SAPIENZA
Università di Roma

PhD Programme in Life Sciences

XXXII Cycle

Integrative modelling of the antibody antigen interactions

Candidate

Francesco Ambrosetti

Supervisor

Prof. Francesca Cutruzzolà

Tutor

Prof. Alexandre MJJ Bonvin

Coordinator

Prof. Marco Tripodi

"*War is peace.*
Freedom is slavery.
Ignorance is strength."

— George Orwell, 1984

Abstract

Background

Over the years the use of monoclonal antibodies (mAbs) for therapeutic purposes has been experiencing a significative boost.

The reasons behind this enormous growth reside in both their inherent structural properties and in technological advances for their production and characterization. Indeed, their intrinsic high affinity and specificity toward a specific antigen, together with their modular anatomy, which largely simplify their engineering, are the most important drivers of this great market expansion.

Antibody-based therapeutics are developed via well-established processes that can be broadly categorized into Lead Identification and Lead Optimization. During Lead Identification animal immunization or surface display technologies are used to generate a large number of 'hit' molecules which need to be further triaged. Following various rounds of screening and design during Lead Optimization, a small number of high affinity lead candidates are selected. During Lead Identification and Optimization, molecules are assessed for unfavourable characteristics such as immunogenicity or poor biophysical properties and experimental methodologies have been developed in order to improve such proprieties. A sound knowledge of the antibody binding site and the relationship between antibody and antigen residues is of paramount importance for the effective design of such strategies.

Structural experimental techniques such as Nuclear Magnetic Resonance (NMR) or X-ray crystallography can be used to study antibody-antigen interactions but they are usually expensive and time consuming and not always applicable. Therefore, the development of computational methods is offering an attractive alternative and a faster approach for the characterization of antibodies and their interactions. Among those, docking approaches offer a valuable

strategy to elucidate the interaction between antibodies and antigens providing a tool to understand the role played by each residue in the binding.

Despite great progresses in protein-protein docking in general, modelling of antibody-antigen complexes, which is a specialized application of the broader field of molecular docking, has been demonstrated to remain challenging.

Aim

This work aimed at investigating how information about the antibody paratope and the antigen epitope can be successfully used into an information-driven docking algorithm to characterize the molecular interactions between antibodies and antigens.

In order to be able to use more accurate information about the antibody binding site, this work also aimed at improving proABC, a paratope prediction tool, and at demonstrating how this information impacts antibody-antigen molecular docking.

The overarching aim is to deepen our understanding of antibody-antigen recognition process and to provide computational tools and strategies that could facilitate antibody design and engineering.

Results

In this work I describe how information on the antibody hypervariable loops and the binding epitope can be effectively used to drive the modelling of their interaction by docking. In particular, I compare the accuracy of four docking software: ClusPro, HADDOCK, LightDock and ZDOCK in predicting antibody-antigen complexes. HADDOCK, which applies a purely data-driven strategy, performs better than the other systems especially when information about the epitope is provided.

I also investigate the accuracy of HADDOCK and LightDock (the only two software which allow flexibility of the components among the tested ones) in predicting the conformational change of the H3 loop, which is crucial for the binding and notoriously challenging to model. The results, described in the related paper "*Modelling antibody-antigen complexes by information-driven docking*", confirm that HADDOCK shows the best accuracy and that, even if to some extent it is not able to correctly model the conformation of the loop, it manages to predict the right H3 interactions with the antigen, providing valuable information for antibody engineering.

To further improve the accuracy of the docking, I developed *proABC-2* (*PRediction Of AntiBody Contacts v2*), which is an update, based on a deep learning framework, of the algorithm implemented in proABC originally developed by Olimpieri and co-workers (Olimpieri et al., 2013). The method is able to accurately predict the specific paratope residues and the type of interaction, starting from the antibody sequence alone and without additional information about the antigen. I demonstrate how its predictions can be effectively used to drive the docking using HADDOCK in order to improve its modelling accuracy. proABC-2 is freely available and accessible for academic purposes as a web server.

Publications

Miotto, M., Olimpieri, P.P., Di Rienzo, L., **Ambrosetti, F.**, Corsi, P., Lepore, R., Tartaglia, G.G., and Milanetti, E. (2019). Insights on protein thermal stability: a graph representation of molecular interactions. Bioinformatics *35*, 2569–2577.

Koukos, P.I., Roel-Touris, J., **Ambrosetti**, **F.**, Geng, C., Schaarschmidt, J., Trellet, M.E., Melquiond, A.S.J., Xue, L.C., Honorato, R.V., Moreira, I., Kurkcuoglu, Z., Vangone A., and Bonvin, A.M.J.J. (2019). An overview of data-driven HADDOCK strategies in CAPRI rounds 38-45. BioRxiv 10.1101/718122v1.

Norman, R.A., **Ambrosetti, F.**, Bonvin, A.M.J.J., Colwell, L.J., Kelm, S., Kumar, S., and Krawczyk, K. (2019). Computational approaches to therapeutic antibody design: established methods and emerging trends. Brief. Bioinform. *Advanced Online Publication.*

Ambrosetti, F., Jiménez-García, B., Roel-Touris, J. and Bonvin, A.M.J.J. (2019). Modelling antibody-antigen complexes by information-driven docking. Structure, *Advanced Online Publication* (2019).

Ambrosetti, F., Olsen, T.H., Olimpieri, P.P., Jiménez-García, B., Milanetti E., Marcatilli, P. and Bonvin, A.M.J.J. (2019). proABC-2: PRediction Of AntiBody Contacts v2 and its application to information-driven docking. *Manuscript in preparation*

1. Chapter I

1.1 General introduction

Our ability to survive even in very unfavourable environments relies on the capacity of our organism to adapt itself according to the external conditions. To achieve this degree of flexibility different systems are finely tuned with one of the most important ones being known as the immune system. It allows us to fight and defend ourselves against pathogens and infections, being able to distinguish between "self" and "non self". Two main responses are used by the immune system to fight extraneous attacks: innate and adaptative. As suggested by the names, the former refers to all of those mechanisms that are innately present to fight different ranges of threats, while the latter consists of all of the specific strategies developed by the organism upon interaction with a given pathogen. The innate system is designed to directly react to a threat, while the adaptive one requires some time to generate and fine tune the protective response. Antibodies (or immunoglobulins) are the result of this adaptive strategy and their role is crucial in order to identify and neutralize foreign organisms by recognizing specific antigens.

The ability of antibodies to recognize and bind with high affinity and specificity a given antigen, along with their highly modular anatomy which largely facilitate their engineering and design, makes them valuable weapons to fight against specific diseases. Therefore, since their discovery in the 20th century by Paul Ehrlich they attracted the attention of researchers and industries in a wide range of different applications.

Nowadays immunoglobulins have revolutionized the medicine with more than 570 antibody therapeutics studied at various clinical phases in 2019, five of the top selling blockbusters are monoclonal antibodies (mAbs) and their market presence is expanding (Kaplon and Reichert, 2019). This rising trend in the use of antibodies for therapeutic purposes mainly relies on faster and more efficient design strategies to study and characterize them. Along with this, the development of in vitro phage display technologies and the generation of transgenic mice expressing human antibodies has considerably boosted and facilitated the production of highly specific antibodies (Larrick et al., 2016).

Over the years, experimental techniques have been developed and optimized in order to get insights into immunoglobulin structures and properties. However, often such experimental methodologies are expensive and time-consuming. Computational approaches are providing a valuable alternative to standard experimental methods, allowing faster, easier and cheaper characterization of immunoglobulins and the way they interact with their cognate antigen. One major driver of this shift is the increasing number of sequence and structural data which is paving the way for the improvement of computational methods in order to increase their accuracy and reliability.

Despite these developments, current computational approaches, as will be discussed in the following chapters, still show limitations due to the inherent properties of antibodies and several challenges (e.g. the prediction of the antibody antigen complexes and the modelling of the H3 loop) that still need to be overcome. Therefore, new approaches are required to tackle these problems taking advantage from the large amount of data presently available.

1.2 Antibody structure

Antibodies, or Immunoglobulins (Ig), are "Y-shaped" globular proteins produced in jawed vertebrates by B-cells in response to an external threat for the organism. Most of the times they are composed of two couples of identical polypeptide chains named heavy (H) and light (L) chain, kept together by disulphide bonds. In humans, the H and L chains can naturally assemble into five isotypes: IgG, IgD, IgE (monomeric forms), IgA (dimers) and IgM (pentamers) according to their heavy chain class, respectively: γ, δ, ε, α or μ. For the light chain, there are four different isotypes but only two of them are present in mammals: κ and λ (Schroeder and Cavacini, 2010).

Antibodies consist of one crystallizable (F_c) and two antigen binding (F_{ab}) fragments as illustrated in Figure 1. The F_{ab} domain contains the heavy and light variable segments (V_H and V_L) which are the regions involved in the recognition of the antigen. The V_H and V_L domains show very high sequence and structural variability among the antibody repertoire. In particular, both heavy and light chain variable segments contain three regions where the sequence displays the highest variability. These regions are known with the name of *Complementarity Determining Regions* (CDRs). The remaining part of the variable domain is called *Framework* and is characterized by a high sequence conservation. Its tertiary structure consists of a 2-layer sandwich of 7-9 antiparallel β-strands arranged in two β-sheets forming the so-called *immunoglobulin-like fold*.

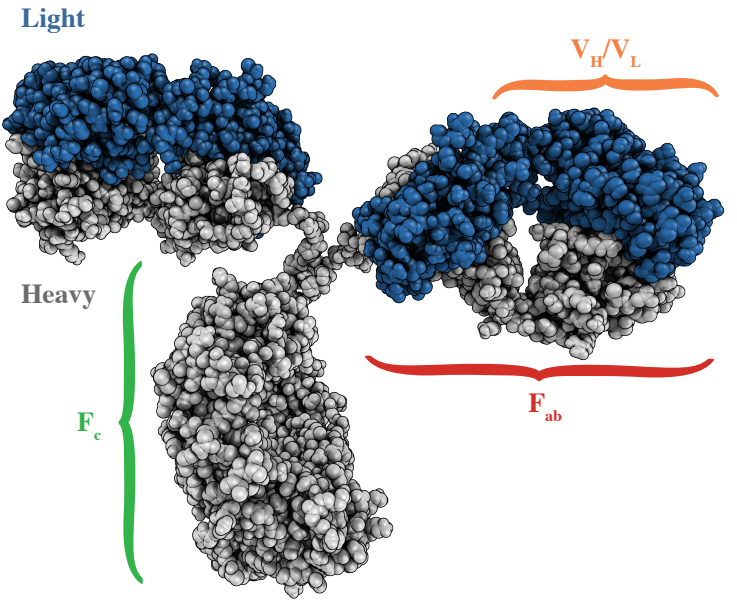

Figure 1: View of an immunoglobulin structure (PDB code: 1IGT). The two heavy chains are shown in grey while the light ones are represented in blue. The variable (VH/VL) domain and the binding (F_{ab}) and crystallizable (F_c) fragments, are highlighted in orange, red and green, respectively.

