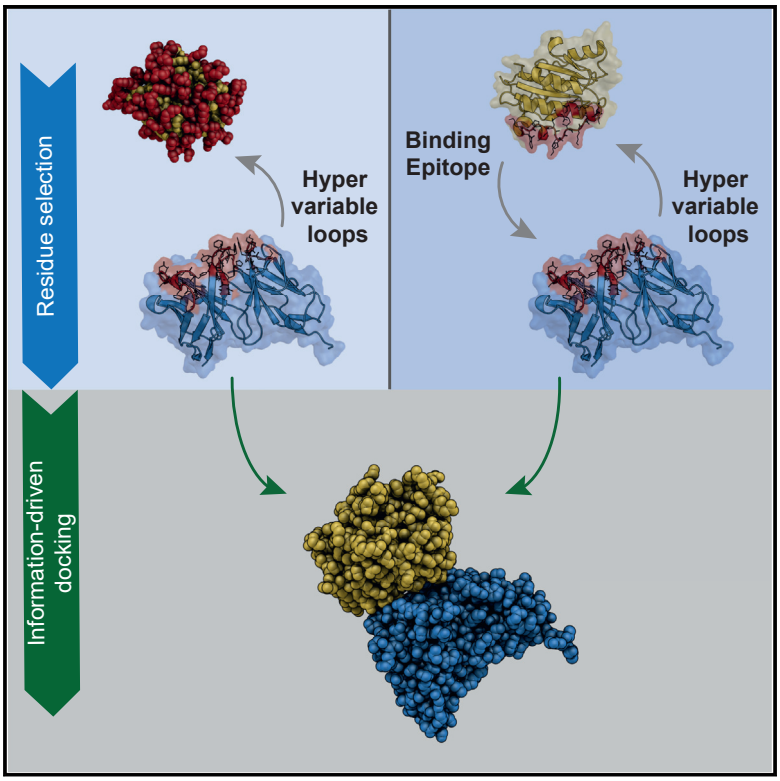
Resource

Structure

Modeling Antibody-Antigen Complexes by Information-Driven Docking

Graphical Abstract



Authors

Francesco Ambrosetti,
Brian Jiménez-García,
Jorge Roel-Touris,
Alexandre M.J.J. Bonvin

Correspondence

a.m.j.j.bonvin@uu.nl

In Brief

Ambrosetti et al. demonstrate that, for the modeling of antibody-antigen complexes, using information about hypervariable loops and, when available, a loose definition of the epitope, improves the docking results. In this context, HADDOCK, which directly uses this information to guide docking, performs better than the other three software applications tested.

Highlights

- Accurate prediction of antibody-antigen structure is still a challenge
- Hypervariable loops of antibodies can be used to bias the modeling process
- Antigen epitope information, even loosely defined, is valuable for docking
- HADDOCK shows better performance and generates higher-accuracy models

Ambrosetti et al., 2020, Structure *28*, 119–129
January 7, 2020
https://doi.org/10.1016/j.str.2019.10.011

CellPress

Structure
Resource

Modeling Antibody-Antigen Complexes by Information-Driven Docking

Francesco Ambrosetti,[1,2] Brian Jiménez-García,[2] Jorge Roel-Touris,[2] and Alexandre M.J.J. Bonvin[2,3,*]
[1]Department of Physics, Sapienza University, Piazzale Aldo Moro 5, 00184 Rome, Italy
[2]Faculty of Science – Chemistry, Computational Structural Biology Group, Bijvoet Center for Biomolecular Research, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands
[3]Lead Contact
*Correspondence: a.m.j.j.bonvin@uu.nl
https://doi.org/10.1016/j.str.2019.10.011

SUMMARY

Antibodies are Y-shaped proteins essential for immune response. Their capability to recognize antigens with high specificity makes them excellent therapeutic targets. Understanding the structural basis of antibody-antigen interactions is therefore crucial for improving our ability to design efficient biological drugs. Computational approaches such as molecular docking are providing a valuable and fast alternative to experimental structural characterization for these complexes. We investigate here how information about complementarity-determining regions and binding epitopes can be used to drive the modeling process, and present a comparative study of four different docking software suites (ClusPro, LightDock, ZDOCK, and HADDOCK) providing specific options for antibody-antigen modeling. Their performance on a dataset of 16 complexes is reported. HADDOCK, which includes information to drive the docking, is shown to perform best in terms of both success rate and quality of the generated models in both the presence and absence of information about the epitope on the antigen.

