generate coarse interaction models and score them. This ability was investigated by performing docking experiments in Rosetta on models generated in InterPred.

The results suggest that InterPred is highly successful in generating good starting points for docking proteins in silico and to distinguish the quality of models.

**Nyckelord**

**Keyword**

**In silico, Rosetta Dock, Bound docking, PPI, Bioinformatics, InterPred**

## Abstract

Protein-protein interactions build large networks which are essential in understanding complex diseases. Due to limitations of experimental methodology there are problems with large amounts of false negative and positive interactions; and a large gap in the amount of known interactions and structurally determined interactions. By using computational methods these problems can be alleviated.

In this thesis the quality of a newly developed pipeline (InterPred) were investigated for its ability to generate coarse interaction models and score them. This ability was investigated by performing docking experiments in Rosetta on models generated in InterPred.

The results suggest that InterPred is highly successful in generating good starting points for docking proteins in silico and to distinguish the quality of models.

## Abbreviations

PPI	protein-protein interaction
FFT	fast fourir transfer
$F_{nat}$	$F_{native}$ fraction native interaction
MC	Monte Carlo
RMSD	Root Mean Square Deviation
RMS	Root Mean Square
Lrms	ligand RMS
Irms	interface RMS
HT	high throughput
Y2H	yeast 2-hybrid
PCA	protein complementation assay
NMR	nuclear magnetic resonance
PDB	Protein databank
CAPRI	Critical Assessment of Predicted Interactions



Abstract .....	iv
Abbreviations .....	iv
1 Introduction.....	1
1.1 Purpose of the study .....	1
1.2 Ethics statement.....	1
2 Theory.....	2
2.1 PPI networks and protein structures in healthcare.....	2
2.2 The in silico Structural Protein Problem.....	2
2.3 In silico protein docking.....	2
2.3.1 Fast Fourier Transfer (FFT) .....	2
2.3.2 Geometric hashing (GH) .....	3
2.3.3 Rosetta.....	3
2.3.4 InterPred.....	4
2.4 Model quality .....	6
2.4.1 Capri.....	6
3 Process.....	8
4 Methods .....	9
4.1 Target structures and preparation .....	9
4.1.1 Relax protocol.....	9
4.1.2 Choosing models and targets .....	9
4.2 Scoring.....	10
4.3 Docking protocols.....	10
4.4 Models found by sequence similarity .....	11
4.5 Determining successfully docked models .....	11
5 Results .....	12
5.1 Preparation.....	12
5.2 Selected models .....	14
5.3 Models found by sequence similarity .....	15
5.4 Docking.....	15
5.4.1 Successfully docked targets and models.....	15
5.4.2 Unsuccessfully docked targets .....	21
5.5 Process analysis .....	25
6 Discussion.....	27

6.1	Process discussion .....	27
6.2	Relaxation .....	27
6.3	Selected models .....	28
6.4	Docking .....	28
6.4.1	Successful docking .....	28
6.4.2	Unsuccessful docking .....	29
6.5	Future prospects and conclusions .....	29
7	Acknowledgments .....	30
	References .....	30

# 1 Introduction

Proteins are the most important macromolecules in organisms. They are responsible for all reactions and all regulation, fulfilling tasks such as catalyzing chemical reactions, giving cells and organelles their shape, all movement and transportation and finally all communication and defense of the organism (Whitford 2005). However many of these functions are dependent on proteins interacting with other proteins forming complexes of at least two proteins. Which makes knowledge about protein-protein interactions (PPI) essential to understanding complex phenotypic behavior (Wang, Marcotte 2010). Even though major efforts have been spent in order to experimentally chart and characterize these PPI networks, interactome coverage remains low (Shoemaker, Panchenko 2007, Melquiond, Karaca et al. 2012). There are many methods for investigating and proving PPIs, however given the amount of predicted PPIs, interactome studies are best investigated with high-throughput (HT) methods, such as yeast 2-hybrid (Y2H), affinity purification and protein complementation assays (PCA). Results from these models do however suffer from high rates of false positives and negatives, with additional problems of few overlaps between the different methods (Wodak, Vlasblom et al. 2013). When it comes to characterizing the PPIs on a molecular level the problem instead lies with the low throughput of the methods, such as NMR, X-ray crystallography or cryo-EM. As these methods are time consuming and comparatively expensive there is a large gap between known proteins which is further expanded between structurally known protein structures and PPI structures.

For the reasons above, computational methods are needed as a complement to fill the gaps of knowledge.

## 1.1 Purpose of the study

In this study a currently unpublished computational pipeline InterPred (Mirabello et al) is investigated, for the sake of determining InterPreds ability to generate starting points for docking proteins in silico. These starting points can then be considered coarse models of the interaction and by evaluating results from docking these coarse models, their quality and InterPreds ability to distinguish model quality will be investigated.

## 1.2 Ethics statement

As this study is strictly performed in-silico on data from the Protein Data Bank (PDB) there is no ethics to be considered for this work.



## 2 Theory

### 2.1 PPI networks and protein structures in healthcare

There are many reasons to investigate PPI networks. To name just a few they can be used to explore the difference between healthy and diseased states. They can be used to find potential disease associated targets for drug development. Disease associated protein information can then be further investigated in order to develop drugs for less connected proteins in disease networks which has the potential to reduce the severity and scope of side effects. PPI networks especially hold the potential to help in the case of complex diseases in which there are no clear genotype-phenotype connections. Cancer is one disease in which the use of PPI networks has been useful in diagnosing and differentiating between different cancers with different outcomes and treatments (Safari-Alighiarloo, Taghizadeh et al. 2014).

As for structural information it can for example reveal disease association (Butler, Gerek et al. 2015). It is furthermore useful as a template in the development of drugs or as basis for deciding which amino acids are of interest for random mutagenesis studies in order to generate better interactions.

### 2.2 The in silico Structural Protein Problem

The computational docking problem is very expensive in terms of calculations and is extremely complex when contemplating it in its entirety, so it is instead divided into many smaller problems. First of all there is the differentiation between finding the structure of a docked complex and simply deciding whether the two proteins are interacting. Secondly the inputs used for docking decide the difficulty as well and they are from hardest to easiest:

- Primary protein structure: in this instance the proteins will first have to be folded, either individually or in relation to each other;
- Unbound proteins are used as inputs: here the structures where the proteins have been determined one at a time are used to dock and then compared against a native complex where their structure was determined while the proteins are interacting with each other;
- Bound Proteins are used as inputs: in this case the structure of the interaction is already known. It is not really part of the docking problem but can be used to determine the efficacy of docking methods.

### 2.3 In silico protein docking

There are many different ways to model proteins and protein interactions with different advantages and disadvantages. The following are some of the more common docking methods of which two were used in this study.

#### 2.3.1 Fast Fourier Transfer (FFT)

FFT docking is based on representing molecules in sets of squares. When docking occurs each position  $x$  (three dimensional position) is given the values  $a(x)$  and  $b(x)$  respectively for the molecules A and B where  $a(x)$  is 0 outside molecule A, 1 on the thin shell of molecule A and a positive value inside of molecule A. whereas  $b(x)$  is the same except it is given a negative value inside. A penalty is

thereby given since any overlap will be the product of multiplication between a positive and a negative value.

The starting point is then evaluated as the following sum:

$$S(0) = \sum_x a(x) * b(x)$$

After translating molecule B to position  $u$  the sum instead becomes:

$$S(u) = \sum_x a(x) * b(x + u)$$

However calculating  $S(u)$  would be very computationally expensive which is why  $a$  and  $b$  are fourier transformed,  $S(u)$  is calculated and reverse Fourier transformed. While the FFT algorithm does all this very efficiently it has to be done for all orientations of molecule B. The weakness of this method is that most translations are incorrect but all must be calculated. After the FFT method has performed the rigid body search the best structures must then be scored and refined, it is common that refinement is performed in Rosetta (see Monte Carlo-based docking) (Janin 2010).

### 2.3.2 Geometric hashing (GH)

Geometric hashing starts with making a dense Connolly-style surface representation with associated normals. Using further algorithms the dense surface is made into a less dense surface of knobs and holes. After further reducing the amount of space in need of searching by integrating biological and surface information the geometric hashing algorithm looks for complementary surfaces, a hole for each knob, with normals pointing in opposite directions. After the hashing is done the models are sorted in order of complementarity. However like FFT based algorithms GH as well needs refinement for biological ranking and scoring (Janin 2010, Schneidman-Duhovny, Inbar et al. 2003) .

### 2.3.3 Rosetta

Rosetta is a Monte Carlo based docking software and in this study Rosetta 2014.8 was used for all docking and scoring of generated models.

In figure 1 a flowchart of the standard docking protocol is provided. The algorithm begins the creation of models of the interaction (decoys) by randomly translating one protein in relation to the other in order to create glancing contacts. Then a rigid-body Monte Carlo search attempts to translate and rotate one protein along the surface of the other protein 500 times. The step sizes are adjusted in order to keep a 50% acceptance rate of the steps, starting from an initial mean of 0.7 Å and 5° spin. At this stage all side chains are replaced by large pseudo-atoms called centroids, to reduce the amount of computations needed to avoid overlaps and calculate residue-residue interactions (see CAPRI).

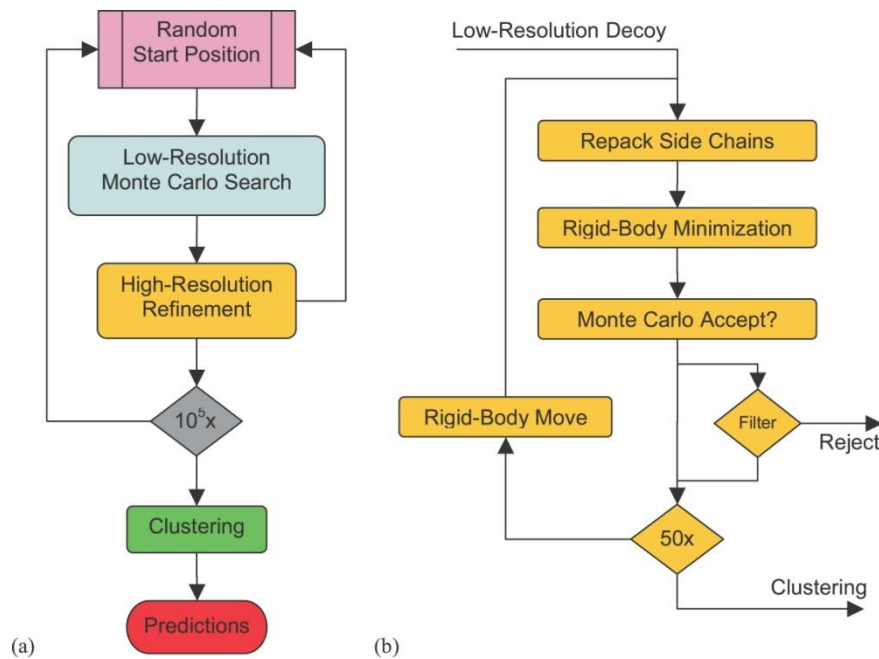


Figure 1: (a) Is a visualization of the docking protocol; (b) shows the high resolution refinement in detail. (Gray, Moughon et al. 2003)

After the low resolution search the centroids are replaced with side chains using a backbone dependent rotamer library, which contains all allowed positional conformations of side chains based on backbone angles. The rotamers are then optimized through a simulated annealing Monte Carlo search. Once the side chains are added the rigid body displacement is optimized. In order to simultaneously optimize rigid-body translation and side chain conformation, the side chain packing and minimization steps are repeated 50 times. Before each cycle one protein is randomly translated 0.1 Å in each direction with a random rotation of 0.05°. After each packing, move and minimization a score is calculated. After the final cycle the lowest scoring conformation is minimized one last time. (Gray, Moughon et al. 2003) Rosetta is also highly modifiable by adding flags which change or add functionality. For more information refer to the Rosetta manual ([www.rosettacommons.org/manuals](http://www.rosettacommons.org/manuals)).