Each CDR harbours one loop, also named hyper variable loop (HV), for a total of six loops, three for the heavy chain (H1, H2, H3) and three for the light chain (L1, L2, L3) (Te Wu and Kabat, 1970). They play a crucial role in the binding of the antigen (Novotný et al., 1983). Despite their intrinsic high sequence variability, as demonstrated by Chothia and co-workers (Chothia et al., 1989), five out of the six HV loops exhibit a small set of well-defined conformations according to the length and to the presence of specific residues located in key positions of the antibody variable domain.

H3 is the loop which shows the highest structural and sequence diversity and it is also the one that is mainly involved in the antigen recognition (Shirai et al., 1996; Weitzner et al., 2015). Its modelling is one of the main challenges in the antibody modelling field. Structurally, is divided into two regions called "*torso*" and "*apex*" respectively more proximal or distal to the antibody framework. While for the apex it is not possible to define any "canonical conformation", the torso can be classified into "*bulged*" on "*non bulged*" depending on the presence and absence of arginine and aspartate in defined positions of the framework (Morea et al., 1998).

1.3 Production of antibodies

Production of antibodies is a complex and very well-regulated procedure. Indeed, the immune system has to be ready to fight with high specificity against a huge range of external (virus, bacteria, parasites) and internal threats (cancer cells). To do that the organism must be able to generate billions of different B-cells capable of releasing antibodies in response to a particular antigen. It has been estimated that around $5x10^9$ B-cells are present in an organism (Briney et al., 2019) producing distinct B-cell receptors (membrane or bound) or antibodies (in soluble form). The mechanism that underlies the generation of a such large volume of immunoglobulins based on the limited genomic space involves the somatic recombination of different gene segments, namely: variable (V), diversity (D), joining (J) and constant (C) segments. The overall process is known with the name of *V(D)J recombination* (Hesslein and Schatz, 2001; Tonegawa, 1983).

These genomic segments are present in clusters and located in different loci situated in different chromosomes. In total in the human organism it is possible to find three different loci: one for the heavy chain, one for the isotype λ of the light chain and one for the κ one (Bassing et al., 2002).

The heavy chain locus shows in sequence the V, D and J segments followed by the C one. The segments V, D and J encode for the entire variable domain of the heavy chain while the C segment is responsible for the constant domain, therefore it defines the immunoglobulin class. The two loci for the light chain include a group of V segments followed by the J ones. They are responsible for the first and last part of the light chain variable domain. The position of the C segments is different depending on the specific light chain isotype. In the lambda locus one single C region is located after the J

segments while in the case of kappa the C segments are alternated with the J ones.

The V(D)J recombination occurs together with the maturation of the B-cell. It starts when a specific protein complex called *V(D)J recombinase* recognizes conserved regions named *recombination signal sequences* (RSS) surrounding two D and J segments on the heavy chain locus. This recognition leads to the formation of a ring structure in which a D and J segment are close to each other. Then the V(D)J recombinase cleaves and joins together two random D and J segments. This process is not accurate, with some nucleotides being added or deleted, and largely contributes to the final variability of the antibody repertoire.

After the DJ recombination the two segments now joined together are successively combined to a random V segment leading to the formation of the so-called *germline sequence* (*IGHV*) which encodes for the full variable domain of the immunoglobulin heavy chain. Finally, the splicing of the corresponding RNA transcript assembles together the VDJ segment and the C one resulting in the final generation of the full functional heavy chain (Schatz and Swanson, 2011).

Once the heavy chain assembling is completed the rearrangement of the light chain can take place following similar steps.

The V(D)J recombination process is already able to introduce some level of variability in the generation of the immunoglobulins. However, it is not able on its own to explain the huge variety of the antibody repertoire. Indeed, upon antigen exposure, the antibody-producing B-cells undergo a natural process of affinity maturation, based on somatic hypermutation (Peters and Storb, 1996). This process introduces mutations primarily in the CDR regions and in the HV loops developing a specific and high-affinity binder. Together with the diversity introduced by V(D)J recombination, somatic

hypermutation can produce an estimated 10^{12}-10^{15} possible antibody sequences (Glanville et al., 2009) increasing the probability of recognizing an arbitrary foreign antigen.

1.4 Antibody numbering

In order to study immunoglobulins on a large scale it is necessary to map the antibody sequence onto a standardized framework to allow the unambiguous identification of residues which show equivalent structural position by assigning a unique identifier to the amino acids of the variable domain. This is particularly relevant given the high similarities between antibody sequences and the presence of conserved residues in specific key positions. The numbering schemes contextualize each position within the structure of an antibody, allowing for rapid delineation of CDRs and framework regions. Since the seminal work to define a standard numbering scheme for antibodies was carried out by Kabat in 1970 (Te Wu and Kabat, 1970), the Chothia (Chothia and Lesk, 1987) and IMGT (international ImMunoGeneTics information system) (Lefranc, 2011) schemes have been adopted as the main alternatives. Additional numbering schemes such as Contact (MacCallum et al., 1996), North (North et al., 2011), WolfGuy (Bujotzek et al., 2015) and Aho (Honegger and Plückthun, 2001) exist but these are less prevalent.

Kabat and IMGT definitions are based on sequence alignments identifying conserved positions in the variable region whereas Chothia, which is a modification of the first Kabat scheme, takes into account the 3D structure of the CDR loops.

In this work we are following the Chothia numbering scheme. More specifically, it assigns a consecutive number to all the framework residues of the light and heavy variable domain up to the ones that are part of one HV loop. From them, it continues by adding to the number of the last framework residues a letter (which identifies an insertion) until the end of the loop, then the consecutive numbering starts again (see Figure 2).

There are three freely available software packages to perform numbering of antibodies; ANARCI (Dunbar and Deane, 2016), Abnum (Abhinandan and Martin, 2008) and AbRSA (Li et al., 2019), to act as the first step in computational antibody analysis.

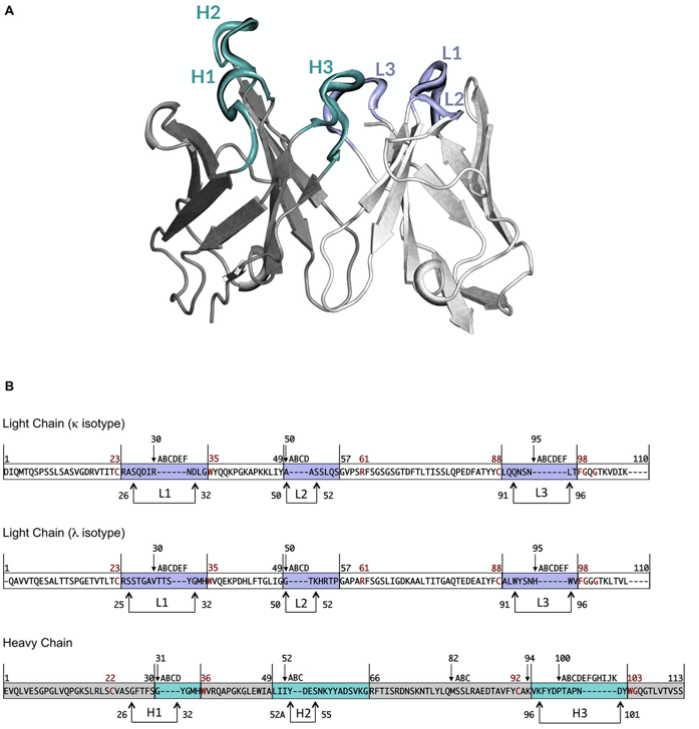

Figure 2: (**A**) View of the variable domain of an antibody. CDRs are highlighted. (**B**) Chothia numbering scheme for the heavy and light chain (κ and λ isotypes). Arrows represent the hypervariable loops. In violet and cyan are represented the CDRs of the light and heavy chain, respectively. Framework residues are coloured in white and grey for the light and heavy chain respectively. Conserved residues are reported in red.
Figure from: *PIGSPro: Prediction of immunoGlobulin structures v2. Nucleic Acids Res., 45, W17–W23*.

2. Chapter II

2.1 Molecular docking

Characterization of the interface between two molecules is a key step in order to elaborate efficient and effective rational design strategies for therapeutics. Several experimental techniques can be used to determine the three-dimensional (3D) structure of molecular complexes. These methods include X-ray crystallography, Nuclear Magnetic Resonance (NMR) and Cryo-Electron Microscopy (Cryo-EM). Despite their reliability and accuracy, experimental methods are generally time consuming and expensive. Therefore, computational approaches can provide a valuable, more rapid solution to gain information about the interacting residues of two molecules.

In this scenario, one suitable approach is "*molecular docking*". It can be defined as the process that, starting from the free form structures of the components, predicts the structure of the complex. It usually involves two different steps: the *sampling step*, during which thousands of possible complex conformations are generated, and the *scoring step*, in which the conformations are ranked according to a specific scoring function to identify models which are closer to the native conformation.

According to the sampling strategy used during the simulation, docking methods can be classified into two categories. The first class includes algorithms which perform a global search around the entire surface of the components without taking into account any information about the binding region (*ab initio docking*). The second class consists of docking methods that can use experimental data, for example coming from hydrogen-deuterium exchange (HDX) coupled with mass spectrometry, from mutational studies (Sevy et

al., 2013), or from predicted information about the binding interface to drive the sampling during docking (*information-driven, local or integrative docking*) (Rodrigues and Bonvin, 2014). Both classes can benefit from available information during the scoring step to select models that are consistent with the available information about the interaction.

Most protein-protein docking algorithms do not consider possible conformational changes occurring upon binding (rigid-body docking). This is the case for software such as ClusPro (Kozakov et al., 2017) and ZDOCK (Chen and Weng, 2002) that are based on the Fast Fourier Transform (FTT) search algorithm (Katchalski-Katzir et al., 1992). In most cases, however, protein flexibility is a crucial factor to be considered (Kotev et al., 2016). Approaches that allow for flexibility of side chains and/or backbone have also been developed such as ATTRACT (De Vries et al., 2015), LightDock (Jiménez-García et al., 2018), Swarmdock (Torchala et al., 2013) SnugDock (Sircar and Gray, 2010) and HADDOCK (De Vries et al., 2010). The former three do that by using normal modes, the latter two by allowing some flexibility along side-chains and backbone during a refinement stage.

The performance of various docking software is regularly assessed by the Critical Assessment of Predicted Interactions (CAPRI) experiment (Méndez et al., 2003), catalysing the effort of researchers towards the development of new and more accurate methods.

2.2 Docking of antibody-antigen complexes

The possibility to provide information either at the sampling and/or at the scoring level is particularly relevant in the case of antibody-antigen docking as CDRs and in particular the HV loops offer a reasonable proxy of the binding interface on the antibody. In fact, some docking methods such as for example ClusPro and PatchDock (Comeau et al., 2004; Schneidman-Duhovny et al., 2005) are able to automatically define the antibody CDRs in order to use this information during the docking process.

Despite great progresses in predicting protein-protein complexes, docking of antibody-antigen complexes remains challenging (Pedotti et al., 2011; Ponomarenko and Bourne, 2007; Vajda, 2005) due to the specific properties of their interfaces (Conte et al., 1999; Sela-Culang et al., 2013).