INTRODUCTION

Antibodies are essential components of the immune response. Their capability to recognize antigens with high specificity along with their modular anatomy, which facilitates their design and engineering, makes them excellent therapeutic targets (Kaplon and Reichert, 2019). For the design of efficient biological drugs (Morea et al., 2000) based on antibodies it is crucial to properly understand the structural basis of antibody-antigen interactions.

Antibodies are Y-shaped proteins usually composed of two pairs of identical polypeptide chains named light and heavy chains. On the basis of their structural and sequence variability, it is possible to identify variable and constant domains, more specifically one variable and one constant domain for the light chain and one variable and three or more constant domains for the heavy one. The variable domain is composed of a very well conserved framework containing six hypervariable loops (HV loops), three from the light chain and three from the heavy chain. These are part of the so-called complementarity-determining regions (CDRs) (Wu, 2004). These regions, and in particular the HV loops, are crucial for antigen recognition and specificity (Novotný et al., 1983). The position of HV loops is known a priori and can be inferred given only the antibody sequence (Al-Lazikani et al., 1997). Despite the majority of the antibody-binding residues being included in the CDRs (Kunik et al., 2012; MacCallum et al., 1996), it has been shown that residues that fall outside these CDRs can also play a crucial role in the antigen recognition process (Narciso et al., 2011). Various experimental methods have been proposed to investigate the role of each residue in the antigen recognition process such as, for example, hydrogen/deuterium (H/D) exchange (Lim et al., 2017), mutagenesis analysis (Fontayne et al., 2007), or the classical structural methods such as nuclear magnetic resonance (NMR) and X-ray crystallography. These provide various levels of information about the key antibody and antigen residues involved in the interaction, but most of them require high cost and effort. Computational approaches able to predict antibody-antigen structures would offer a valuable and fast alternative. 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 identification or prediction of the set of antigen residues that are recognized by the antibody is, however, the most challenging part. Although 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), epitope prediction remains an open issue (Ponomarenko and Bourne, 2007). In this context, docking approaches could present a valuable alternative to the available epitope prediction methods, provided near-native solutions can be generated and recognized.

Different docking algorithms have been developed over the years to predict the three-dimensional (3D) structure of biological complexes starting from the free, unbound structures of the components (Moreira et al., 2010). These approaches rely on the generation of thousands of possible complex conformations or models, which are successively ranked according to a specific scoring function to identify or predict the models that are close to the real conformation (near-native solutions). 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