### 2.3.4 InterPred

InterPred is a pipeline being developed for the sake of generating full interaction models from protein sequence information (Mirabello et al).

InterPred is modular and contains three modules the first generates protein structure from sequence information. The second module takes 2 protein structures as input, then finds structural homologs for each of the protein structures, which it uses as templates. The input structures are then structurally aligned to their template structures and in case a template for each of the inputs is known to interact a coarse interaction model is built based on the alignment. Lastly that coarse model is evaluated in order to sort the quality of the coarse models which can be seen as a starting point for further model refinement. The last module of InterPred takes a PDB file with 2 proteins as input and uses the docking protocol in Rosetta (Gray, Moughon et al. 2003) to refine the input structure and lastly find the best interaction model.

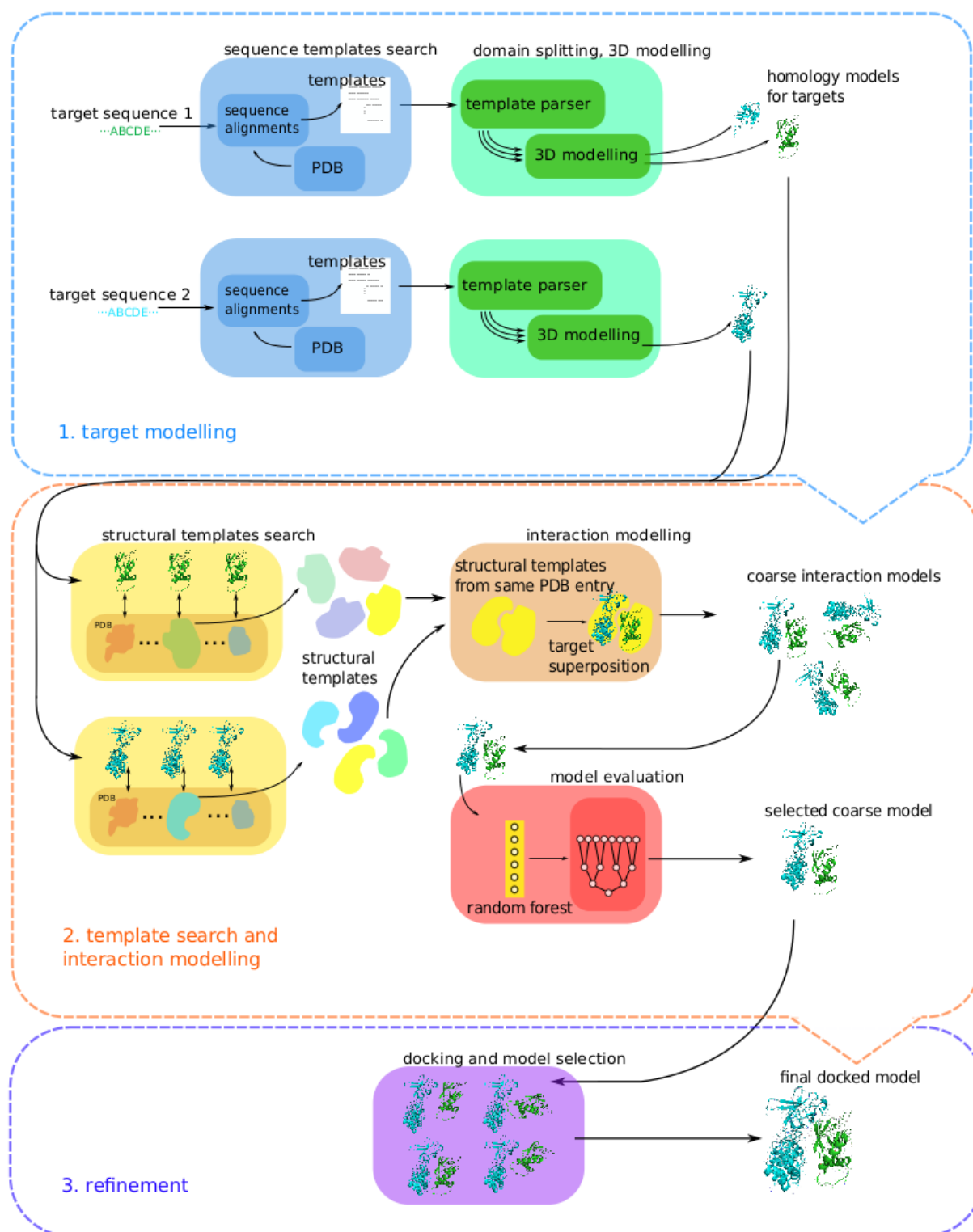


Figure 2: Overview of the methodology in InterPred which has the functionality to begin from a protein sequence model it into a protein structure. Then it finds structural homologs and generates coarse interaction models through superimposing modeled targets on templates. The coarse interaction models are then evaluated and selected models are

docked to refine the structure, thereby generating structural models for the interactions between the target sequences (Mirabella et al).

## 2.4 Model quality

### 2.4.1 Capri

CAPRI is short for Critical Assessment of PRedicted Interactions, a community-wide experiment designed to test the reliability and accuracy of in silico docking methods. CAPRI aims to be as close to a real prediction as possible and tries to achieve this by primarily using unbound targets. Meaning that the components of the target have to be available as free structures in the PDB, in some cases it has been acceptable that the structure can be model-built from close homologous structures. But the complex must not be available. This criteria have however been expanded to bound/unbound docking in order to allow for more targets to be available. In the more relaxed case of unbound/bound docking one of the components is known in advance and the other is taken from the complex (Janin 2010).

When evaluating the complexes in CAPRI there is primarily 3 different scores which are of interest.

#### **Fnat**

The first score is based on the amount of conserved residue-residue contacts in the predicted complex with regards to the target complex. This score is a fraction of predicted residue-residue contacts divided by native residue-residue contacts. Thereby giving it the name Fraction native interactions (Fnat). A residue is considered to be in contact with a residue on the other protein if they have one atom within 5 Å of each other (see figure 3).

#### **Lrms**

The second score is the Ligand Root Mean Square (Lrms). It is calculated by superimposing the predicted receptor (larger of the 2 proteins) on the target receptor and comparing the positions of the ligands (the smaller protein) backbone atoms (C $\alpha$ , N, C, O) (see figure 3).

#### **Irms**

The third score is the Interface Root Mean Square (Irms). It is counted by optimally superimposing the predicted interface on the native interface and comparing backbone atom placement. A residue was counted as part of the interface if it had an atom within 10Å of the other protein in the complex (see figure 3) (Méndez, Leplae et al. 2003).

## RMSD

This score is calculated based on heavy backbone atoms and simply the Root Mean Square Deviation between those atoms on the model and the target.

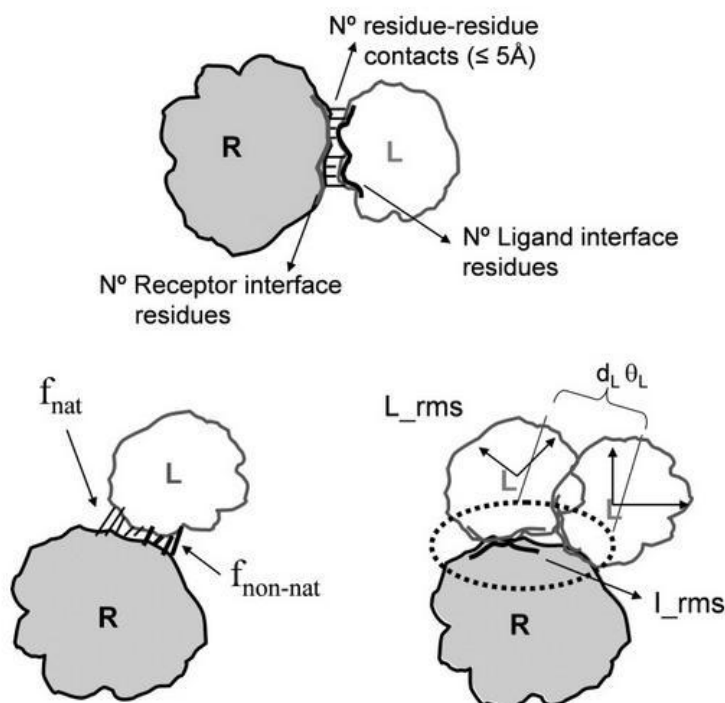


Figure 3: Illustration of the quality measures in CAPRI used to evaluate predicted models. Where R is Receptor and L is Ligand. For information about how the measures Fnat, Lrms and I rms are calculated see 2.4.1 (Méndez, Leplae et al. 2003).

The measures are then used to qualify the predicted models into the following classes of high medium and low quality as seen in table1. A model which does not qualify into the low quality category is considered a random solution.

Quality	Capri Quality Measure				
	Fnat	and	(Lrms	or	I rms)
High	≥0.5		≤1.0		≤1.0
Medium	≥0.3		≤5.0		≤2.0
Acceptable	≥0.1		≤10		≤4.0

Table 1: For a certain model quality the Fnat criteria and at least one of the Lrms and I rms must be fulfilled for a model to attain that quality.

### 3 Process

At the beginning of the project a timeplan was established (figure 4) with estimated time for each activity. The main tasks were introductory work, laboratory work and work on the report. All work was performed in close proximity with examiner and supervisor in order to systematically follow up the work and adapt to changing circumstances. Because it was not known how long the docking experiments was going to take it was simply planned to use a docking software for each docking methodology, Monte-Carlo based, FFT based and geometric hashing based. The methods were overlapped so that work on the next docking method was to start while scripts to generate decoys and scores for the previous method were being computed. It was also planned that writing the results and methodology as they were performed would be a efficient way to record the work and complete the report.

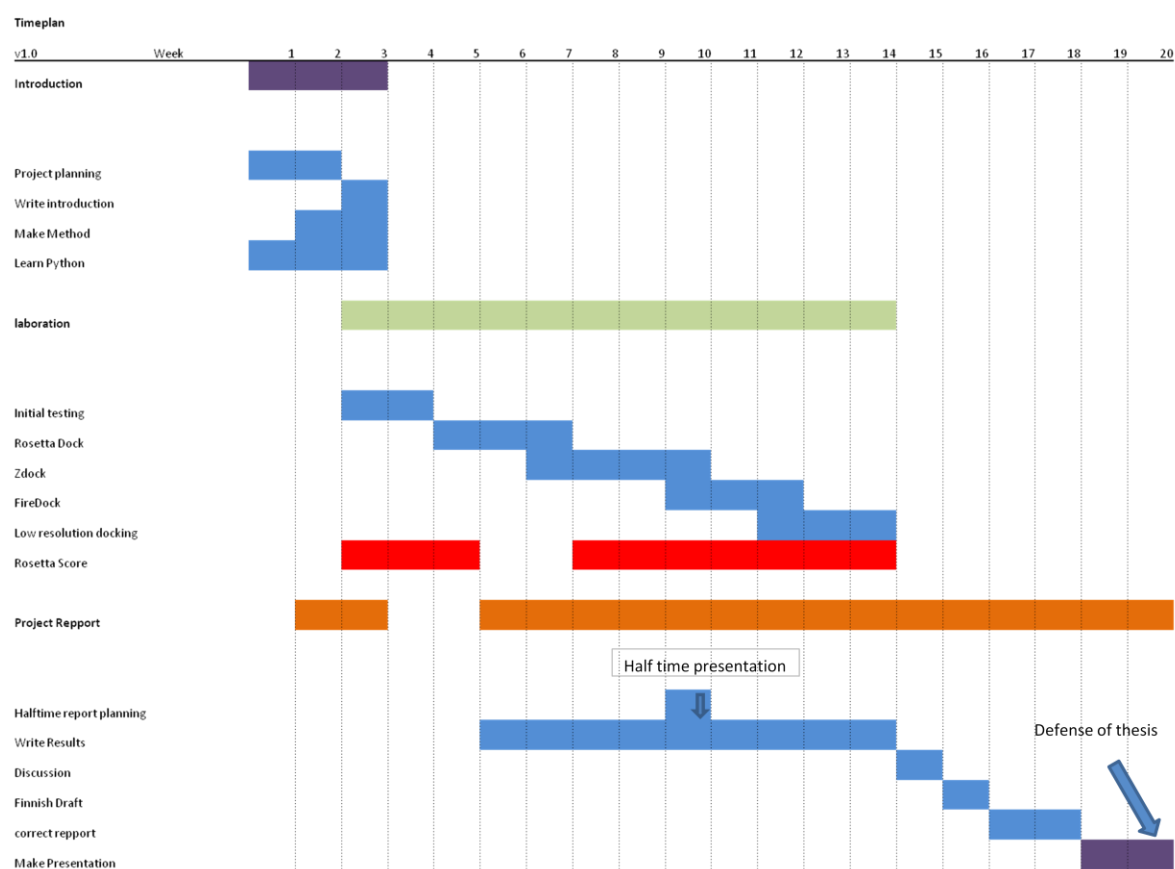


Figure 4: All activities are shown in blue except activities relating to scoring (red) and the work on the presentation (purple). Any week where lab work occurred is mint green and any week where work on the report is done is orange. The first three weeks were supposed to be introductory work and is colored purple.