Understanding the structural basis of the antibody-antigen interaction would pave the way to the design of new efficient biological drugs (Lippow et al., 2007). For the antibody, the residues involved in the binding – the so called paratope residues – can be predicted quite accurately through various computational approaches (Krawczyk et al., 2013; Kunik et al., 2012; Liberis et al., 2018; Olimpieri et al., 2013). The information provided by those methods has been demonstrated to be valuable to drive the docking (Liberis et al., 2018).

The identification or prediction of the set of antigen residues that are recognized by the antibody is however the most challenging part. Several methods have been reported (Ansari and Raghava, 2010; Jespersen et al., 2017; Krawczyk et al., 2014; Kringelum et al., 2012; Liang et al., 2010; Qi et al., 2014; Rubinstein et al., 2009; Sela-Culang et al., 2015), but existing epitope prediction systems still do

not provide reliable results, limiting their applicability in molecular docking (Ponomarenko and Bourne, 2007). In this context, docking approaches could present a valuable alternative to the available epitope prediction methods provided that near-native solutions can be generated and recognized. Additionally, docking approaches can elucidate the relationship between antibody and antigen residues facilitating the rational design and engineering of immunoglobulins.

3. Chapter III

3.1 Modelling of antibody-antigen complexes

In this chapter I present an assessment of the performance of ClusPro (Comeau et al., 2004; Kozakov et al., 2017), HADDOCK (De Vries et al., 2010), LightDock (Jiménez-García et al., 2018) and ZDOCK (Pierce et al., 2011) in predicting antibody-antigen structures. All those software allow to use in various ways a-priori knowledge, e.g. the hyper variable loops, into the modelling process to drive or limit the sampling and/or score the docking models.

ClusPro and ZDOCK are based on the Fast Fourier Transform (FTT), they treat the molecules as rigid systems and thus they are unable to account for conformational changes upon binding. On the other hand, LightDock and HADDOCK allow for flexibility during the simulation, the former by using normal modes, the latter by allowing motions of both side chain and backbone during a refinement stage.

3.2 Dataset

The dataset used in this work includes 16 complexes (see Table 1), all with available unbound structures, which represent the new antibody-antigen entries of the protein-protein benchmark version 5.0 (Vreven et al., 2015). These were selected because none were used for training/scoring optimization of any of the docking software considered in this work. Only the antibody variable domain was used for the docking. Antibodies and antigens were each randomly translated and rotated in order to avoid any bias related to the starting orientation (this is required since the structures in the docking benchmark are pre-oriented onto their reference bound complex). The structures were renumbered using a consecutive numbering as not all the software used in this work are able to deal with the insertion format of the Chothia scheme.

Table 1: Summary of the structures used in this work. i-RMSD(Å) indicates the interface root mean square deviation of the Ca atoms of the interface residues calculated after finding the best superposition of bound and unbound interfaces. Complexes are grouped into two classes: Rigid and Medium according to the i-RMSD and the faction of non-native residue contacts (not reported).

Complex	PDB ID 1	Protein 1	PDB ID 2	Protein 2	i-RMSD (Å)
Rigid					
2VXT	2VXU	Murine reference antibody 125-2H FAB	1J0S	Interleukin-18	1.33
2W9E	2W9D	ICSM 18 FAB fragment	1QM1	Prion protein fragment	1.13
3EOA	3EO9	Efalizumab FAB fragment	3F74	Integrin alpha-L I domain	0.39
3HMX	3HMW	Ustekinumab FAB	1F45	Interleukin-12	0.73
3MXW	3MXV	Anti-Shh 5E1 chimera FAB fragment	3M1N	Sonic Hedgehog N-terminal domain	0.48
3RVW	3RVT	4C1 FAB	3F5V	DER P 1 allergen	0.5
4DN4	4DN3	CNTO888 FAB	1DOL	MCP-1	0.81
4FQI	4FQH	CR9114 FAB	2FK0	H5N1 influenza virus hemagglutinin	1.08
4G6J	4G5Z	Canakinumab antibody fragment	4I1B	Interleukin-1 beta	0.61
4G6M	4G6K	Gevokizumab antibody fragment	4I1B	Interleukin-1 beta	0.49
4GXU	4GXV	1F1 antibody	1RUZ	1918 H1 Hemagglutinin	0.78
Medium					
3EO1	3EO0	GC-1008 FAB fragment	1TGJ	Transforming Growth Factor-Beta 3	1.37
3G6D	3G6A	CNTO607 FAB	1IK0	Interleukin-13	1.86
3HI6	3HI5	AL-57 FAB fragment	1MJN	Integrin alpha-L I domain	1.65
3L5W	3L7E	C836 FAB	1IK0	Interleukin-13	0.48
3V6Z	3V6F	FAB E6	3KXS	Capsid protein assembly domain	1.83

3.3 Docking scenarios

Antibody-antigen docking was performed by providing specific options for the antibody antigen-modelling. Indeed, while for the antibody the HV loops provide a valuable proxy of the binding region, as already mentioned, the definition of the antigen epitope is challenging. Therefore, I elaborated three different scenarios in order to mimic different levels of available information about the binding interface on the antigen (see Figure 3):

1. The first scenario (HV – Surf) includes information about the antibody HV loops, defined according to the Chothia numbering scheme (Al-Lazikani et al., 1997) but no information about the epitope. For HADDOCK this was complemented by all solvent-exposed antigen residues defined by selecting those with a relative accessible surface area (RSA) ≥ 40% as calculated with NACCESS (Hubbard SJ, 1993).

2. In the second scenario (HV – Epi 9) a vague definition of the epitope is provided based on all residues having any atom within 9Å from any atom of the antibody in the reference structure.

3. The third scenario (Real interface) represents the ideal case where both interfaces are well characterized. All interface residues, selected using a distance cutoff of 4.5Å between any antibody and antigen atom, were given to the docking software.

This information was used differently in the various docking software depending on their ability to deal with it. In short, HADDOCK follows a data-driven sampling strategy in which the information is encoded into ambiguous restraints to drive the

docking; LightDock uses the information both to limit the sampling to specific regions and in scoring, while ClusPro and ZDOCK include this information into their scoring functions in order to select the correct models.

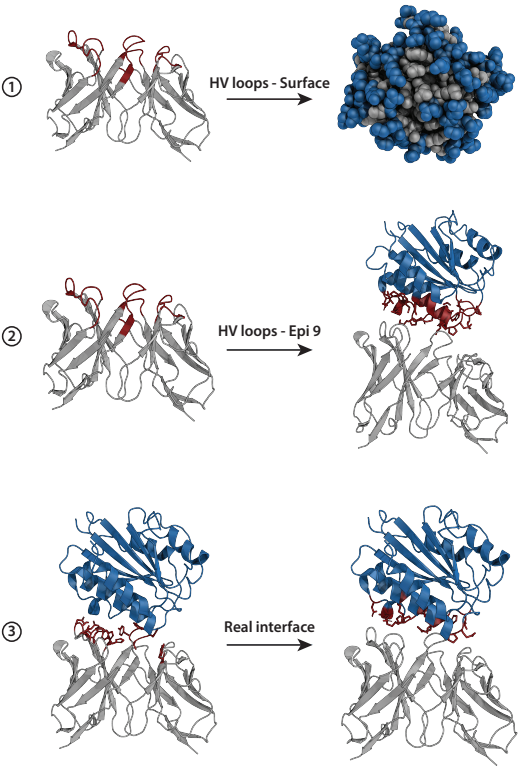

Figure 3: Summary of the three docking scenarios used in this work. The first case represents the situation in which no previous information about the epitope is known so the docking is performed exploring the whole surface of the antigen while for the antibody the HV loops are provided. In the second scenario the antibody HV loops and a loose epitope definition corresponding to the antigen residues within 9Å from the antibody are used to drive the docking. Finally, in the third scenario the real interfaces of both the antigen and the antibody (defined at 4.5Å distance) are used.

3.4 Docking settings

Four docking methods were compared in this work: ClusPro (Kozakov et al., 2017), HADDOCK (De Vries et al., 2010), LightDock (Jiménez-García et al., 2019) and ZDOCK (Pierce et al., 2011).

The ClusPro web server (https://cluspro.org) was used in the Antibody Mode (Brenke et al., 2012) using default settings. Information was provided in the form of attractive residues.

ZDOCK predictions were obtained using a local installation of version 3.0.2. The sampling was set to 2000 models. ZDOCK allows the user to assign a highly unfavourable contact energy to the residues which are known not to be involved in the binding. Accordingly, all residues not included in the defined interfaces were blocked.

For LightDock I used release 0.5.6 of the software (Jiménez-García et al., 2019), which provides a mechanism for including residue restraints. At the receptor level, the surface swarms used at the start of the simulations are filtered according to the Euclidean distance of the restraints on the provided receptor residues. Only the ten closest swarms for each receptor residue restraint are kept. An additional energy term is added to the scoring function used in this work, DFIRE (Zhou and Zhou, 2009), that accounts for the percentage of satisfied restraints. The predictions are filtered with a minimum 40% cutoff of satisfied restraints, at both receptor and ligand levels. For the remaining parameters default settings were used: Anisotropic Network Model (ANM) enabled (10 first non-trivial modes for both receptor and ligand), 400 swarms before filtering by restraints, 200 glowworms per swarm and 100 simulation steps.

Finally, HADDOCK version 2.2 was used with default settings

except that the rigid-body (it0) sampling was increased to 5000 models for the HV – Epi9 and the Real interface runs and to 10000 for the HV – Surf scenario (default is 1000). The flexible (it1) and water refinement sampling were set to 400 models for all scenarios (default is 200) (Dominguez et al., 2003). This corresponds to an increased sampling compared to the default settings. In general, the least information is available to drive the docking in HADDOCK the larger the sampling should be. The docking was performed using the web server version of HADDOCK (https://haddock.science.uu.nl) (Van Zundert et al., 2016). In the case of the Real interface scenario, the random removal of restraints (by default 50% of restraints are randomly discarded for each docking trial) was turned off. The information about the binding interface was encoded in the form of active and passive residues. The antibody HV loops and paratope residues were provided as active, while, for the antigen, the surface and the epitope residues, selected using a 9Å cutoff, were defined as passive. For the third, ideal scenario, the true interface epitope residues selected at 4.5Å distance cutoff were classified as active.

The distinction between active and passive means that an active residue not at the interface (defined as the union of active and passive residues of the partner molecule) will result in an energy penalty while this is not the case for passive residues.

In all the methods, the antibody was treated as the receptor partner and the antigen as ligand.

3.5 Docking performance

I analysed the performance of the four docking methods in predicting antibody-antigen complexes in terms of success rate calculated as the percentage of cases in which at least one acceptable, medium or high-quality model is found in the top N ranked solutions. The model quality was defined according to the CAPRI criteria (Méndez et al., 2003). Those are based on three parameters; briefly:

1. Interface root mean square deviation (i-RMSD): calculated on the backbone atoms of all interface residues of the native complex defined using a 10Å cutoff.

2. Ligand root mean square deviation (L-RMSD): calculated by superimposing on the backbone atoms of the antibody and calculating the RMSD of the antigen backbone atoms.

3. Fraction of native contacts (F_{nat}): number of native contacts in a docking model divided by the total number of contacts in the reference structure. These are defined using a 5Å cutoff.