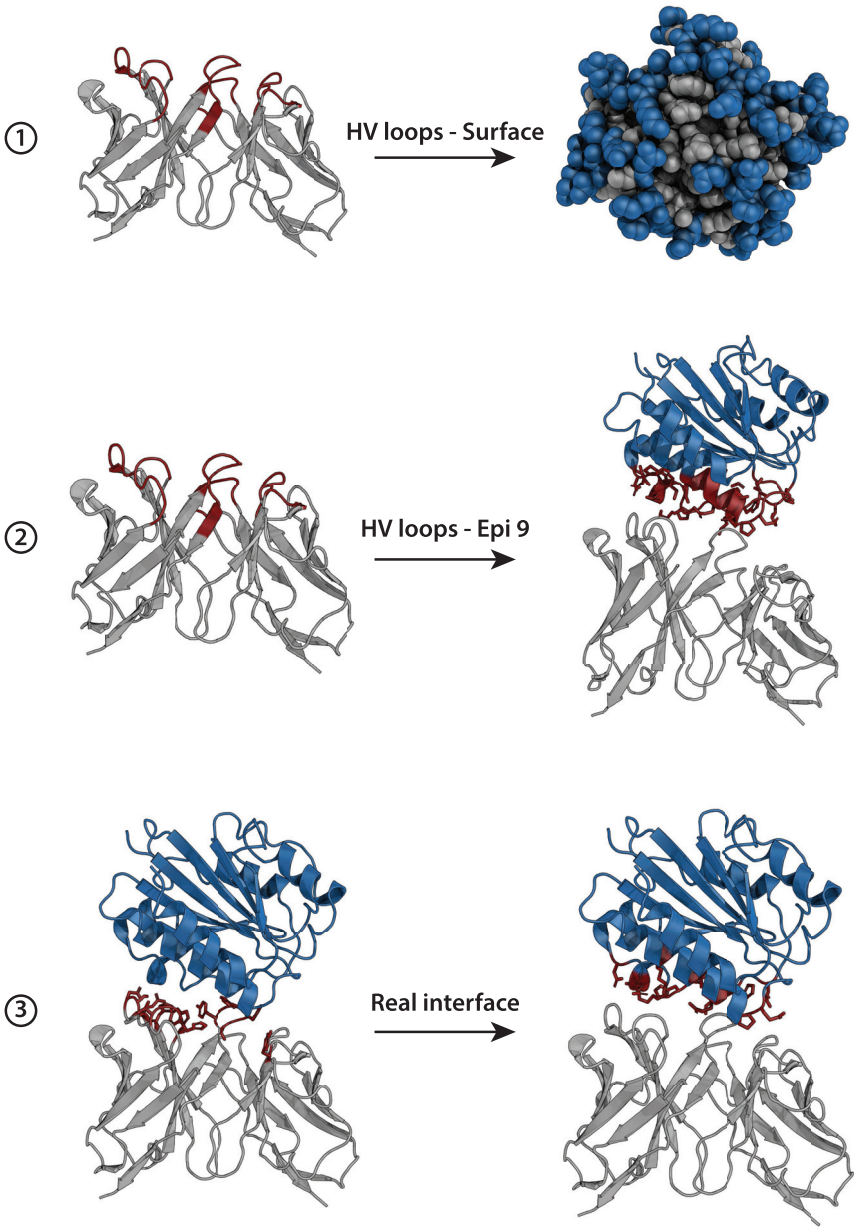


Figure 1. Summary of the Three Docking Scenarios Used in This Work

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 9 Å 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 4.5 Å distance) are used.

Fourier transform 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 so by using normal modes, the latter two by allowing some flexibility along side chains and the backbone during a refinement stage.

Docking methods can be classified into two categories according to the sampling strategy applied during the simulation (excluding the template-based docking approaches when a homologous interface can be identified, a less relevant approach for antibody-antigen complexes). The first class includes algorithms that do not take into account any previous information about the putative binding interfaces and perform an exhaustive search of the interaction space, the so-called *ab initio* docking methods (Ritchie, 2008). Nonetheless, in many cases experimental information can be used during the scoring step to select near-native models. The second class consists of docking approaches where sampling is driven by experimental data, coming, for example, from mutagenesis, mass spectrometry (MS) (H/D exchange and/or crosslinking experiments), NMR analysis, or predicted information about the binding interface. These fall under the so-called information-driven or integrative modeling approaches (Rodrigues and Bonvin, 2014). The performances of docking methods is continuously assessed by the Critical Assessment of Predicted Interactions (CAPRI) experiment (Janin et al., 2003; Méndez et al., 2003), stimulating researchers' efforts toward the development of more accurate docking and scoring algorithms.

All docking approaches rely on the availability of 3D structures or models of the components. Since antibodies have a very conserved framework and the conformation of five out of six loops can be quite reliably predicted (Chothia et al., 1989), modeling methods specific for antibodies are able to generate reasonably accurate structures (Leem et al., 2016; Lepore et al., 2017; Weitzner et al., 2017; Yamashita et al., 2014). The main problems in this field are related to predicting the conformation of the H3 loop, which remains challenging due to its high structural and length variability (Shirai et al., 1996; Weitzner et al., 2015). Methods specifically tailored to predict its conformation have been developed (Choi and Deane, 2010; Messih et al., 2014) to improve the accuracy of the antibody modeling systems.