## 4 Methods

### 4.1 Target structures and preparation

All experiments are performed on native structures (targets) from the large model set on Dockground (Anishchenko, Kundrotas et al. 2015). The set was created from a binary set of hetero complexes, crystallographically determined structures with a resolution below 3.5 Å, an interface larger than 250 Å<sup>2</sup> of buried solvent accessible area per chain and more than three secondary structure elements per chain. The set is then limited to 165 targets. (Anishchenko, Kundrotas et al. 2015, Roy, Kucukural et al. 2010) These targets then had their side chains and backbone optimized by using a Relaxation protocol in Rosetta (see 4.1.1). The relaxed structures were used as inputs to InterPreds second module which found structural homologs and aligned the relaxed target structures to their corresponding homolog template which yielded 44509 coarse interaction models (see 2.3.4), these were then scored in Rosetta (see 4.2). The amount of coarse models varied greatly for the targets, from 0 to 4024. So targets and models were selected in order to reduce bias (see 4.1.2 for the full procedure). After selection 2006 coarse models were chosen for 115 targets.

#### 4.1.1 Relax protocol

The native structures were relaxed in order to solve internal clashes and thereby prepare them for usage as native structures in scoring and docking experiments. The relaxation was performed using the following flags in Rosetta.

- relax.linuxgccrelease : flag for the relaxation protocol which refines the fullatom Rosetta model.
- database path to database : a flag to find the rosetta database
- relax:constrain\_relax\_to\_start\_coords : Flag which constrains to backbone heavy atoms
- relax:ramp\_constraints false : flag which cancels ramping constraints
- constraints:cst\_fa\_weight 1.0 : sets the weight for the relaxation flag
- in:file:s infile : where the infile is a PDB of the native target (native)
- in:file:native nativefile : in this case the nativefile is the same as the infile (native)
- outfile:scorefile scorefile : flag used to specify name and directory of the scorefile
- evaluation:DockMetrics : adds a couple of scores to the scorefile
- nstruct 1: Number of constructs (decoys) desired

#### 4.1.2 Choosing models and targets

Targets with 10 suggested models or more were selected for docking procedures which reduced native structures to 115. As there still was a large difference in coarse models per target, the amount was reduced in order to minimize bias caused by targets with many models. Therefore all models for a target were sorted into 5 bins of InterPred scores, with the intervals 0.50-0.60, 0.61-0.70, 0.71-0.80, 0.81-0.90, 0.91-1.00. Then models were chosen according to the following criteria.

- 5 models were randomly chosen for each interval for each target



- Target must not be aligned to itself

These constraints generated between 5 and 25 models for each target for a total of 2006 targets. The distribution of the models can be seen in table 2.

## 4.2 Scoring

All scoring of structures were performed in Rosetta using the following flags

- score\_jd2.linuxgccrelease : scoring protocol
- database databasedir: Flag responsible to add the Rosetta Database pathway
- in:file:s infile : Input target file(s) (native)
- in:file:native nativefile : provide a native structure for all RMSD calculations (Also the native file)
- outfile:scorefile target.score : flag used to assign name and directory of scorefiles generated
- evaluation:DockMetrics : adds a couple of scores to the scorefile needed for CAPRI evaluation

## 4.3 Docking protocols

The selected models were docked in two instances one default full protocol Rosetta docking and one with the additional flag `-pert 3 8`. Lastly all relaxed targets were docked in the same way as the normal docking but the target was entered both as an infile and a nativefile, to create decoys with the target as starting point and score them against the target. The last docking was performed to determine the difficulty of finding the target structure.

- docking\_protocol.linuxgccrelease : docking protocol
- evaluation:DockMetrics : flag for additional scores
- database databasedir : simply the database used for Rosetta
- in:file:s infile : the infile was a PDB file with 2 chains (coarse interaction model)
- in:file:native nativefile : the native file is a PDB file containing the structure against which all generated decoys are compared (target (native3) from which the coarse interaction model is created).
- out:file:silent silentfile : saves all output decoys to a single file from which interesting models can be extracted
- out:file:scorefile scorefile : simply a scorefile where all scores for the docking protocol and evaluation are stored
- nstruct 1000 : The number of decoys produced in the docking. In this study all docking experiments used 1000 decoys

-docking: dock\_pert 3 8 (used only for perturbed docking)

#### **4.4 Models found by sequence similarity**

After docking had been performed InterPreds second module was used again on the target structures, however this time sequence constraints instead of structural constraints were used to find templates (Mirabello et al). Because structure is more conserved than sequence, templates found using sequence homology are very similar in structure to the targets and all models generated based on sequence homologs are considered easier than models generated from structure homologs.

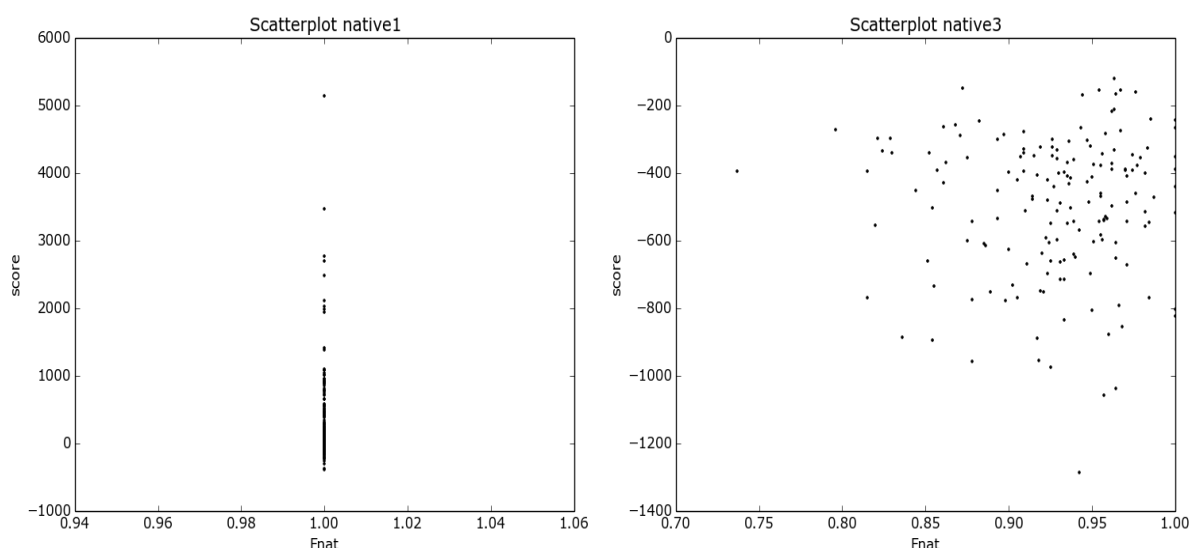
#### **4.5 Determining successfully docked models**

Each model was evaluated based on the 10 decoys with the best Rosetta energy score. The decoys were then evaluated against the CAPRI criteria. As soon as a model fulfilled the requirements for a CAPRI class the model was considered successfully docked at this quality and if no decoy fulfilled any of the CAPRI criteria (see table1) the model was considered unsuccessful.

## 5 Results

### 5.1 Preparation

When scoring the native targets in Rosetta it was found that they contained large clashes which led to high energy scores. Therefore the structures had to be relaxed to get rid of the worst clashes while keeping the structures as close as possible to the native structures. This yielded the relaxed native set (native3) seen in figure 5.



**Figure 5:** The figure shows the starting score of all native structures. Where  $F_{nat}$  is plotted against the Rosetta energy score. The left figure shows the data for the targets before relaxation was performed and the right figure shows the results after relaxation was performed.

When all models and relaxed models had been generated from native and native3 targets respectively the starting points for the models were compared in order to see that it is not vastly different and that it would be possible to select starting points of all combinations of  $F_{nat}$  and InterPred. A density comparison of the coarse-model starting points is seen in figure 6 where blue is low density and dark red signifies high density of starting points. Figure 7 more clearly shows the differences between the relaxed set and the targets.

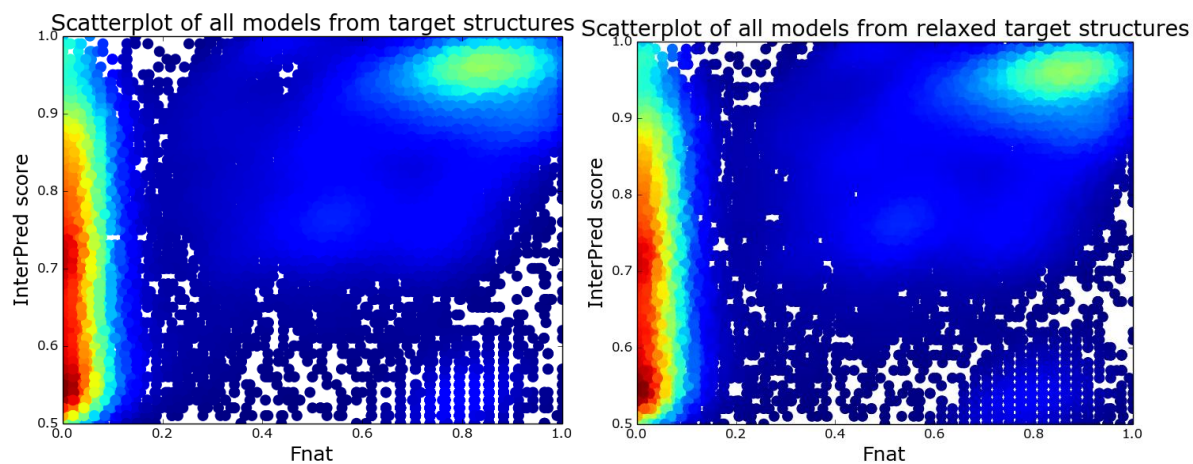


Figure 6: Above a comparison of the starting points before and after the relaxation is shown. Dark blue means that there is a low density of starting points and dark red that there is a high concentration of starting points.