The ranges used to define the classes are shown in Table 2.

Table 2: Classification of docking models in the classes: ***, **, * according to F_{nat}, and either L-RMSD or i-RMSD measures.

Class	F_{nat}	L-RMSD[Å]	i-RMSD[Å]
High (***)	> 0.5	≤ 1.0	or ≤ 1.0
Medium (**)	> 0.3	≤ 5.0	or ≤ 2.0
Acceptable (*)	> 0.1	≤ 10.0	or ≤ 4.0

The success rate for top 1, 5, 10, 20, 50 and 100 is shown in Figure 4 for each docking method and scenario as described in the paragraph 3.3. The first panel refers to the HV – Surf scenario, the second to the HV – Epi 9 one and finally the third shows the success rate obtained using the real interface information in the docking. The latter represents the gold standard achievable by each docking approach, i.e. the best accuracy that can be reached for this dataset given a perfect interface definition (but no specific contacts) and starting from the unbound structures of the components.

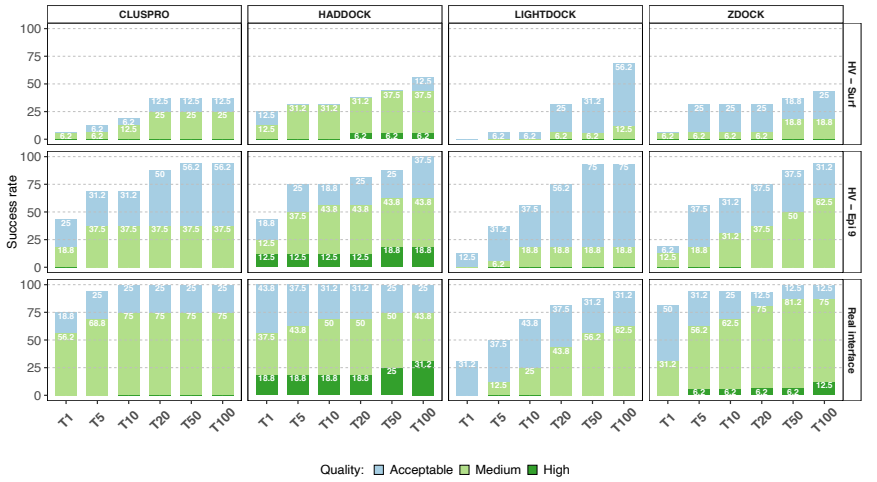

Figure 4: ClusPro, HADDOCK, LightDock and ZDOCK success rate for the three scenarios described in this work as a function of the top 1, 5, 10, 20, 50 and 100 ranked models. The top row (HV - Surf) shows the success rate using the antibody HV loops and the entire antigen surface as restraints. The second represents the success rate achieved by driving the docking with the antibody HV loops and a loose epitope definition using a 9Å cutoff. The third panel shows the docking results using the true interfaces (defined at 4.5Å). The colour coding indicates the quality of the models according to CAPRI criteria.

In the absence of any kind of information about the epitope (HV loops – surface – top row in Figure 4) the overall performance is rather low for all methods. HADDOCK reaches a success rate of 25% in the top 1 which is higher than ClusPro (6.2%), ZDOCK (6.2%)

and LightDock (0%). Note that considering the limited size of the benchmark, a difference of 6.2% only corresponds to one more successfully predicted complex. The differences are smaller for the top 10 (the typical number of models evaluated in CAPRI), with HADDOCK and ZDOCK leading with 31.2%, followed by ClusPro with 18.7%. Considering the top 100, in this scenario LightDock outperforms the other methods with a success rate of 68.7%. This is linked to the fact that LightDock is based on a very effective sampling strategy, but the scoring function used is not accurate for this type of complexes.

By providing a low accuracy definition of the epitope region (HV – Epi 9, middle row in Figure 4) the success rate increases significantly. For example, HADDOCK and ClusPro are able to predict correct models respectively for 75% and 68.7% of the cases already in the top 5 (43.8% in the top1 for both), followed by ZDOCK (56.3%) and LightDock (37.4%).

The bottom row in Figure 4 (Real interface) shows the results when both interfaces are perfectly characterized such that the exact residues involved in the binding are used in the docking. In this case HADDOCK ranks acceptable models in the top 1 position for all 16 complexes of the dataset (100% success rate), while ZDOCK, ClusPro and LightDock reach success rates of 81.2%, 75% and 31.2%, respectively.

Overall, Figure 4 shows that HADDOCK is performing best in every scenario. Even in the cases where ClusPro, LightDock and ZDOCK are able to reach comparable results (e.g. top 50 HV – Epi 9 scenario) the quality of the generated models is usually not as good as those produced by HADDOCK. This can be attributed to the different strategies of using information between those software, with HADDOCK directly using restraints during the sampling/refinement stages, and not only for filtering and/or scoring as is the case in the other software.

3.6 HADDOCK performance – Cluster based

Many approaches perform a clustering after docking in order to group together similar models and simplify the analysis. This has been demonstrated to significantly improve the accuracy of the scoring. The most widely used parameter to measure similarities among different structures is the positional RMSD. The fraction of common contacts (FCC) has been introduced as a fast and valuable alternative to classical RMSD-based methods (Rodrigues et al., 2012). FCC clustering is used by default in HADDOCK (with 4 as minimum cluster size) to cluster the docking models using a default cutoff of 0.6. This has been optimized on classical protein-protein systems. Taking into account the result of the cluster analysis it is possible to express the success rate as the percentage of cases in which there is at least one acceptable, medium or high-quality model in the top 4 cluster members of the top 1, 2, 3, 4 and 5 clusters. In this work clusters were ranked by the average HADDOCK score of their top 4 models (the default scoring scheme of the HADDOCK server (De Vries et al., 2010)). The cluster-based success rate of HADDOCK is shown in Figure 5 for the three different scenarios.

Comparing Figure 4 and Figure 5, and in particular the success rate for top 1 and top 5, one can clearly see how cluster-based scoring increases the success rate of HADDOCK when information about the epitope is provided, but reduces it when no information on the antigen is available and the entire antigen surface is used to drive the docking. This is due to the fact that the sampling around the entire surface of the antigen leads to the generation of many possible different conformations. This results in many local minima of the energy landscape, which the HADDOCK scoring function is not able to distinguish properly. Also, a slightly lower number of models do fall into clusters in this case as illustrated by the clustering coverage calculated as the fraction of clustered models with average values of

0.83±0.10, 0.93±0.04 and 0.99±0.003 respectively for the HV – Surf, HV – Epi 9 and the Real interface scenarios. However, even with a rather loose definition of the epitope (HV – Epi 9 scenario), clustering leads to an improvement in scoring performance from 43.8% to 56.3% for top 1. These results indicate that different scoring strategies should ideally be followed depending on the availability of epitope information or not.

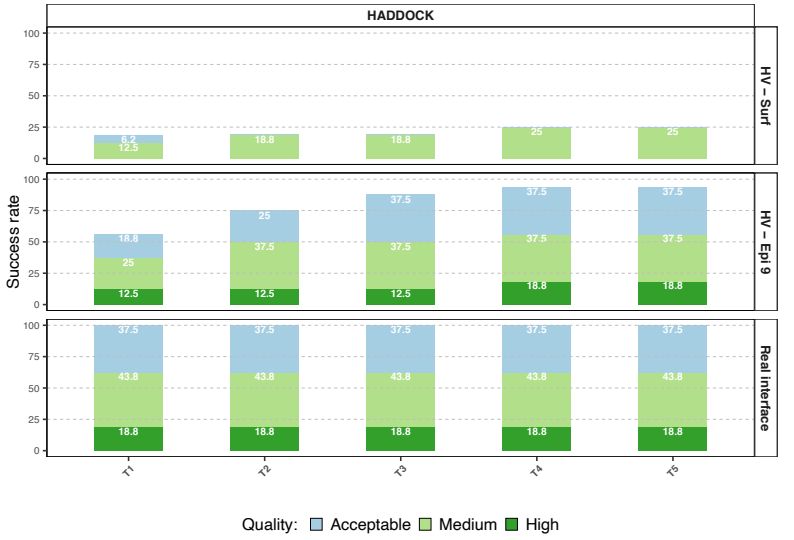

Figure 5: HADDOCK cluster-based success rate for the three docking scenarios as a function of the top 1,2,3,4 and 5 ranked clusters. The colour coding indicates the quality of the models according to CAPRI criteria.

3.7 Sampling performance

Docking involves two different steps, the sampling for the generation of thousands of models, and the scoring to select the best (near-native) models according to a specific scoring function. Most software include the information about the binding interface at the scoring stage, while HADDOCK is the only system which uses this information to drive the sampling (the information is encoded into an additional energy term that generates forces to drive the energy minimization and molecular dynamics steps). The effect of those different strategies can be noticed by calculating the number of acceptable, medium or high-quality models generated out of the total number of produced models. This number is summarized for each software in Figure 6.

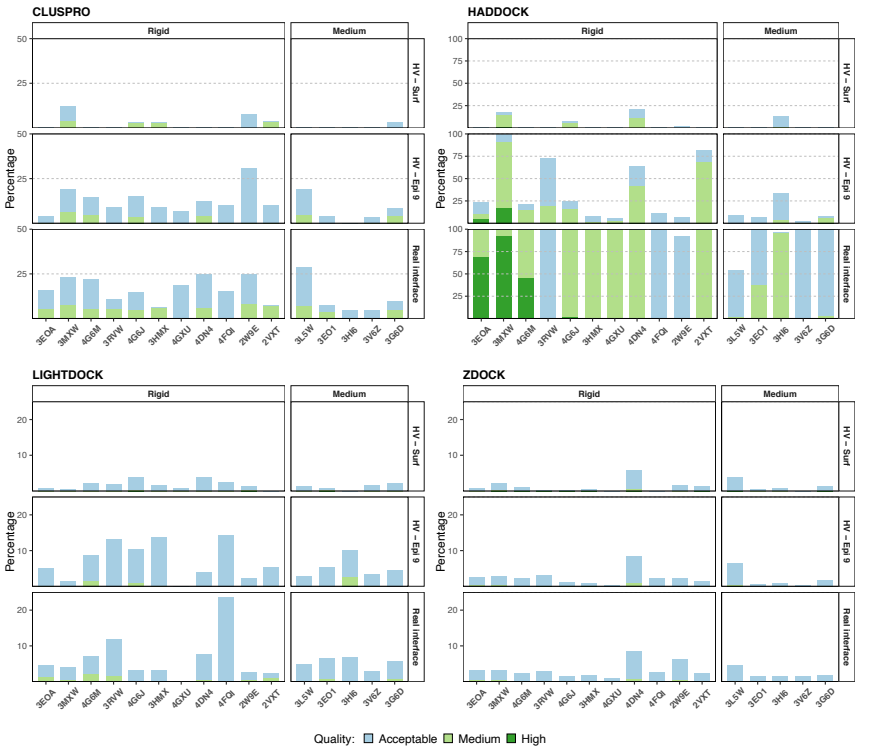

Figure 6: Percentages of acceptable, medium and high-quality models generated by each software per complex and for each scenario. Complexes are split into rigid and medium categories according to the Docking Benchmark5 definition which is based on the size of the conformational change of the unbound molecules upon binding. Note that the Y axis scales are different for each docking method for better readability. The colour coding indicates the quality of the models according to CAPRI criteria.