Despite great progress in predicting protein-protein complexes, docking of antibody-antigen complexes is still challenging (Pedotti et al., 2011; Ponomarenko and Bourne, 2007; Vajda, 2005) due to the inherent properties of their interfaces

Table 1. Classification of Docking Models in the Classes *, **, and * according to F_{nat} and Either Ligand RMSD or i-RMSD Measures**

Class	F_{nat}	Ligand 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

(Lo Conte et al., 1999; Sela-Culang et al., 2013). In this work, we present an assessment of the performance of ClusPro, HADDOCK, LightDock, and ZDOCK in predicting antibody-antigen structures. All of these software packages allow the use in various ways of a priori knowledge, e.g., the hypervariable loops, into the modeling process to drive or limit the sampling and/or score the docking models. For the antigen, we use different levels of information to define the epitope. Using a set of 16 antibody-antigen complexes corresponding to the new entries from the docking benchmark 5 (BM5) (Vreven et al., 2015) for which unbound structures are available, we compare the performance of the various docking software applications following several strategies to include information about the binding regions.

RESULTS

Antibody-antigen docking was performed following three different scenarios in order to mimic different levels of information that can be obtained about the antibody and antigen residues involved in the binding. The first scenario (HV-Surf) includes information about the antibody (HV loops) but no information about the epitope; in the second scenario (HV-Epi 9), a vague definition of the epitope is provided based on all residues within 9 Å from the antibody in the reference structure; finally, the third scenario (Real interface) represents the ideal case whereby both interfaces are well characterized (see Figure 1). Further information about the scenarios can be found in STAR Methods. This information was used differently in the various docking software depending on their ability to deal with it. In short (for details see STAR Methods), HADDOCK follows a data-driven sampling strategy whereby 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 in their scoring functions in order to select the correct models.

Docking Performance: Single Structure

We analyzed 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 (see Table 1) is found in the top N ranked solutions. The success rate for the top 1, 5, 10, 20, 50, and 100 is shown in Figure 2 for each docking method and scenario as described in STAR Methods. The first panel refers to the HV-Surf scenario and the second to the HV-Epi 9, and 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.

In the absence of any kind of information about the epitope (HV loops—surface; top row in Figure 2) 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 complex.

By providing a low-accuracy definition of the epitope region (HV-Epi 9; middle row in Figure 2), the success rate increases significantly. For example, HADDOCK and ClusPro are able to predict correct models for 75% and 68.7%, respectively, of the cases already in the top 5 (43.8% in the top 1 for both), followed by ZDOCK (56.3%) and LightDock (37.4%).

The bottom row in Figure 2 (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.

ClusPro offers an automated masking of non-CDR regions for antibodies. We had, however, manually defined the HV loops for consistency between all software. We therefore evaluated whether this affected the docking performance by repeating the docking using the automated masking of non-CDR regions for all docking scenarios. The automatic identification of the CDR regions in the first scenario led to a higher success rate (e.g., top 10 increased from 18.7% to 31.3%) (see Figure S1). On the other hand, in the second scenario the automated masking of non-CDR regions, while overall leading to an improvement of the quality of the generated models, resulted in a decrease of the success rate (except for the top 10 of scenario 2 that increased from 67.7% to 75%). For the third scenario, as expected, driving the docking using the automatic masking of the non-CDR regions rather than the real interface led to a significant decrease in both the success rate (e.g., top 10 drops from 100% to 87.5%) and quality of the generated models.

Overall, Figure 2 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 the various 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.

Figure 3 shows the performance for each complex. Complexes are grouped into two classes, rigid and medium, according to the classification provided in docking benchmark BM5, which is based on the interface root-mean-square deviation