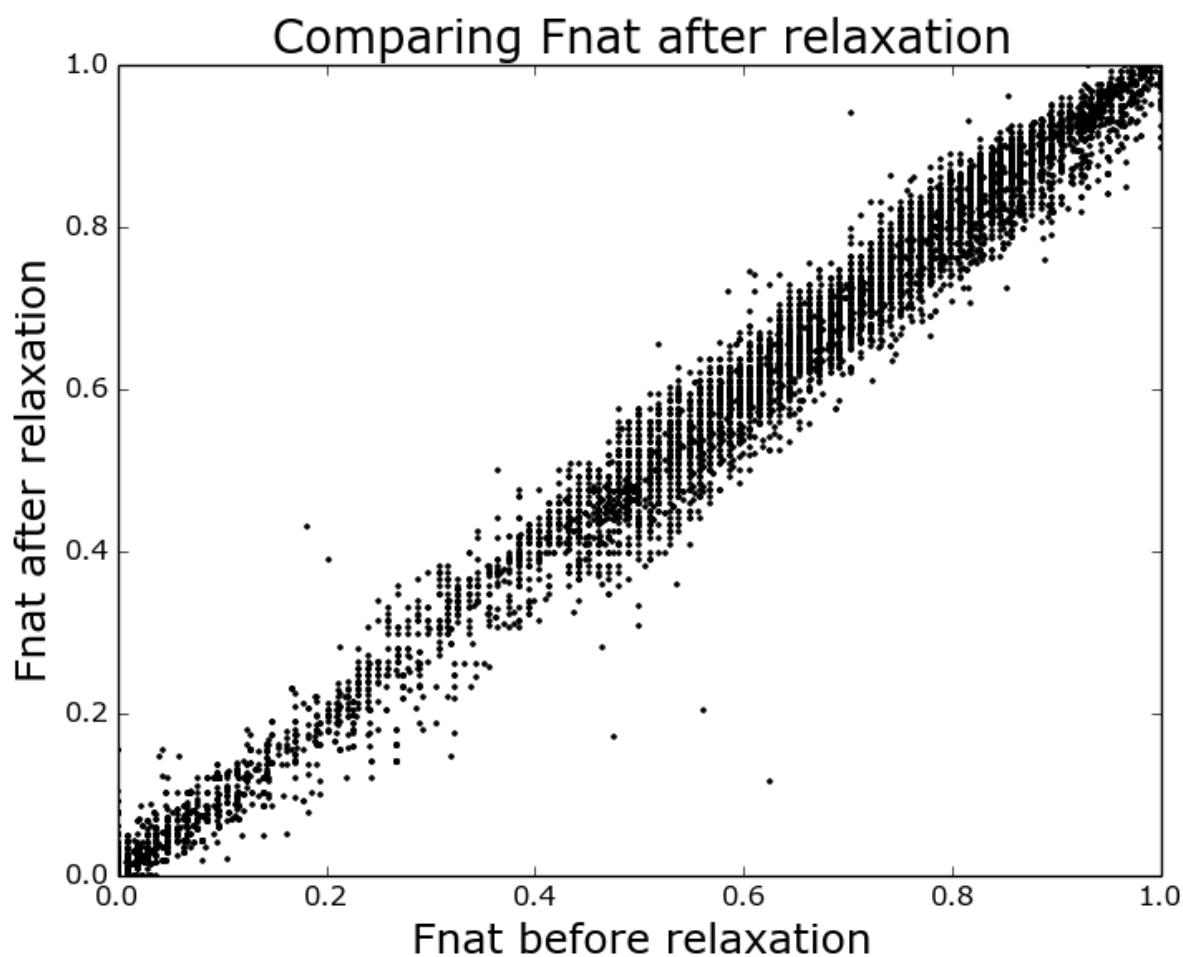


Figure 7: Shown in this figure is the Fnat before relaxation on the x-axis plotted against Fnat after relaxation on the y-axis.

## 5.2 Selected models

When choosing models only 115 targets fulfilled the CAPRI criteria (table 2) and yielded a total of 2006 models to dock. Table 2 shows the distribution of all the models chosen and also how many targets are in each InterPred bin.

Distribution of Chosen models					
InterPred Intervall	0,5-0,60	0,61-0,70	0,71-0,80	0,81-0,90	0,91-1,00
Number of models	329	373	413	412	479
Number of targets	78	87	97	101	107

Table 2: The chosen models and targets were distributed as shown above.

Comparison between InterPred and starting point Fnat is shown in figure 7.

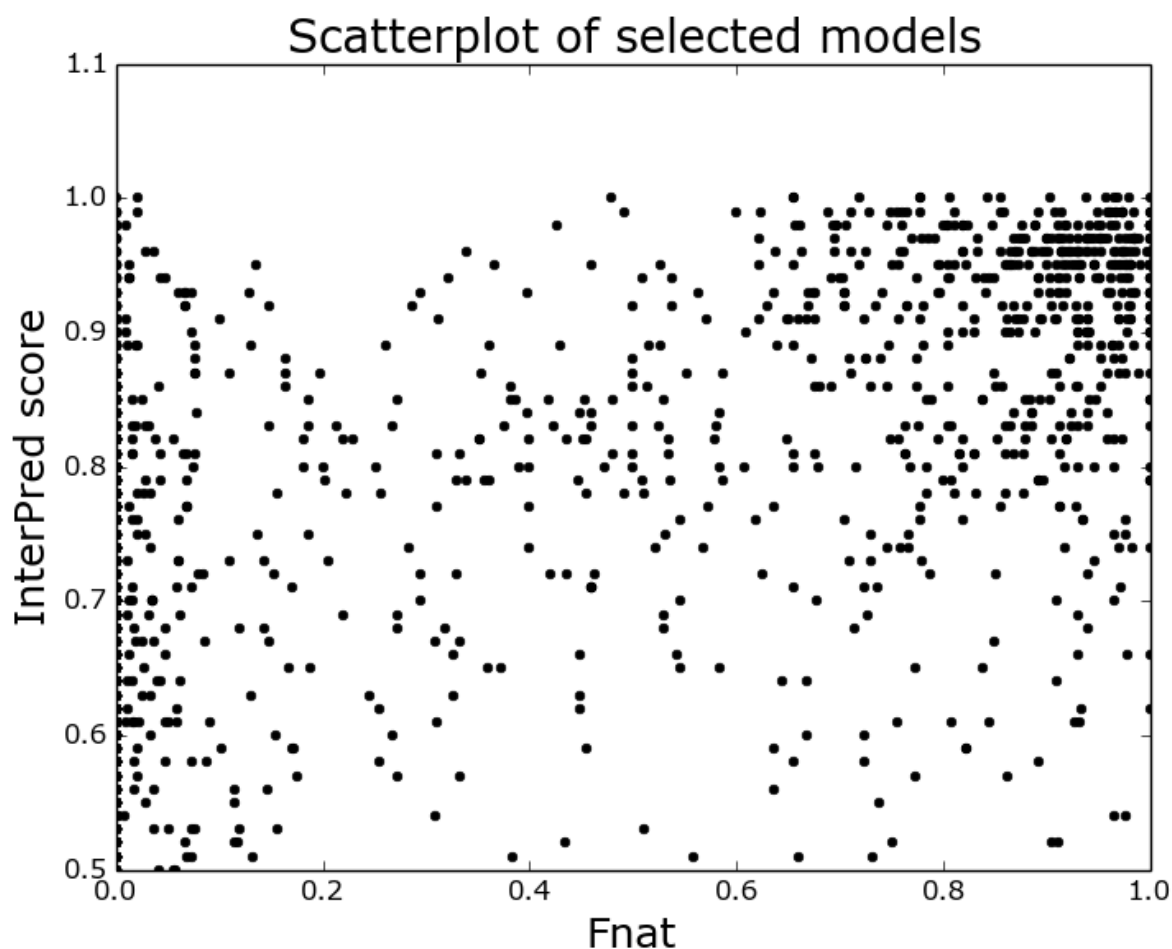


Figure 8: The starting points of all selected models can be seen in this figure. There is a spearman correlation of 0.547 between InterPred and starting point Fnat.

### 5.3 Models found by sequence similarity

Using sequence homology 18135 coarse interaction models were found. Of these 18,135 models 7,608 models overlapped with models found when InterPred used structural homology for predicting interactions. These models are considered easy models when represented in table 3 and 4.

### 5.4 Docking

#### 5.4.1 Successfully docked targets and models

The models are sorted according to their Rosetta score and the 10 lowest Rosetta energies were then tried according to the CAPRI quality assessment (table 1).

All models were then divided into bins either according to their Fnat or InterPred score before docking was performed (see table 3 and 4 respectively). The models were further divided depending on whether the model could have been found using sequence homology (easy) and models found using structure homology (default). Lastly the models were either docked using the pert 3 8 flag (perturbed) or docked without any additional conditions (default). A successful model is counted once in the CAPRI class of its highest scoring decoy unless none of the top ten decoys manages to fulfill the acceptable criteria in which case the model were counted as an unsuccessfully docked model. In figure 9 and 10 comparisons between InterPred and either docking success or docking quality can be seen for table 3 and 4 respectively.

Fnat limits	0.00-0.20	0.20-0.40	0.40-0.60	0.60-0.80	0.80-1.00
High quality CAPRI models					
Default	5	24	30	29	123
Easy default	0	13	30	81	267
Perturbed	12	35	33	33	120
Easy perturbed	2	16	36	102	293
Medium quality CAPRI models					
Default	0	2	2	0	2
Easy default	0	0	0	1	1
Perturbed	3	0	1	0	4
Easy perturbed	1	1	0	1	0
Acceptable quality CAPRI models					
Default	4	4	1	5	0
Easy default	1	4	1	1	7
Perturbed	1	0	2	4	5
Easy perturbed	7	0	1	4	7
Total amount of models fulfilling acceptable quality					
Default	9	30	33	34	125
Easy default	1	17	31	83	275
Perturbed	16	35	36	37	129
Easy perturbed	10	17	37	107	300
Total amount of easy and normal models of acceptable quality					
Total default	10	47	64	117	400
Total perturbed	26	52	73	144	429
Total number of tested models					
Models	729	43	38	40	137
Easy models	553	19	38	107	302
Total models	1282	62	76	147	439
Number of unsuccessfully docked models					
Undocked default	720	13	5	6	12
Undocked easy default	552	2	7	24	27
Undocked perturbed	713	8	2	3	8
Undocked easy perturbed	543	2	1	0	2
Total default undocked	1272	15	12	30	39
Total perturbed undocked	1256	10	3	3	10

**Table 3:** All models were divided into bins according to their Fnat score before docking was performed. The models were further divided depending on whether the model could have been found using sequence homology (easy) and models found using structure homology (normal). Lastly the models were either docked using the pert 3 8 flag (perturbed) or docked without any additional conditions (normal). A model is counted once in the CAPRI class of its highest scoring decoy.