One can clearly see (Figure 6) how the driving strategy implemented in HADDOCK leads to the generation of a much higher number of good models when information about the interface is provided (HV – Epi 9 and Real Interface scenarios). It has however the danger that no single acceptable model might be generated in case of bad information. The other software, ClusPro, LightDock and ZDOCK

use the interface information only at the scoring level (except for LightDock that filters starting swarms to sample around the provided binding site). These have the advantage that they perform an exhaustive search of the interaction space, but this comes at the cost of a small number of near-native models generated. In that case, the scoring becomes crucial to identify the acceptable models.

3.8 H3 loop modelling performance

The H3 loop of antibodies is the most important loop involved in antigen recognition. Its accurate modelling is still a challenge due to its high structural and sequence variability. Of the four docking software used in this work, two allow for conformational changes during the docking, namely HADDOCK and LightDock. I analysed their capability of inducing the right conformational changes of the loop upon binding with the antigen. For this I superimposed the antibody framework residues of the bound and the unbound structure and calculated the RMSD of H3 ($H3_{unbound}$). Then I repeated the same procedure for each docking model compared to the native complex ($H3_{model}$). For both HADDOCK and LightDock, models produced from the different scenarios were merged and split into correct (i-RMSD ≤ 4Å) and wrong models (i-RMSD > 4Å). Figure 7 shows the distribution of $H3_{model}$ versus $H3_{unbound}$ for correct and wrong models. Values below the diagonal correspond to an improvement of the conformation of the H3 loop. Overall, for HADDOCK (Figure 7A) the flexible refinement tends to increase the RMSD of the H3 loop for complexes that show a low H3 conformational change upon binding but, in contrast, for complexes undergoing larger conformational changes of $H3_{unbound}$, the refinement leads to improvement in the H3 conformation, especially in the scenarios where information about the epitope is provided (HV – Epi 9 and Real interface), with a maximum observed improvement of 1.25Å. In the case of LightDock (Figure 7B) the final selected H3 loop conformation from normal modes remains very close to the unbound form with no remarkable changes in terms of RMSD.

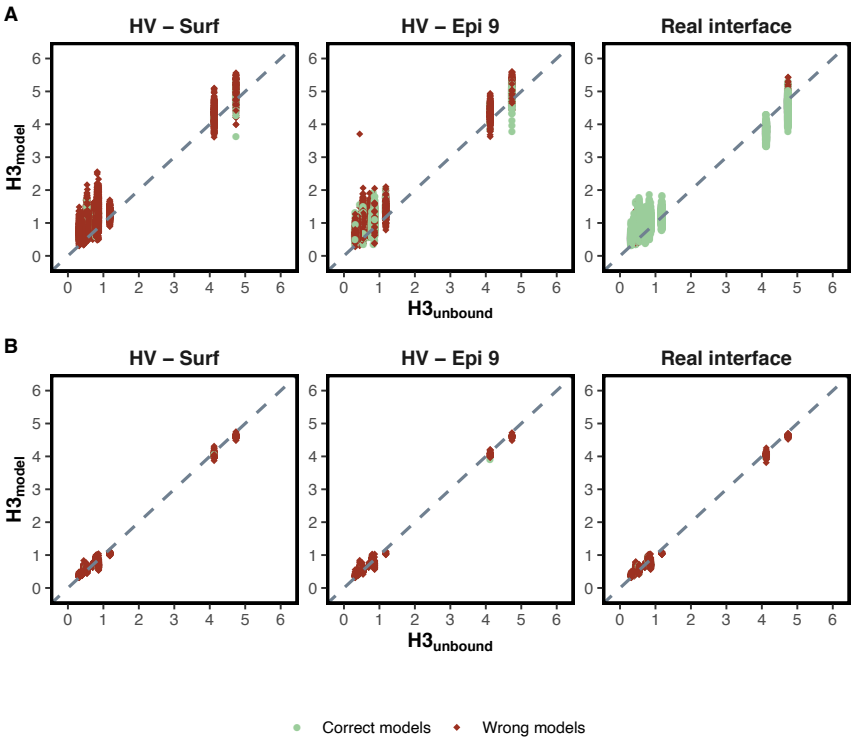

Figure 7: H3 loop RMSD [Å] from the bound conformation for the docked models ($H3_{model}$) versus the starting unbound conformation ($H3_{unbound}$). (**A**) HADDOCK models, (**B**) LightDock models. Correct and wrong models were defined according to their i-RMSD from the reference structure using a 4Å cutoff.

To further investigate the impact of the HADDOCK flexible refinement stage on the H3 loop conformation I analysed the fraction of native contacts (F_{nat}) that H3 makes at the rigid body docking stage ($H3_{it0}$) and after flexible refinement ($H3_{water}$). Figure 8 plots $H3_{water}$ versus $H3_{it0}$ for the three different scenarios discussed in this work, taking into account the quality of the models. In this case all points above the diagonal correspond to an improvement in F_{nat} after flexible refinement. Figure 8 clearly shows that for most cases the flexible refinement improves the number of native contacts made by

H3, with a maximum improvement observed of 0.72. This is more evident for the last two scenarios (HV – Epi 9 and Real interface) indicating that an accurate selection of the native interface is crucial in order to improve the H3 conformation during the simulation.

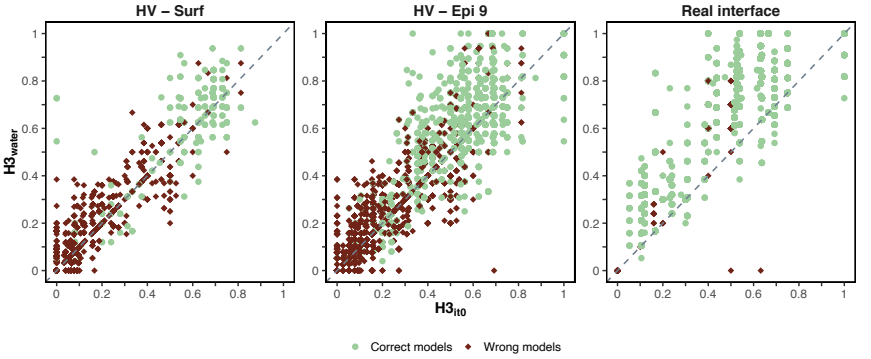

Figure 8: Comparison of the fraction of native contacts (F_{nat}) made by the H3 loop after the rigid body docking stage ($H3_{it0}$) and after flexible/water refinement ($H3_{water}$) of the HADDOCK runs. Values above the diagonal correspond to an improvement in F_{nat}.

4. Chapter IV

4.1 proABC-2: PRediction Of AntiBody Contacts v2 and its application to information-driven docking

In this chapter I present an update of the method implemented in *proABC* (*PRediction Of AntiBody Contacts*) that allows the prediction of which residues of an antibody are forming intermolecular contacts with its cognate antigen, as well as the nature of the contacts distinguishing between hydrogen bonds and hydrophobic interactions.

proABC is based on a random forest algorithm, using the antibody heavy and light chain sequence, the hypervariable loop canonical structures and lengths (Chothia and Lesk, 1987) and the germline family as features (Schatz and Swanson, 2011). The performance of this pipeline has been validated by us (Olimpieri et al., 2013) and others (Peng et al., 2014), demonstrating good accuracy and reliability.

Here I aimed at developing *proABC-2*, an update of the original algorithm, based on a deep learning framework. In particular I trained a *Convolutional Neural Network* (CNN), which has been shown to be successful for similar purposes (Deac et al., 2019; Liberis et al., 2018). proABC-2 has been trained on the aforementioned features. Additionally, I show how the proABC-2 predictions can be used to drive the docking algorithm implemented in HADDOCK (Dominguez et al., 2003), which, as I demonstrated in Chapter III, is performing best in the context of antibody-antigen modelling, giving useful insights into their binding mode. The method is integrated in a freely available web server that predicts paratope residues forming general contacts as well as those involved in hydrogen bonds and hydrophobic interactions.

4.2 Dataset and interaction calculation

The full protein data bank (PDB) was scanned using in-house Hidden Markov Models (HMM) in order to identify all the antibody structures deposited. Immunoglobulins having only one chain (nanobodies), resolution lower than 3Å and not solved with an antigen were excluded. Finally, using cd-hit (Fu et al., 2012) all the antibodies sharing a sequence identity higher than 95% with any other immunoglobulin of the dataset were removed.

I ended up with a dataset of 769 complexes (CNN-dataset) which was used to train the model.

For the docking studies a dataset of 16 complexes (Docking-set), all with available unbound structures, corresponding to the new antibody-antigen entries of the protein-protein benchmark version 5.0 was used (Vreven et al., 2015).

For all the complexes of the CNN-dataset, the non-covalent interactions including intermolecular hydrogen bonds and hydrophobic interactions were calculated. Non-covalent interactions were determined using a distance cut-off of 3.9Å. Hydrogen bonds were calculated by defining donors (D) as any N/O/F/S connected to a hydrogen atom and acceptors (A) as any Nitrogen (N), Oxygen (O), Fluorine (F) or Sulphur (S) within a distance threshold (2.5Å) of that hydrogen and by filtering the matches for D-H-A triplets with a minimum angle of 120 degrees (Baker and Hubbard, 1984). Finally, hydrophobic interactions were computed using a distance cutoff of 4.4Å between any heavy atom of two hydrophobic residues (Bissantz et al., 2010).

General contact were calculated using an in-house R script while H-bond and hydrophobic interactions were determined using *interfacea* (Rodrigues et al., 2019).

4.3 Neural network features

In order to train the CNN a specific set of features was used:

1. *Light and heavy chain sequences* aligned using HMM profiles. In particular for the H3 alignment, according to the previously described method (Morea et al., 1998), I put the insertions in the middle between the conserved residues Cys92 and Gly104 defined according to the Chothia numbering scheme (Chothia and Lesk, 1987). Each sequence position was considered as a variable. To allow the textual information of a sequence to be processed by an algorithm, each residue was converted into numerical values using one-hot encoding, the representation of categorical variables (i.e. a residue) as binary vectors. Here, I used a 20x1 vector consisting of all zeros except at the index of the given residue, which was marked with a 1. Concurrent, a 20x1 vector of only zeros represented a gap. The heavy and light chains were represented by a 297x20 array.

2. *Hypervariable loops canonical structures* calculated according to the key residues found within and outside the loops (Chothia and Lesk, 1987; Morea et al., 1998; Vargas-Madrazo and Paz-García, 2002). One-hot encoding was used.

3. *Length of the hypervariable loops* defined according to the Chothia numbering scheme.

4. *Germline family* and *source organism* determined using *igblastp* (Ye et al., 2013). One-hot encoding was used.

4.4 Convolutional Neural Network (CNN)

The neural network used by proABC-2 consists of three convolutional modules (Conv11, Conv12 and Conv2), a fully connected feed-forward module (Ff1) and an output layer (Figure 9). Conv11 and Conv12 are identical and consists of three parts; a 1D convolutional layer with 32 filters of size 3x1 and a stride of 1, followed by a 1D max pooling layer of size 10x1 and a stride of 3 and finally a dropout layer with a dropout rate of 0.15. Conv2 also consists of three parts; a 1D convolutional layer with 64 filters of size 3x1 and a stride of 1, followed by a 1D average pooling layer of size 6x1 and a stride of 3 and finally a dropout layer with a dropout rate of 0.15. Ff1 consists of a fully connected layer with 512 nodes followed by a dropout layer with a dropout rate of 0.10. The final output layer has for each of the 297 residues 3 nodes, predicting the general interactions, H-bonds and hydrophobic interactions, amounting to 891 nodes. The model was constructed using the python package *Tensorflow* (Abadi et al., 2016).

These modules are combined in the following way. The one-hot encoded heavy and light chains are connected to Conv11 and Conv12 respectively. The extracted features of the heavy and light chains are then concatenated and enter Conv2 for a deeper feature extraction. The final extracted features from Conv2 are then flattened (reduced to one dimension) and concatenated with the additional features (germline, loop lengths and canonical structures) before entering Ff1 and finally into the output nodes. The purpose of Ff1 is to learn each individual residue's role in the paratope based on the extracted features and the additional ones. The architecture is shown in Figure 9.

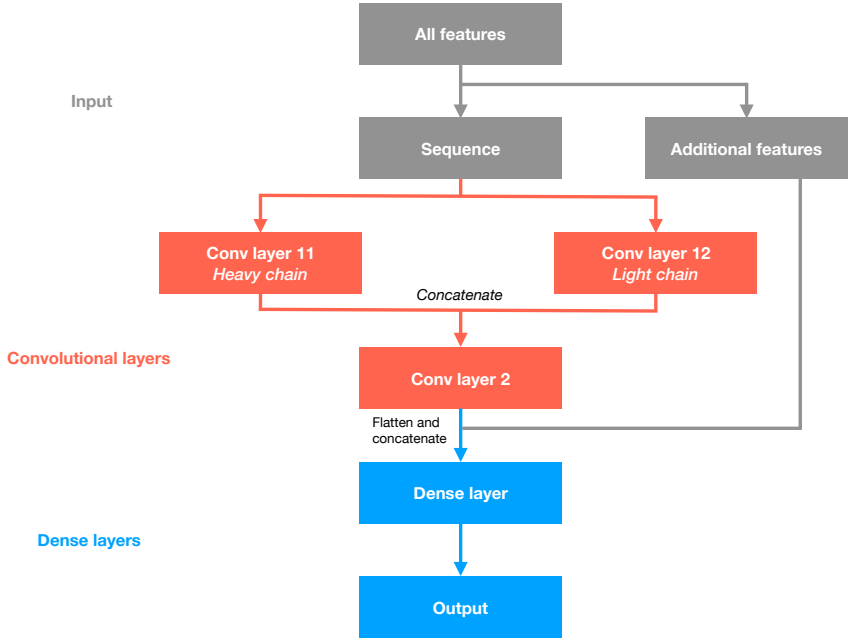

Figure 9: The CNN architecture implemented in the proABC-2 method.

The network was optimized with a focal loss (Lin et al., 2017) and a stochastic gradient descent (SGD) optimizer. The learning rate followed a one-cycle learning rate policy (Smith and Topin, 2017) with a max learning rate of 0.5, a minimum learning rate of 0.1% of the max one and maximum momentum of 0.9. Exponential Linear Units (ELU) were used as activation functions for Conv11, Conv12, Conv2 and Ff1, and sigmoid on the final output. Dropout and early stopping were used throughout training as regularization techniques. All hyperparameters (i.e. nodes, filter sizes, dropout rate etc.) mentioned above were found empirically.

4.5 Model performance

The evaluation of the model was performed using a 10-fold nested cross validation on the full CNN-set (769 complexes). The performance was measured taking into account three different metrics: Area under the Receiver Operating Characteristic (ROC) curve (AUC), Matthew Correlation Coefficient (MCC) and F-score. MCC and F-score were calculated using probability thresholds of 0.40, 0.30 and 0.30 for the general interaction (Pt), hydrophobic interaction (Hy) and hydrogen bond (Hb) predictions, respectively. These cutoffs were selected by averaging the thresholds that, for each fold of the cross validation, gave the best MCC.

AUC, MCC and F-score values corresponding to Pt, Hy and Hb predictions are reported in Table 3. One can clearly see how the performance of the model is higher for general interactions (Pt) (0.96, 0.57 and 0.59 respectively for AUC, MCC and F-score) and it decreases respectively for the hydrophobic interactions (0.95, 0.44 and 0.41) and hydrogen bonds (0.94, 0.33 and 0.27). This is due to the smaller number of Hb and Hy interactions made by the complexes of the CNN-set compared to the Pt ones, which makes their predictions more difficult.

Table 3: Performance of proABC-2 for the three different types of predicted interactions: general contacts (Pt), hydrophobic interactions (Hy) and Hydrogen bonds (Hb).

Interaction	AUC	MCC	F-score
Pt	0.96	0.57	0.59
Hy	0.95	0.44	0.41
Hb	0.94	0.33	0.27

I also compared proABC-2 with *Parapred* (Liberis *et al.*, 2018) which currently is one of the best paratope prediction methods. For a fair comparison proABC-2 was trained on the same dataset used to develop Parapred and the AUC, MCC and F-score were calculated on the same residues used by Parapred to make the predictions (CDRs defined according to the Chothia numbering scheme plus two extra residues) after a 10-fold nested cross validation. For proABC-2 the MCC and F-score were calculated using a threshold of 0.37 (determined as previously explained), while the values from the work of Liberis et al. (Liberis *et al.*, 2018) are reported for Parapred. Results shown in Table 4 indicate that proABC-2 outperforms Parapred in terms of AUC and MCC but shows a lower performance in terms of F-score.

Table 4: Performance comparison between Parapred and proABC-2.

Method	AUC	MCC	F-score
proABC-2	0.91	0.56	0.62
Parapred	0.88	0.55	0.69

4.6 Docking scenarios and settings

To assess the impact of the proABC-2 predictions on HADDOCK's docking performance I designed different scenarios:

1. *Pred Para – Surf*: No previous information about the epitope are provided to HADDOCK. The docking was performed by using the residues predicted to be in contact by proABC-2, defined as active, and the antigen residues having a relative accessible surface areas (RSA) ≥ 40% (calculated with NACCESS (Hubbard SJ, 1993)), provided as passive. Default docking settings were used except for the sampling that was increased to 10000, 400, 400 for it0, it1 and water respectively.

2. *Pred Para – Epi 9*: In this case I provided a loose definition of the epitope region by selecting all the antigen residues within a 9Å distance from the antibody in the reference structure. I provided to HADDOCK the residues predicted by proABC-2 as active and the defined antigen residues as passive. Default docking settings were used except for the sampling that was increased to 5000, 400, 400 for it0, it1 and water respectively.

These settings recapitulate the ones described in Chapter III in order to allow the comparison of the docking results.

The docking models were analysed following the same criteria described in paragraph *3.5 Docking performance* of Chapter III.

4.7 Prediction-driven docking accuracy

I also investigated whether the predictions obtained from proABC-2 can be used to successfully drive antibody-antigen docking using the HADDOCK 2.2 web server (De Vries et al., 2010; Van Zundert et al., 2016).

For unbiased predictions the model was trained on a subset of the CNN-set in which all the structures that were part of the Docking-set and those sharing a 95% sequence identity with them were removed. Only residues making general contacts (Pt) were used to drive the docking. Those were selected using a 0.40 cutoff.

The docking results were compared to those shown in Figure 4 of Chapter III and are reported here in terms of success rate in Figures 10 and 11.

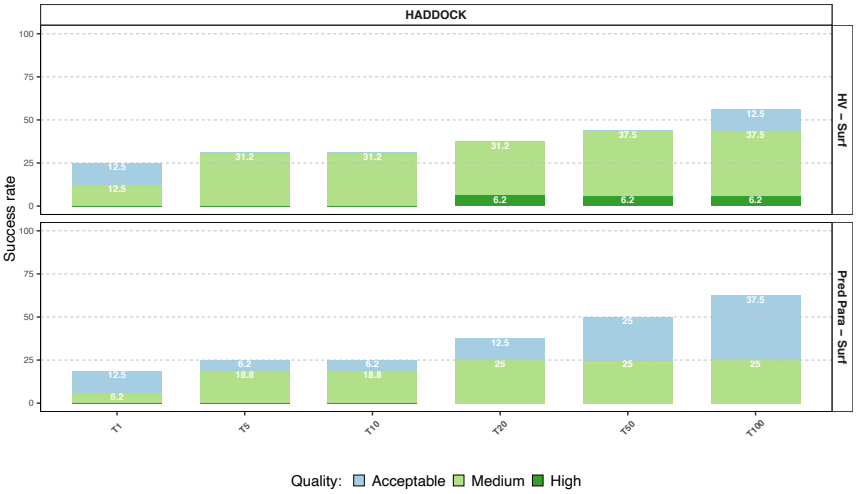

Figure 10: HADDOCK success rate as a function of the top 1, 5, 10, 20, 50 and 100 ranked models. The top row (HV - Surf) shows the success rate using the antibody HV loops and the entire antigen surface as restraints. The second row represents the success rate achieved by driving the docking with the proABC-2 predictions and the full surface of the antigen. The colour coding indicates the quality of the models according to CAPRI criteria.

More specifically, Figure 10 shows the docking results for the first scenario described in this chapter (Pred Para - Surf) in comparison with the first one described in *Chapter III* (HV - Surf). In both cases no information about the epitope is provided to the system. Using the antibody hyper variable loops (HV) to drive the docking led to slightly better results for the top 1,5 and 10 with success rates of 25%, 31.2% and 31.2% respectively, compared to 18.7%, 25% and 25% for the Pred Para - Surf scenario. On the other hand, the use of the predicted antibody residues (Pred Para - Surf) gives better results for the top 50 and 100 with a success rate of 50% and 62.5% respectively (note that 6.2% corresponds to a difference of only one complex). Thus, even if during the sampling HADDOCK is able to generate correct models, the scoring is not able to rank them in the top positions.

As for the quality of the models generated, using the HV loop led to better quality models overall.

The docking performance obtained by providing to the algorithm a loose definition of the epitope following the definition provided in Chapter III and either the antibody hypervariable loops or the proABC-2 predictions are compared in Figure 11. In this scenario driving the docking with the proABC-2 predictions led to a significative improvement of the Top1 success rate which increased from 43.8% (using the HV) to 62.5%. Overall, even if the success rate is comparable between the two scenarios the use of the proABC-2 predictions generally resulted in an improvement of the quality of the generated models, which is mainly reflected in an increase in the number of medium quality ones.