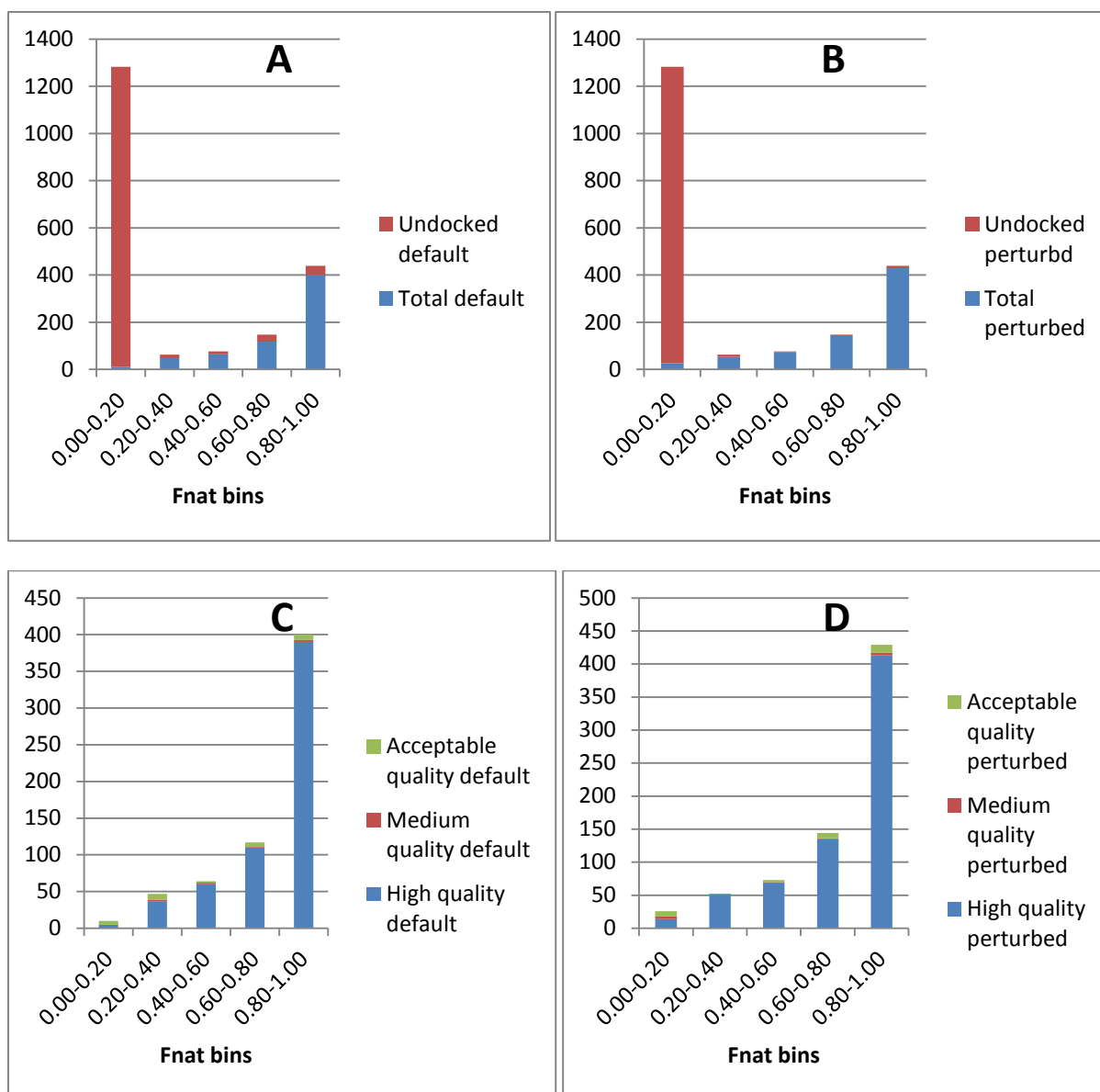
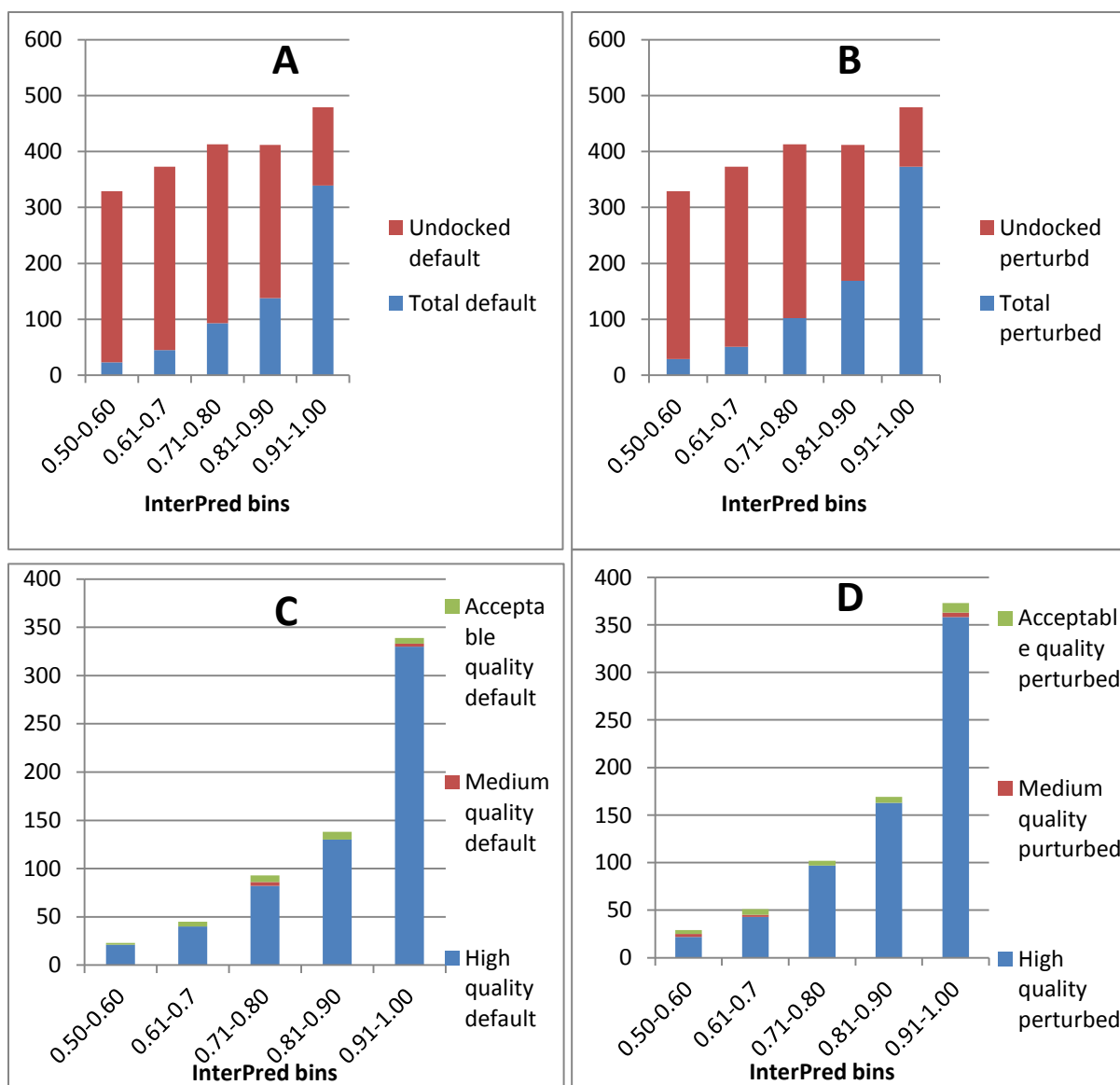


Figure 9: A) Models were divided into bins based on Fnat of the starting point and the amount of successfully docked models is compared to the amount of tried models. Red represents models which did not find any successfully docked decoys and blue shows successfully docked models. B) Same as A but for results yielded by perturbed docking instead of default docking. C) The CAPRI quality of the successfully docked models was decided and yielded figure C for the default docking. D) Same as C but for results that stem from the perturbed docking.



InterPred Limits	0.50-0.60	0.61-0.7	0.71-0.80	0.81-0.90	0.91-1.00
High quality CAPRI models					
Default	13	27	40	48	84
Easy default	8	13	42	82	246
Perturbed	13	29	49	57	85
Easy perturbed	9	14	48	106	273
Medium quality CAPRI models					
Default	0	0	3	0	2
Easy default	0	0	1	0	1
Perturbed	2	1	0	0	5
Easy perturbed	1	1	0	0	0
Acceptable quality CAPRI models					
Default	2	4	5	2	1
Easy default	0	1	2	6	5
Perturbed	2	1	2	3	4
Easy perturbed	2	5	3	3	6
Total amount of models fulfilling acceptable quality					
Default	15	31	48	50	87
Easy default	8	14	45	88	252
Perturbed	17	31	51	60	94
Easy perturbed	12	20	51	109	279
Total amount of Easy and normal models of acceptable quality					
Total default	23	45	93	138	339
Total perturbed	29	51	102	169	373
Total number of tested models					
Models	235	236	209	165	142
Easy models	94	137	204	247	337
Total models	329	373	413	412	479
Number of unsuccessfully docked models					
Undocked default	220	205	161	115	55
Undocked easy default	86	123	159	159	85
Undocked perturbed	218	205	158	105	48
Undocked easy perturbed	82	117	153	138	58
Total default undocked	306	328	320	274	140
Total perturbed undocked	300	322	311	243	106

**Table 4:** The models in this table are divided in the same way as table 3 except that the models were divided into InterPred bins instead of Fnaf. A model is counted once in the CAPRI class of its highest scoring decoy. For full legend see table 3



**Figure 10:** A) Models were divided into bins based on the InterPred score of the starting structure and the amount of successfully docked models is compared to the amount of tried models. Red represents models which did not find any successfully docked decoys and blue shows successfully docked models. B) Same as A but for results yielded by perturbed docking instead of default docking. C) The CAPRI quality of the successfully docked models was decided and yielded figure C for the default docking. D) Same as C but for results that stem from the perturbed docking.

Most targets had a mixture of successful and unsuccessful targets, one example of this is seen in figure 11 where all decoys for all models of the target 3k1i can be seen. The structure of one successful decoy can be seen in figure 12 where the docking algorithm had found a solution shown in green based on a difficult starting point seen in purple. The structures of the starting point and the decoy are aligned to the native structure shown in red.

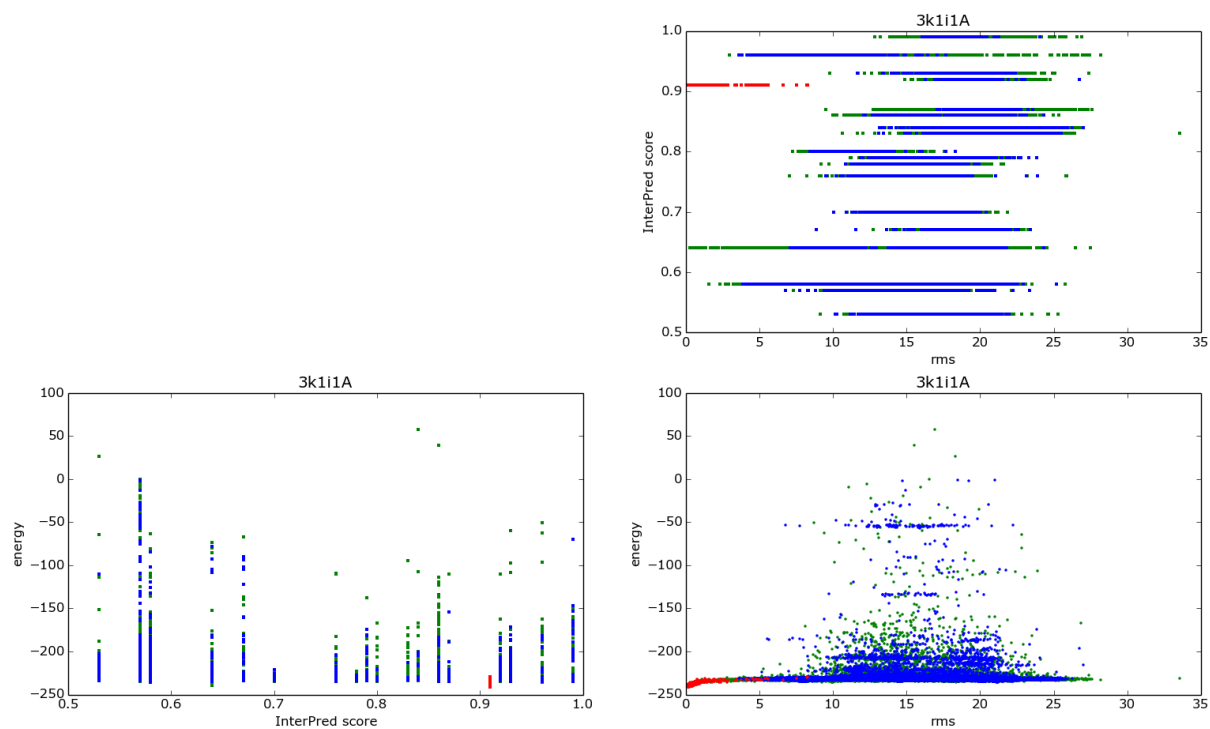


Figure 11: For target 3k1i one model manages to dock from a difficult position. The starting point, best decoy and native structure can be seen in figure 12. In this figure native decoys are red default docking decoys are blue and perturbed docking decoys are green.

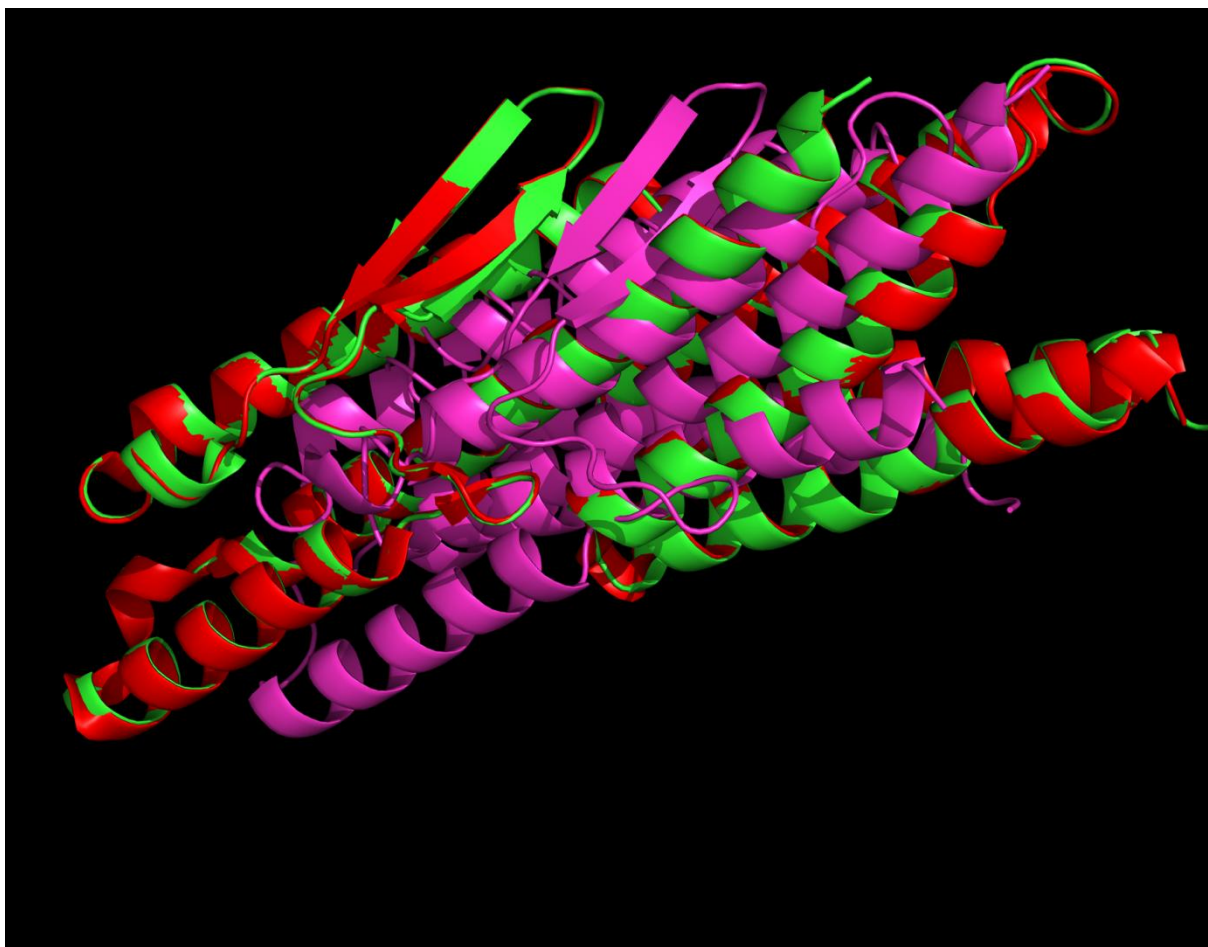


Figure 12: Shown in the figure is the successfully docked model structure of 3k1i. Where purple is the starting structure which was docked in Rosetta, the green structure is the best decoy yielded from the perturbed docking procedure and red is the target structure.

#### 5.4.2 Unsuccessfully docked targets

When comparing successfully docked models with a list of all models tried, it yielded table 5 that shows all targets where none of the top 10 energy decoys for any of the docked models associated with the target were good enough to fulfill the acceptable CAPRI criteria.

Targets which never docked		
Normal docking	Perturbed docking	Probable reason why no docking occurred
1qav	1qav	Difficult interaction site
1sb2	1sb2	Difficult interaction site
1wmh	1wmh	Unknown
2pqa	2qpa	Missing large ligand
2qby		Missing large ligand and DNA-chains Native never docks
2wbw	2wbw	Missing large ligand
2xg4		Missing large ligand and Native never docks properly
3e33	3e33	Missing large ligand and Native never docks properly
3kli		Huge interaction site and complex interaction

Table 5: All targets which never docked and a probable reason as to why they did not dock is provided in this table.

In figures 13-15 all decoys of all models for each selected target is shown. These 3 pictures depict cases where it was not possible to dock the target model (native3).

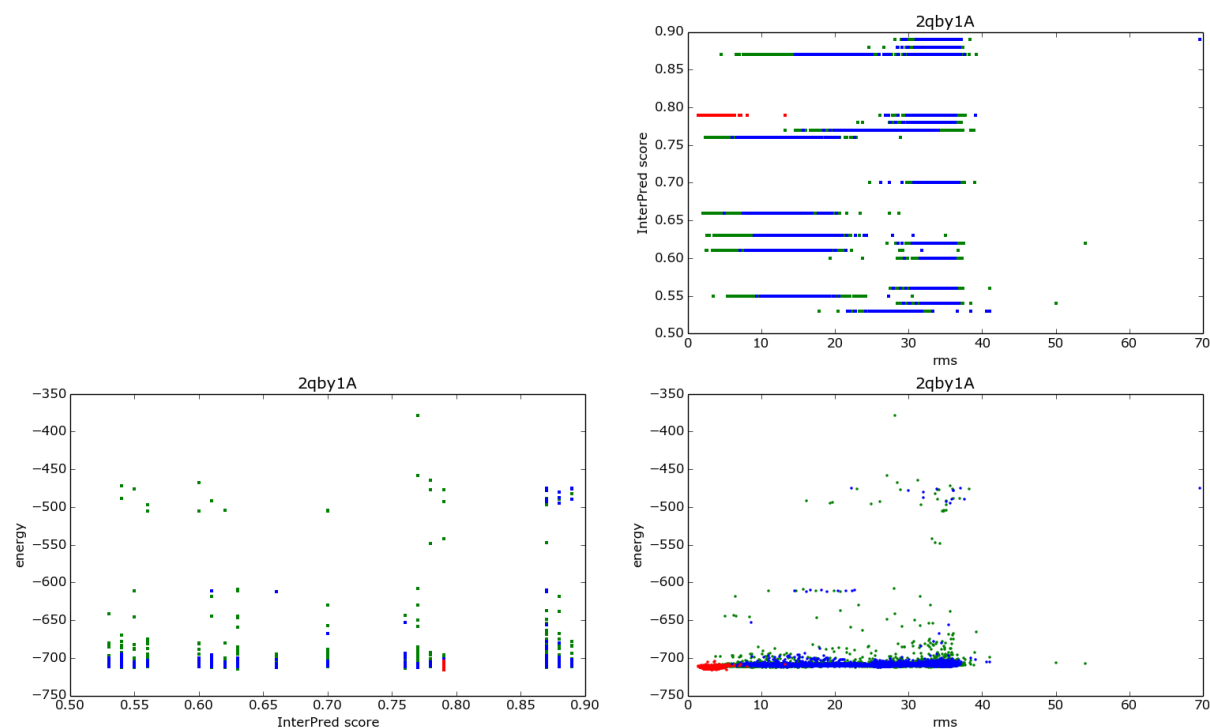


Figure 13: For target 2qby even though no docking funnel is observed there is at least one perturbed decoy which fulfills the acceptable CAPRI criteria. In the figure Native docking decoys are red, normal docking decoys are blue and perturbed docking decoys are green

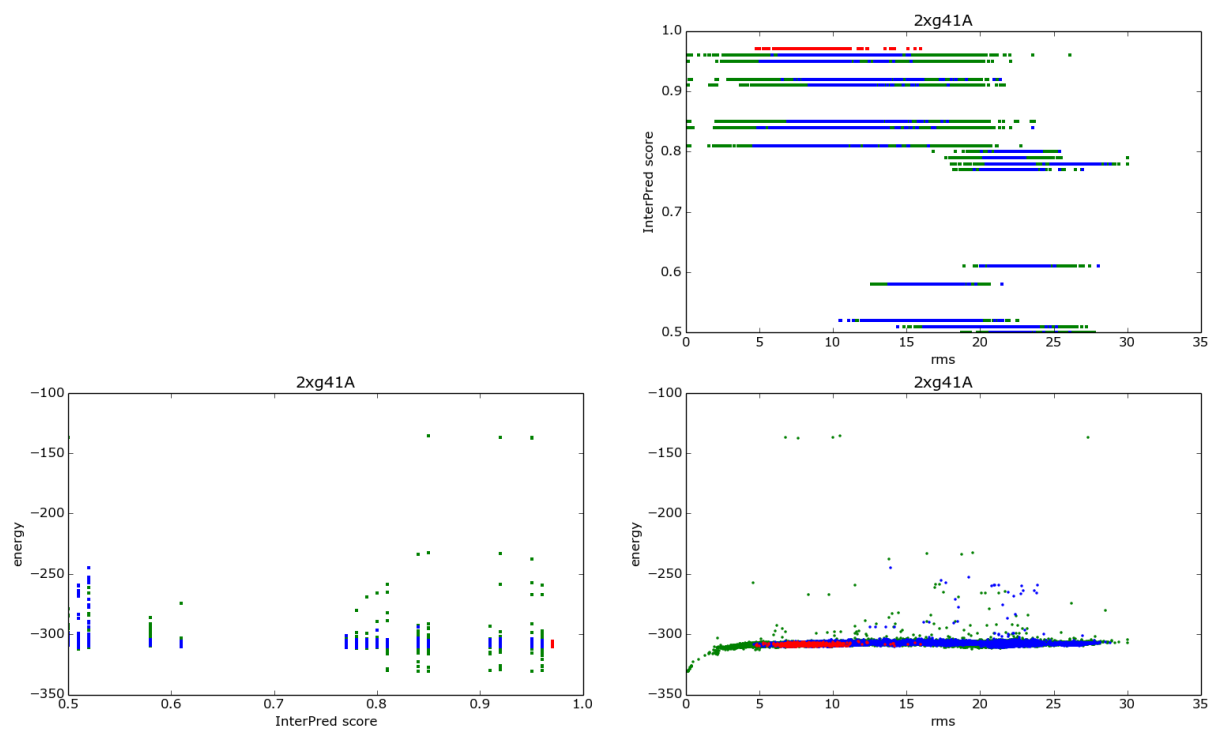


Figure 14: For target 2xg4 only perturbed models are able to dock. In the figure Native docking decoys are red, normal docking decoys are blue and perturbed docking decoys are green