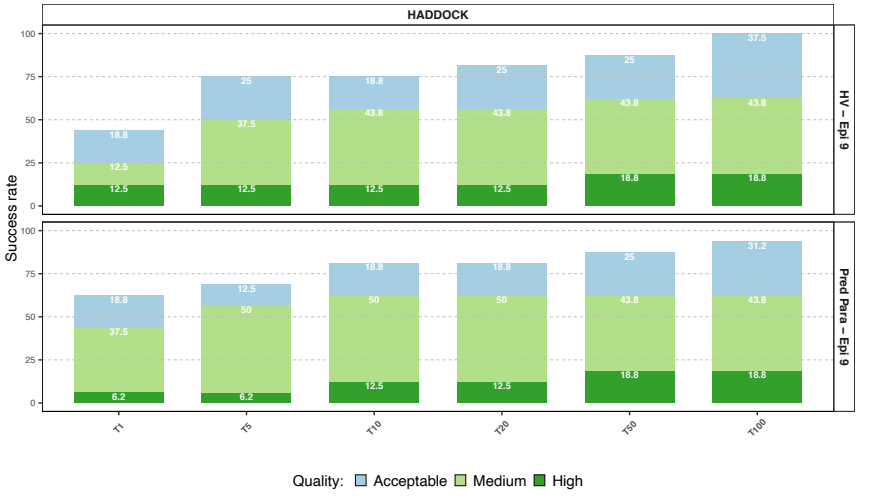

Figure 11: HADDOCK success rate as a function of the top 1, 5, 10, 20, 50 and 100 ranked models. The top row (HV – Epi 9) shows the success rate using the antibody HV loops and a loose definition of the epitope using a 9Å cutoff. The second represents the success rate achieved by driving the docking with the proABC-2 predictions and the same definition for the epitope on the antigen. The colour coding indicates the quality of the models according to CAPRI criteria.

4.8 Web server

proABC-2 is freely available as a web server at https://bianca.science.uu.nl/proabc2 (still under development). It only requires the user to provide the sequences of the heavy and light chains of an antibody. The input is subsequently processed in order to calculate all of the sequence-derived features (germline, canonical structures and length of the HV loops) and these are passed to the CNN to make the predictions. The computations only take a few seconds. The results page reports a barplot (see Figure 12) showing, for each antibody residue, the probability of making a general interaction (pt), H-bond (hb) and a hydrophobic contact (hy). The final output of the web server consists of two files (one for the heavy and one for the light chain) in which for each antibody residue are indicated the probability of making a general interaction, a H-bond and a hydrophobic interaction.

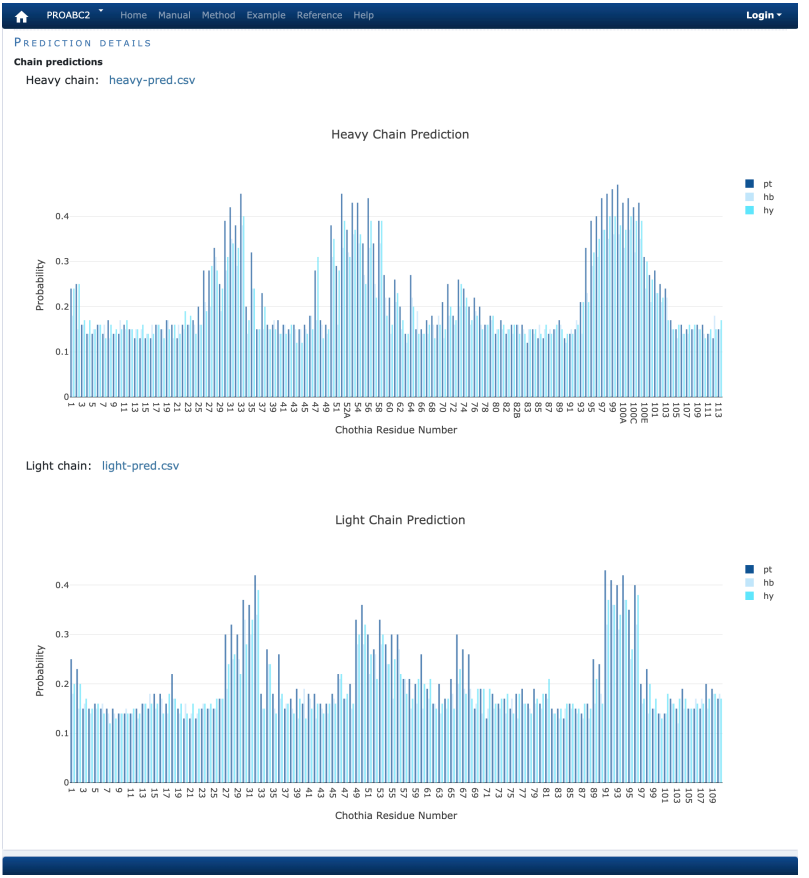

Figure 12: Output page of the *proABC-2* web server. It shows the interaction probability of the antibody residues belonging to the heavy and light chain.

Conclusions and future perspectives

Antibodies continue to dominate the field of biotherapeutics with an increasing number of new clinical approvals each year. Current approaches to bring these molecules to the market have remained experimentally focused, with animal immunization and surface display technologies accounting for the majority of the antibodies developed to date.

The increasing amount of antibody-specific data in the public domain is catalysing the maturation of computational methods for antibody design, resulting in a growing uptake as part of standard pharmaceutical discovery processes. The switching from a purely experiment-based approach to the application of computational techniques could provide more cost-efficient and faster ways for rational design and engineering of immunoglobulins for therapeutic purposes. The final goal is not to completely bypass the experimental pipelines but rather offer a reliable framework which could act as a starting point for further experimental developments.

With this work I demonstrated how the use of molecular docking coupled to machine learning techniques can be beneficial for the identification and the study of the key residues involved in the antigen recognition process.

More specifically, one of the main benefits of this work is to offer researchers a clear overview of the state of the art of antibody-antigen structure prediction (for the various software considered) and of the various strategies that can be followed depending on the available information. Provided that at least a vague definition of the epitope can be obtained, reasonably accurate models can be generated, with HADDOCK performing best among the four software compared.

I also developed and described a machine learning-based method able to predict the specific residues that will make an interaction with

the antigen and I characterised the impact of those predictions on the docking accuracy using HADDOCK. My results show that if a loose definition of the epitope region is provided, the use of the proABC-2 predictions to guide the docking leads to an improvement of both the success rate and the quality of the HADDOCK models compared to the use of the HV loops.

Altogether, my analysis indicates that there is still plenty of opportunities for improvements, especially in modelling conformational changes, with the H3 loop as particular challenge, but also in scoring considering that all software achieve a rather good performance in the top 100, but this significantly drops in most cases when only the top 10 or less are considered.

In the context of scoring, machine learning and in particular deep learning offer a great potential but they strongly rely on the availability of enough reliable data. An accurate benchmarking is therefore of paramount importance for the application of such techniques. In addition, development of scoring schemes specifically tailored for the immune complexes might also increase the accuracy of the docking methods.

Moreover, taking into account the low accuracy of the docking methods when no information about the antigen epitope is available, further progresses are required in the development of antibody-specific epitope prediction methods. This might be achieved in the future by leveraging on the large amount of Next Generation Sequencing (NGS) data that are nowadays becoming available.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., et al. (2016). TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems.

Abhinandan, K.R., and Martin, A.C.R. (2008). Analysis and improvements to Kabat and structurally correct numbering of antibody variable domains. Mol. Immunol. *45*, 3832–3839.

Al-Lazikani, B., Lesk, A.M., and Chothia, C. (1997). Standard conformations for the canonical structures of immunoglobulins. J. Mol. Biol. *273*, 927–948.

Ansari, H.R., and Raghava, G.P. (2010). Identification of conformational B-cell Epitopes in an antigen from its primary sequence. Immunome Res. *6*, 6.

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

Bassing, C.H., Swat, W., and Alt, F.W. (2002). The Mechanism and Regulation of Chromosomal V(D)J Recombination. Cell *109*, S45–S55.

Bissantz, C., Kuhn, B., and Stahl, M. (2010). A Medicinal Chemist's Guide to Molecular Interactions. J. Med. Chem. *53*, 5061–5084.

Brenke, R., Hall, D.R., Chuang, G.Y., Comeau, S.R., Bohnuud, T., Beglov, D., Schueler-Furman, O., Vajda, S., and Kozakov, D. (2012). Application of asymmetric statistical potentials to antibody-protein docking. Bioinformatics *28*, 2608–2614.

Briney, B., Inderbitzin, A., Joyce, C., and Burton, D.R. (2019). Commonality despite exceptional diversity in the baseline human

antibody repertoire. Nature *566*, 393–397.

Bujotzek, A., Dunbar, J., Lipsmeier, F., Schäfer, W., Antes, I., Deane, C.M., and Georges, G. (2015). Prediction of VH-VL domain orientation for antibody variable domain modeling. Proteins Struct. Funct. Bioinforma. *83*, 681–695.

Chen, R., and Weng, Z. (2002). Docking unbound proteins using shape complementarity, desolvation, and electrostatics. Proteins Struct. Funct. Genet. *47*, 281–294.

Chothia, C., and Lesk, A.M. (1987). Canonical structures for the hypervariable regions of immunoglobulins. J. Mol. Biol. *196*, 901–917.

Chothia, C., Lesk, A.M., Tramontano, A., Levitt, M., Smith-Gill, S.J., Air, G., Sheriff, S., Padlan, E.A., Davies, D., Tulip, W.R., et al. (1989). Conformations of immunoglobulin hypervariable regions. Nature *342*, 877–883.

Comeau, S.R., Gatchell, D.W., Vajda, S., and Camacho, C.J. (2004). ClusPro: a fully automated algorithm for protein-protein docking. Nucleic Acids Res. *32*, W96–W99.

Conte, L. Lo, Chothia, C., and Janin, J. (1999). The atomic structure of protein-protein recognition sites. J. Mol. Biol. *285*, 2177–2198.

Deac, A., VeliČković, P., and Sormanni, P. (2019). Attentive Cross-Modal Paratope Prediction. J. Comput. Biol. *26*, 536–545.

Dominguez, C., Boelens, R., and Bonvin, A.M.J.J. (2003). HADDOCK: A Protein−Protein Docking Approach Based on Biochemical or Biophysical Information. J. Am. Chem. Soc. *125*, 1731–1737.

Dunbar, J., and Deane, C.M. (2016). ANARCI: Antigen receptor

numbering and receptor classification. Bioinformatics *32*, 298–300.

Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). CD-HIT: Accelerated for clustering the next-generation sequencing data. Bioinformatics *28*, 3150–3152.

Glanville, J., Zhai, W., Berka, J., Telman, D., Huerta, G., Mehta, G.R., Ni, I., Mei, L., Sundar, P.D., Day, G.M.R., et al. (2009). Precise determination of the diversity of a combinatorial antibody library gives insight into the human immunoglobulin repertoire. Proc. Natl. Acad. Sci. *106*, 20216–20221.

Hesslein, D.G.T., and Schatz, D.G. (2001). Factors and Forces Controlling V(D)J Recombination. In Advances in Immunology, pp. 169–232.

Honegger, A., and Plückthun, A. (2001). Yet another numbering scheme for immunoglobulin variable domains: An automatic modeling and analysis tool. J. Mol. Biol. *309*, 657–670.

Hubbard SJ, T.J. (1993). NACCESS. Comput. Progr.

Jespersen, M.C., Peters, B., Nielsen, M., and Marcatili, P. (2017). BepiPred-2.0: Improving sequence-based B-cell epitope prediction using conformational epitopes. Nucleic Acids Res. *45*, W24–W29.

Jiménez-García, B., Roel-Touris, J., Romero-Durana, M., Vidal, M., Jiménez-González, D., and Fernández-Recio, J. (2018). LightDock: A new multi-scale approach to protein-protein docking. Bioinformatics *34*, 49–55.

Jiménez-García, B., Vidal, M., and Roel-Touris, J. (2019). brianjimenez/lightdock: Release 0.5.6.

Kaplon, H., and Reichert, J.M. (2019). Antibodies to watch in 2019. MAbs *11*, 219–238.

Katchalski-Katzir, E., Shariv, I., Eisenstein, M., Friesem, A.A., Aflalo, C., and Vakser, I.A. (1992). Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques. Proc. Natl. Acad. Sci. U. S. A. *89*, 2195–2199.

Kotev, M., Soliva, R., and Orozco, M. (2016). Challenges of docking in large, flexible and promiscuous binding sites. Bioorganic Med. Chem. *24*, 4961–4969.

Kozakov, D., Hall, D.R., Xia, B., Porter, K.A., Padhorny, D., Yueh, C., Beglov, D., and Vajda, S. (2017). The ClusPro web server for protein-protein docking. Nat. Protoc. *12*, 255–278.

Krawczyk, K., Baker, T., Shi, J., and Deane, C.M. (2013). Antibody i-Patch prediction of the antibody binding site improves rigid local antibody-antigen docking. Protein Eng. Des. Sel. *26*, 621–629.

Krawczyk, K., Liu, X., Baker, T., Shi, J., and Deane, C.M. (2014). Improving B-cell epitope prediction and its application to global antibody-antigen docking. Bioinformatics *30*, 2288–2294.

Kringelum, J.V., Lundegaard, C., Lund, O., and Nielsen, M. (2012). Reliable B Cell Epitope Predictions: Impacts of Method Development and Improved Benchmarking. PLoS Comput. Biol. *8*, e1002829.

Kunik, V., Ashkenazi, S., and Ofran, Y. (2012). Paratome: an online tool for systematic identification of antigen-binding regions in antibodies based on sequence or structure. Nucleic Acids Res. *40*, W521-4.

Larrick, J.W., Alfenito, M.R., Scott, J.K., Parren, P.W.H.I., Burton, D.R., Bradbury, A.R.M., Lemere, C.A., Messer, A., Huston, J.S., Carter, P.J., et al. (2016). Antibody Engineering & Therapeutics 2016: The Antibody Society's annual meeting, December 11–15,

2016, San Diego, CA. MAbs *8*, 1425–1434.

Lefranc, M.P. (2011). IMGT unique numbering for the variable (V), constant (C), and groove (G) domains of IG, TR, MH, IgSF, and MhSF. Cold Spring Harb. Protoc. *6*, 633–642.

Li, L., Chen, S., Miao, Z., Liu, Y., Liu, X., Xiao, Z.X., and Cao, Y. (2019). AbRSA: A robust tool for antibody numbering. Protein Sci. *28*, 1524–1531.

Liang, S., Zheng, D., Standley, D.M., Yao, B., Zacharias, M., and Zhang, C. (2010). EPSVR and EPMeta: Prediction of antigenic epitopes using support vector regression and multiple server results. BMC Bioinformatics *11*, 381.

Liberis, E., Velickovic, P., Sormanni, P., Vendruscolo, M., and Lio, P. (2018). Parapred: Antibody paratope prediction using convolutional and recurrent neural networks. Bioinformatics *34*, 2944–2950.

Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollár, P. (2017). Focal Loss for Dense Object Detection.

Lippow, S.M., Wittrup, K.D., and Tidor, B. (2007). Computational design of antibody-affinity improvement beyond in vivo maturation. Nat. Biotechnol. *25*, 1171–1176.

MacCallum, R.M., Martin, A.C.R., and Thornton, J.M. (1996). Antibody-antigen interactions: Contact analysis and binding site topography. J. Mol. Biol. *262*, 732–745.

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 Struct. Funct. Genet. *52*, 51–67.

Morea, V., Tramontano, A., Rustici, M., Chothia, C., and Lesk, A.M. (1998). Conformations of the third hypervariable region in the VH domain of immunoglobulins. J. Mol. Biol. *275*, 269–294.

North, B., Lehmann, A., and Dunbrack, R.L. (2011). A new clustering of antibody CDR loop conformations. J. Mol. Biol. *406*, 228–256.

Novotný, J., Bruccoleri, R., Newell, J., Murphy, D., Haber, E., and Karplus, M. (1983). Molecular anatomy of the antibody binding site. J. Biol. Chem. *258*, 14433–14437.

Olimpieri, P.P., Chailyan, A., Tramontano, A., and Marcatili, P. (2013). Prediction of site-specific interactions in antibody-antigen complexes: The proABC method and server. Bioinformatics *29*, 2285–2291.

Pedotti, M., Simonelli, L., Livoti, E., and Varani, L. (2011). Computational docking of antibody-antigen complexes, opportunities and pitfalls illustrated by influenza hemagglutinin. Int. J. Mol. Sci. *12*, 226–251.

Peters, A., and Storb, U. (1996). Somatic Hypermutation of Immunoglobulin Genes Is Linked to Transcription Initiation. Immunity *4*, 57–65.

Pierce, B.G., Hourai, Y., and Weng, Z. (2011). Accelerating protein docking in ZDOCK using an advanced 3D convolution library. PLoS One *6*, e24657.

Ponomarenko, J. V., and Bourne, P.E. (2007). Antibody-protein interactions: benchmark datasets and prediction tools evaluation. BMC Struct. Biol. *7*, 64.

Qi, T., Qiu, T., Zhang, Q., Tang, K., Fan, Y., Qiu, J., Wu, D., Zhang, W., Chen, Y., Gao, J., et al. (2014). SEPPA 2.0—more refined server

to predict spatial epitope considering species of immune host and subcellular localization of protein antigen. Nucleic Acids Res. *42*, W59–W63.

Rodrigues, J.P.G.L.M., and Bonvin, A.M.J.J. (2014). Integrative computational modeling of protein interactions. FEBS J. *281*, 1988–2003.

Rodrigues, J., Valentine, C., and Jimenez, B. (2019). JoaoRodrigues/interfacea: First beta version of the API.

Rodrigues, J.P.G.L.M., Trellet, M., Schmitz, C., Kastritis, P., Karaca, E., Melquiond, A.S.J., and Bonvin, A.M.J.J. (2012). Clustering biomolecular complexes by residue contacts similarity. Proteins Struct. Funct. Bioinforma. *80*, 1810–1817.

Rubinstein, N.D., Mayrose, I., Martz, E., and Pupko, T. (2009). Epitopia: a web-server for predicting B-cell epitopes. BMC Bioinformatics *10*, 287.

Schatz, D.G., and Swanson, P.C. (2011). V(D)J Recombination: Mechanisms of Initiation. Annu. Rev. Genet. *45*, 167–202.

Schneidman-Duhovny, D., Inbar, Y., Nussinov, R., and Wolfson, H.J. (2005). PatchDock and SymmDock: servers for rigid and symmetric docking. Nucleic Acids Res. *33*, W363–W367.

Schroeder, H.W.J., and Cavacini, L. (2010). Structure and Function of Immunoglobulins. J. Allergy Clin. Immunol. *125*, S41–S52.

Sela-Culang, I., Kunik, V., and Ofran, Y. (2013). The Structural Basis of Antibody-Antigen Recognition. Front. Immunol. *4*, 302.

Sela-Culang, I., Ashkenazi, S., Peters, B., and Ofran, Y. (2015). PEASE: Predicting B-cell epitopes utilizing antibody sequence. Bioinformatics *31*, 1313–1315.

Sevy, A.M., Healey, J.F., Deng, W., Spiegel, P.C., Meeks, S.L., and Li, R. (2013). Epitope mapping of inhibitory antibodies targeting the C2 domain of coagulation factor VIII by hydrogen-deuterium exchange mass spectrometry. J. Thromb. Haemost. *11*, 2128–2136.

Shirai, H., Kidera, A., and Nakamura, H. (1996). Structural classification of CDR-H3 in antibodies. FEBS Lett. *399*, 1–8.

Sircar, A., and Gray, J.J. (2010). SnugDock: Paratope Structural Optimization during Antibody-Antigen Docking Compensates for Errors in Antibody Homology Models. PLoS Comput. Biol. *6*, e1000644.

Smith, L.N., and Topin, N. (2017). Super-Convergence: Very Fast Training of Neural Networks Using Large Learning Rates.

Tonegawa, S. (1983). Somatic generation of antibody diversity. Nature *302*, 575–581.

Torchala, M., Moal, I.H., Chaleil, R.A.G., Fernandez-Recio, J., and Bates, P.A. (2013). SwarmDock: A server for flexible protein-protein docking. Bioinformatics *29*, 807–809.

Vajda, S. (2005). Classification of protein complexes based on docking difficulty. In Proteins: Structure, Function and Genetics, pp. 176–180.

Vargas-Madrazo, E., and Paz-García, E. (2002). Modifications to canonical structure sequence patterns: Analysis for L1 and L3. Proteins Struct. Funct. Genet. *47*, 250–254.

Vreven, T., Moal, I.H., Vangone, A., Pierce, B.G., Kastritis, P.L., Torchala, M., Chaleil, R., Jiménez-García, B., Bates, P.A., Fernandez-Recio, J., et al. (2015). Updates to the Integrated Protein-Protein Interaction Benchmarks: Docking Benchmark Version 5 and Affinity Benchmark Version 2. J. Mol. Biol. *427*, 3031–3041.

De Vries, S.J., Van Dijk, M., and Bonvin, A.M.J.J. (2010). The HADDOCK web server for data-driven biomolecular docking. Nat. Protoc. *5*, 883–897.

De Vries, S.J., Schindler, C.E.M., Chauvot De Beauchêne, I., and Zacharias, M. (2015). A web interface for easy flexible protein-protein docking with ATTRACT. Biophys. J. *108*, 462–465.

Weitzner, B.D., Dunbrack, R.L., and Gray, J.J. (2015). The origin of CDR H3 structural diversity. Structure *23*, 302–311.

Te Wu, T., and Kabat, E.A. (1970). An analysis of the sequences of the variable regions of bence jones proteins and myeloma light chains and their implications for antibody complementarity. J. Exp. Med. *132*, 211–250.

Ye, J., Ma, N., Madden, T.L., and Ostell, J.M. (2013). IgBLAST: an immunoglobulin variable domain sequence analysis tool. Nucleic Acids Res. *41*, W34–W40.

Zhou, H., and Zhou, Y. (2009). Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction. Protein Sci. *11*, 2714–2726.

Van Zundert, G.C.P., Rodrigues, J.P.G.L.M., Trellet, M., Schmitz, C., Kastritis, P.L., Karaca, E., Melquiond, A.S.J., Van Dijk, M., De Vries, S.J., and Bonvin, A.M.J.J. (2016). The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. J. Mol. Biol. *428*, 720–725.