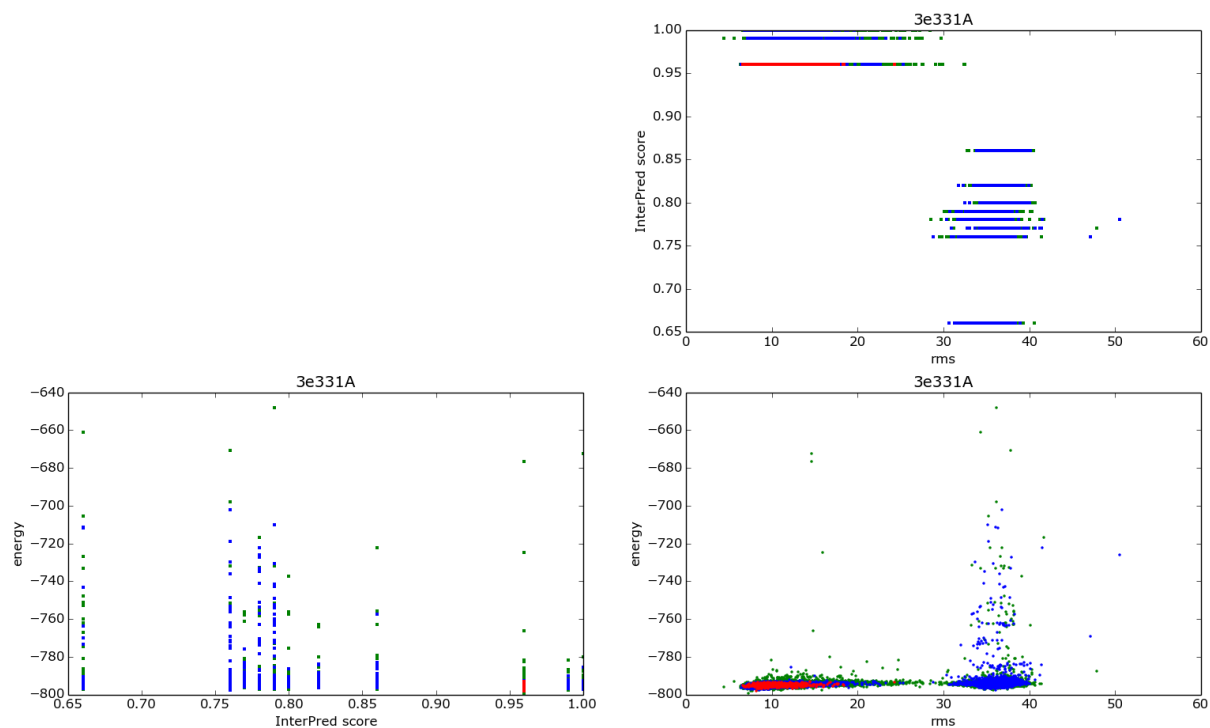


Figure 15: For target 3e33 no successful decoys are observed. In the figure Native docking decoys are red, normal docking decoys are blue and perturbed docking decoys are green

In figure 16 and 17 docking results for protein 1sb2 and 3e33 are shown. 1sb2 docked successfully from the native structure but neither normal docking nor perturbed docking yielded any successful decoy.

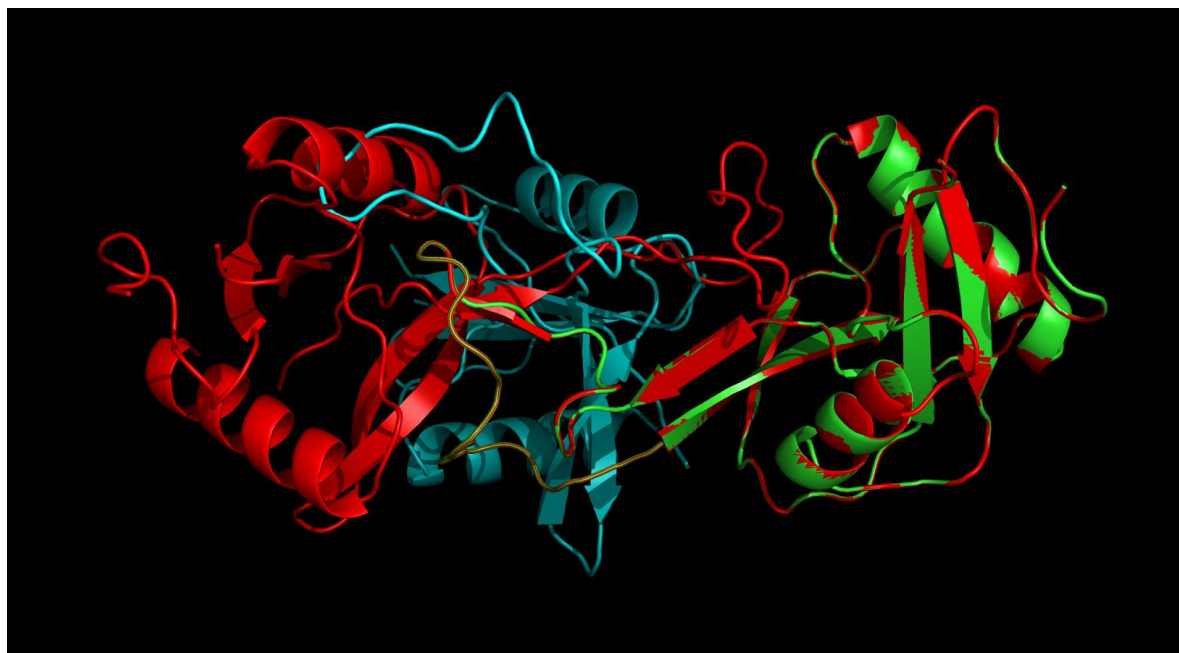


Figure 16: Protein structure of 1sb2 where red is the relaxed native, green and cyan show chain A and B respectively of the highest scoring decoy from the perturbed docking experiment. In this picture Chain A in the decoy were aligned to chain A in the relaxed native of 1sb2.

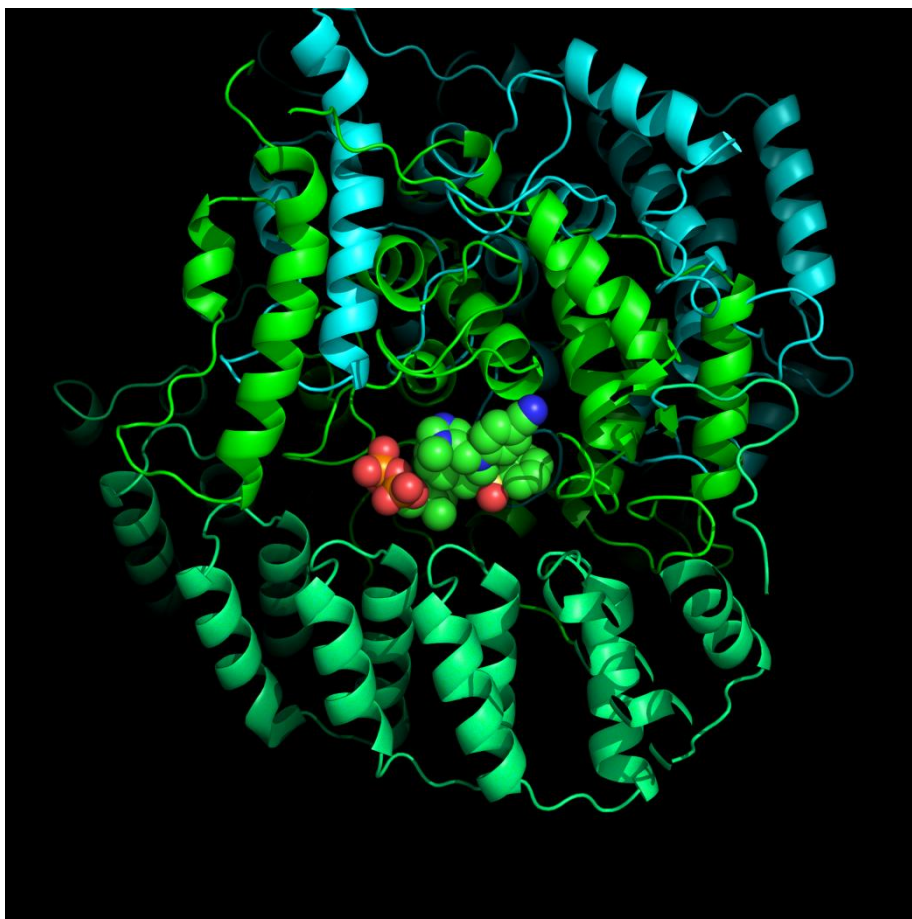


Figure 17: Protein structure of relaxed 3e33 and highest scoring decoy from docking with native3 as input. Overlapping proteins are showed in mint green, the green is part of the 3e33 native3 complex and the blue is the highest scored decoy of the 3e33 native. Shown in ballform in the figure are 2 large ligands which stabilize the crystal structure, but these are not present in the docking because Rosetta can only handle 2 chains.

## 5.5 Process analysis

The actual process (figure 18) is somewhat differentiated from the original planning. As it was found out in the initial testing a lot of work had to be done in order to prepare the data for the docking and as the initial testing dragged out on time it was decided to only perform the docking experiments in Rosetta, but use 2 different docking protocols in order to get a bit more information about docking of the coarse models. The work on methodology and the docking experiments as well as extracting results suitable for this report was all very time and focus consuming activities and it is for this reason that there is a lot less multitasking in the actual work than was planned. As I had close contact with both examiner and supervisor it became natural to hold the half-time presentation when the docking experiments were done instead of at the middle of the project.



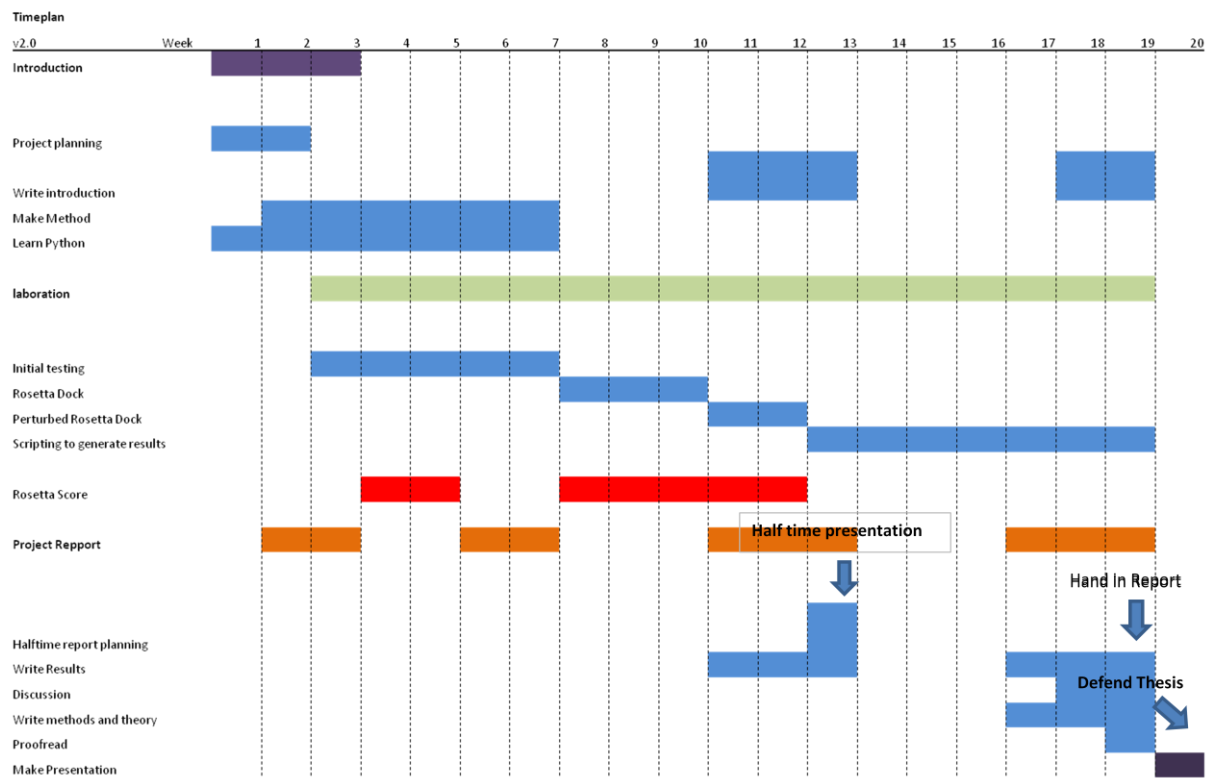


Figure 18: Followup of the timeplan. For full legend and comparison see figure 2.

## 6 Discussion

### 6.1 Process discussion

Because of the easy availability of both supervisor and examiner, the iterative process of working with computational methods and the dynamic nature of the project much of the planning work almost immediately became outdated. This stole a bit of time from the project and while it added a bit of structure and partial goals to work toward this might instead have been replaced by just that, partial goals and deadlines for certain activities. While the time between these goals could have been less planned. For example the timeplan could have been replaced with a list of goals and when certain checkpoints must be reached. The followup could have represented the actual work better as a tree with binary checkpoints. That would have better represented all the times progress in the work was dependent on some intermediate results and might have looked as figure 12 where the all crossroads in the project could be represented. The project could also have been planned in this form as well.

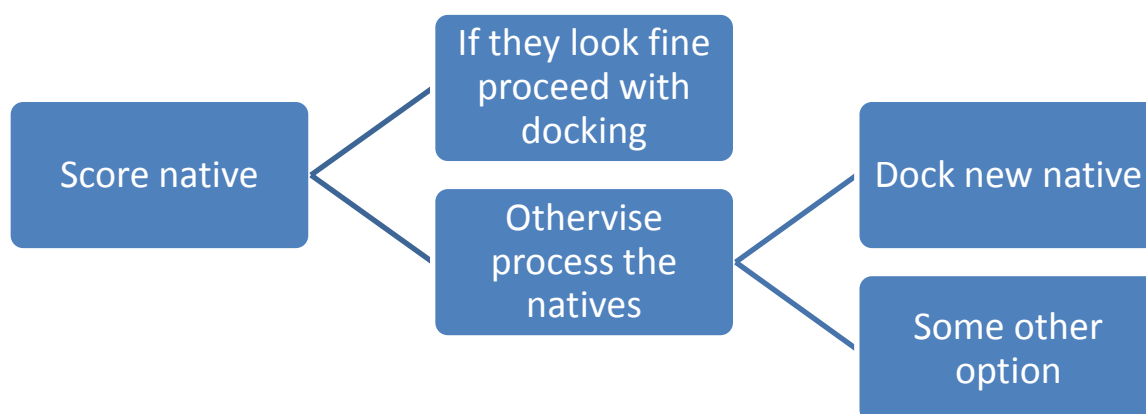


Figure 19: Example of how the flow of the process could have been represented.

### 6.2 Relaxation

Had the native structures not been relaxed the contained clashes would have made it extremely difficult to actually find good decoys since their scores would contain large clashes and models which scored well might be artifacts of the clashes. When relaxing the structures there were initially 3 different relaxed sets of native structures however the one represented in the methods was the one most straightforward and also the best one. It was also considered superfluous to the report to include information for the other relaxed sets, instead native3 was selected as the native set to build the coarse models from. In figure 5 where the distribution of native and native3 complexes it can be

seen that native3 has strictly negative scores which is good, however the deviations in Fnat is not optimal. But given that CAPRI considers a model with Fnat above 0.5 of high quality and the fact that the lowest Fnat is  $\sim 0.74$  all results can be considered a good representation of the native data.

When comparing the starting points for all 44509 coarse models made from native3 and native respectively in figure 6 the density of starting points is highly conserved. This is further visualized by figure 7 where all data seems to have a normal distribution around a ratio of 1:1 between relaxed and native Fnat. Even though there are 2 areas of higher density in figure 6 there are starting points distributed over the whole area. So there was a great foundation for choosing models to investigate the entire spectrum of InterPred.

### 6.3 Selected models

While both the number of models and targets shows a bias to the high InterPred interval it would be difficult to solve without either reducing the amount of targets to investigate or by removing the limitation on only 5 models for each InterPred bin. Choosing only targets with 5 models in each InterPred bin would have ensured that both the number of targets and the number of models were evenly distributed but would have been reduced the amount of targets greatly. The second solution would only have evened out the difference in the amount of models in the bins and would most likely have introduced a heavy bias towards the targets which had thousands of predicted homology models.

### 6.4 Docking

The docking in general shows a strong relation between easy models and the InterPred score. This is to be expected since InterPred gives higher scores the more similar the structures are and since structure is more conserved than sequence it becomes natural to extrapolate this.

Higher InterPred scores led to higher percent of successfully docked models as seen in figure 10. InterPred also showed a strong correlation with starting point Fnat (see figure 8), which signifies that InterPred is very good at predicting the quality of the templates and thereby the coarse model.

Looking at figure 9 and 10 it becomes apparent that there are more models which manages to fulfill the high quality criteria than there are models which only pass the medium or acceptable CAPRI criteria. This would suggest that there is plenty of room to use less good protein structures as inputs to InterPred i.e. protein structures which have large deviations from the target proteins in the interaction. Deviations could be caused by induced fit which changes the structure of proteins when they interact with each other. This could be used in a 2 step investigation where unbound targets are first docked and the best decoys are then refined allowing backbone atoms to be moved in order to model not only how the proteins interact but also how their interaction changes their structure.

#### 6.4.1 Successful docking

The success rate of the dockings are very high for the highest InterPred bin at 70.8% and 77.8% for normal docking and perturbed docking respectively when compared to the dockings of 14 different software reviewed in (Huang 2015) that report between 1.71% and 30.68% when investigating the top 10 predictions. In the review they do however use a different Benchmark (Chaudhury, Berrondo et al. 2011) that uses unbound targets. But considering that they generate more predictions for each

target and that 123 of the targets are considered easy targets, called rigid-docking targets in the benchmark, it suggests that InterPred based Rosetta docking is much more efficient. If one only consider the rigid-body results Huang reports slightly higher success rate but not enough to come close to this study. Also if a model fulfilling the high CAPRI standard in this study the model has Fnat of at least 0.5 and it is probable that Irms is at most 1.0 so if one simply added the average distortion of the medium quality in Huang's benchmark all high quality models in this study would still fulfill the acceptable CAPRI criteria and thereby still be counted as successfully docked. Making these comparisons to unbound docking it becomes apparent that CAPRI based comparison cannot be used for the most difficult cases. For example in the benchmark by Huang the difficult targets have an average Irms distortion of 3.45 Å, meaning a perfectly docked decoy from a model would yield an Irms score of 3.45 Å.

When models were divided according to Fnat of the starting coarse-model as seen in table 3 there is a much more pronounced difference between the bins both in terms of investigated models and the success rate (see figure 9) than when the models were divided based on InterPred (see figure 10). Thereby it would seem that choosing models to dock based on their initial Fnat is preferable to selecting models with high InterPred scores. So whenever Fnat of the initial starting position is available it should be used. This is however only viable for these benchmark studies where docked target structures exist. This option is therefore of little use for the sake of using computational methods to fill the holes in the knowledge about structural PPIs.

#### 6.4.2 Unsuccessful docking

There are primarily 2 reasons that no model was able to find the native structure of its target, either the lack of a large ligand or a very large/complex interaction between the proteins in a target.

Large interactions are simple to identify, using software to simply count the distances between atoms in the structures and then simply classify them as residue-residue contacts if they are within 5 Å from each other. The complexity however is harder to predict and can only be estimated by observing a large number of decoys with large repulsion scores. And in both these cases it would probably be necessary to add constraints to the docking protocol in order to successfully dock such targets.

It is noteworthy that using the `-pert 3 8` flag made it possible to successfully dock at least one decoy in one model in 2 out of the three cases when docking the native could not yield a single decoy with RMSD lower than 1.0 Å as seen in figure 11-13. As the native3 docking decoys usually generated decoys with RMSD values of close to 0 Å.

### 6.5 Future prospects and conclusions

For future prospects it would be of interest to try docking unbound targets from other Benchmarks or the models of lesser qualities in the Dockgrounds model set, in order to simulate the docking of induced fit interactions.

It would be especially interesting to try using the models with 3 or 4 Å deviation from the crystal structure in Dockground model set v2.0. In order to see if there is an equal amount of acceptable models in that study as there are high quality models in this study.

The method of docking with the –pert 3 8 flag works really well for models with high InterPred and it should be sufficient to dock the top 5 models with a thousand decoys each and then compare the top 5 generated decoys to generate a successful interaction model.

Whenever possible the structure of a known ligand interacting with the complex, it should be added to the chain of one of the proteins in order to reduce the chance of unsuccessful dockings because of missing ligands (see figure 17).

Given the low success rate of docking with decreasing InterPred it could be extrapolated that using random starting points should make docking near impossible with only 1000 decoys. However to prove this chosen models should be docked with random starting orientations in order to verify this.

## 7 Acknowledgments

First of all I would like to thank Björn Waller for the opportunity to perform my master thesis in his group.

Secondly I would like to thank Claudio Mirabello for being patient, educational and providing excellent support.

I would also like to thank my two roommates Robert Pilstål and Sankar Basu for help and good company.

Lastly I would like to thank my opponent Fredrik Söderqvist for scrutinizing my report and giving valuable feedback.

## References

ANISHCHENKO, I., KUNDROTAS, P.J., TUZIKOV, A.V. and VAKSER, I.A., 2015. Protein models docking benchmark 2. *Proteins: Structure, Function, and Bioinformatics*, **83**(5), pp. 891-897.

BUTLER, B.M., GEREK, Z.N., KUMAR, S. and OZKAN, S.B., 2015. Conformational dynamics of nonsynonymous variants at protein interfaces reveals disease association. *Proteins: Structure, Function, and Bioinformatics*, **83**(3), pp. 428-435.

CHAUDHURY, S., BERRONDO, M., WEITZNER, B.D., MUTHU, P., BERGMAN, H. and GRAY, J.J., 2011. Benchmarking and analysis of protein docking performance in Rosetta v3.2. *PLoS one*, **6**(8), pp. e22477.

GRAY, J.J., MOUGHON, S., WANG, C., SCHUELER-FURMAN, O., KUHLMAN, B., ROHL, C.A. and BAKER, D., 2003. Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. *Journal of Molecular Biology*, **331**(1), pp. 281-299.

HUANG, S., 2015. Exploring the potential of global protein–protein docking: an overview and critical assessment of current programs for automatic ab initio docking. *Drug discovery today*, **20**(8), pp. 969-977.

JANIN, J., 2010. Protein-protein docking tested in blind predictions: the CAPRI experiment. *Molecular bioSystems*, **6**(12), pp. 2351-2362.

MELQUIOND, A.S.J., KARACA, E., KASTRITIS, P.L. and BONVIN, A.M.J.J., 2012. Next challenges in protein–protein docking: from proteome to interactome and beyond. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, **2**(4), pp. 642-651.

MÉNDEZ, R., LEPLAE, R., DE MARIA, L. and WODAK, S.J., 2003. Assessment of blind predictions of protein–protein interactions: Current status of docking methods. *Proteins: Structure, Function, and Bioinformatics*, **52**(1), pp. 51-67.

Mirabello, C; Wallner, B; Hyvönen, M; InterPred: A pipeline to predict and model protein-protein interactions, manuscript in preparation

ROY, A., KUCUKURAL, A. and ZHANG, Y., 2010. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat.Protocols*, **5**(4), pp. 725-738.

SAFARI-ALIGHIARLOO, N., TAGHIZADEH, M., REZAEI-TAVIRANI, M., GOLIAEI, B. and PEYVANDI, A.A., 2014. Protein-protein interaction networks (PPI) and complex diseases. *Gastroenterology and Hepatology From Bed to Bench*, **7**(1), pp. 17-31.

SCHNEIDMAN-DUHOVNY, D., INBAR, Y., POLAK, V., SHATSKY, M., HALPERIN, I., BENYAMINI, H., BARZILAI, A., DROR, O., HASPEL, N., NUSSINOV, R. and WOLFSON, H.J., 2003. Taking geometry to its edge: Fast unbound rigid (and hinge-bent) docking. *Proteins: Structure, Function, and Bioinformatics*, **52**(1), pp. 107-112.

SHOEMAKER, B.A. and PANCHENKO, A.R., 2007. Deciphering protein-protein interactions. Part I. Experimental techniques and databases. *PLoS computational biology*, **3**(3), pp. e42.

WANG, P.I. and MARCOTTE, E.M., 2010. It's the machine that matters: Predicting gene function and phenotype from protein networks. *Journal of proteomics*, **73**(11), pp. 2277-2289.

WHITFORD, D., 2005. *Proteins: Structures and Function*. First Edition edn. John Wiley & Sons.

WODAK, S.J., VLASBLOM, J., TURINSKY, A.L. and PU, S., 2013. Protein–protein interaction networks: the puzzling riches. *Current opinion in structural biology*, **23**(6), pp. 941-953.