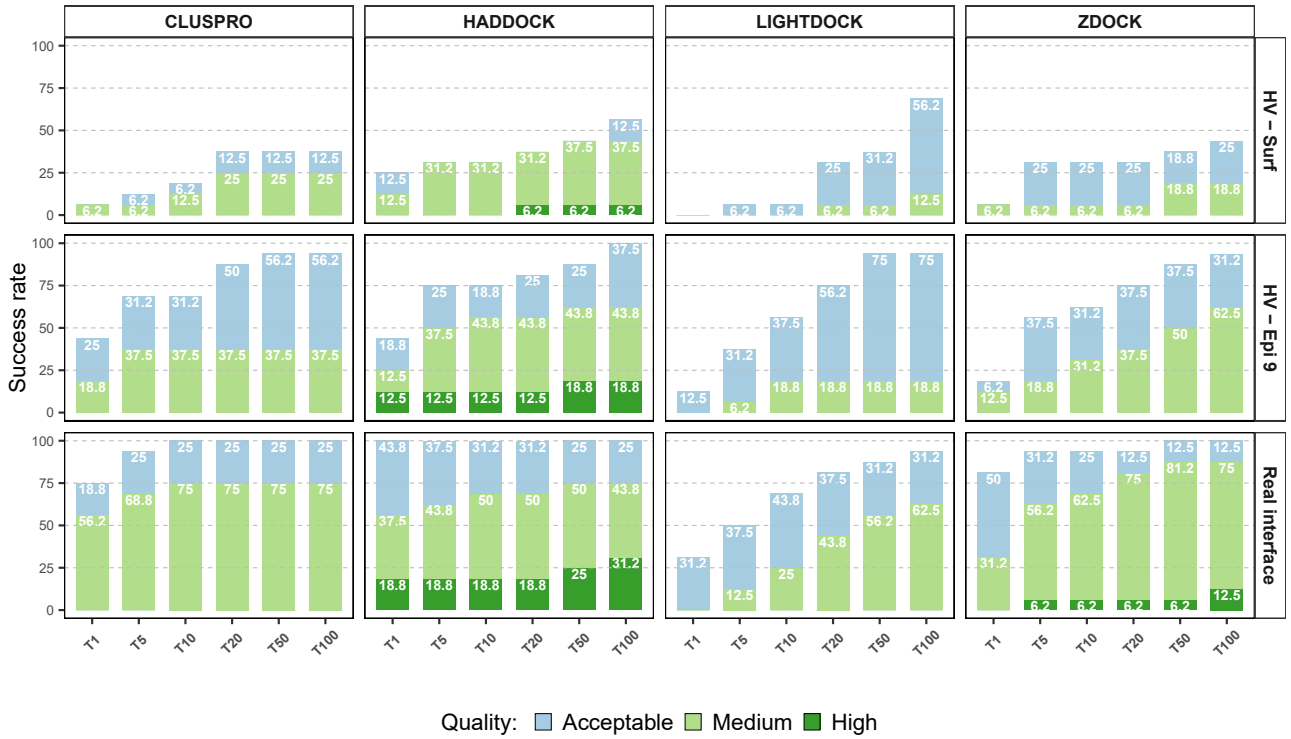


Figure 2. ClusPro, HADDOCK, LightDock, and ZDOCK Success Rate for the Three Scenarios Described in This Work as a Function of the Top 1, 5, 10, 20, 50, and 100 Ranked Models

The top row (HV-Surf) shows the success rate using the antibody HV loops and the entire antigen surface as restraints. The second row represents the success rate achieved by driving the docking with the antibody HV loops and a loose epitope definition using a 9 Å cutoff. The third row shows the docking results using the true interfaces (defined at 4.5 Å). The color coding indicates the quality of the models according to CAPRI criteria (see STAR Methods). See also Tables S1–S4 and Figures S1–S4.

(i-RMSD) and the fraction of non-native residue contacts (Vreven et al., 2015).

Overall, all methods achieve good results for the most rigid structures (PDB: 3EOA, 3MXW, 4G6M, and 3RVW) when a vague definition of the epitope is provided. HADDOCK is able to provide for 3MXW and 4G6M sub-angstrom high-quality models as top-ranking models. Although HADDOCK and LightDock allow flexibility of the molecules, the difference in terms of accuracy between them and ClusPro or ZDOCK for rigid and medium complexes is not striking. This is probably due to the rather limited conformational change that occurs upon binding of the antibody to the antigen (average i-RMSD of 0.97 ± 0.50 Å for all entries in this dataset, with minimum and maximum values of 0.39 Å and 1.86 Å, respectively) (Vreven et al., 2015). For each complex and scenario, the i-RMSD and fraction of native contacts (F_{nat}) values for the first acceptable and best models for all of the software used in this study are reported in Tables S1–S4.

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 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 four 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 4 for the three different scenarios.

Comparing Figures 4 and 2, 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

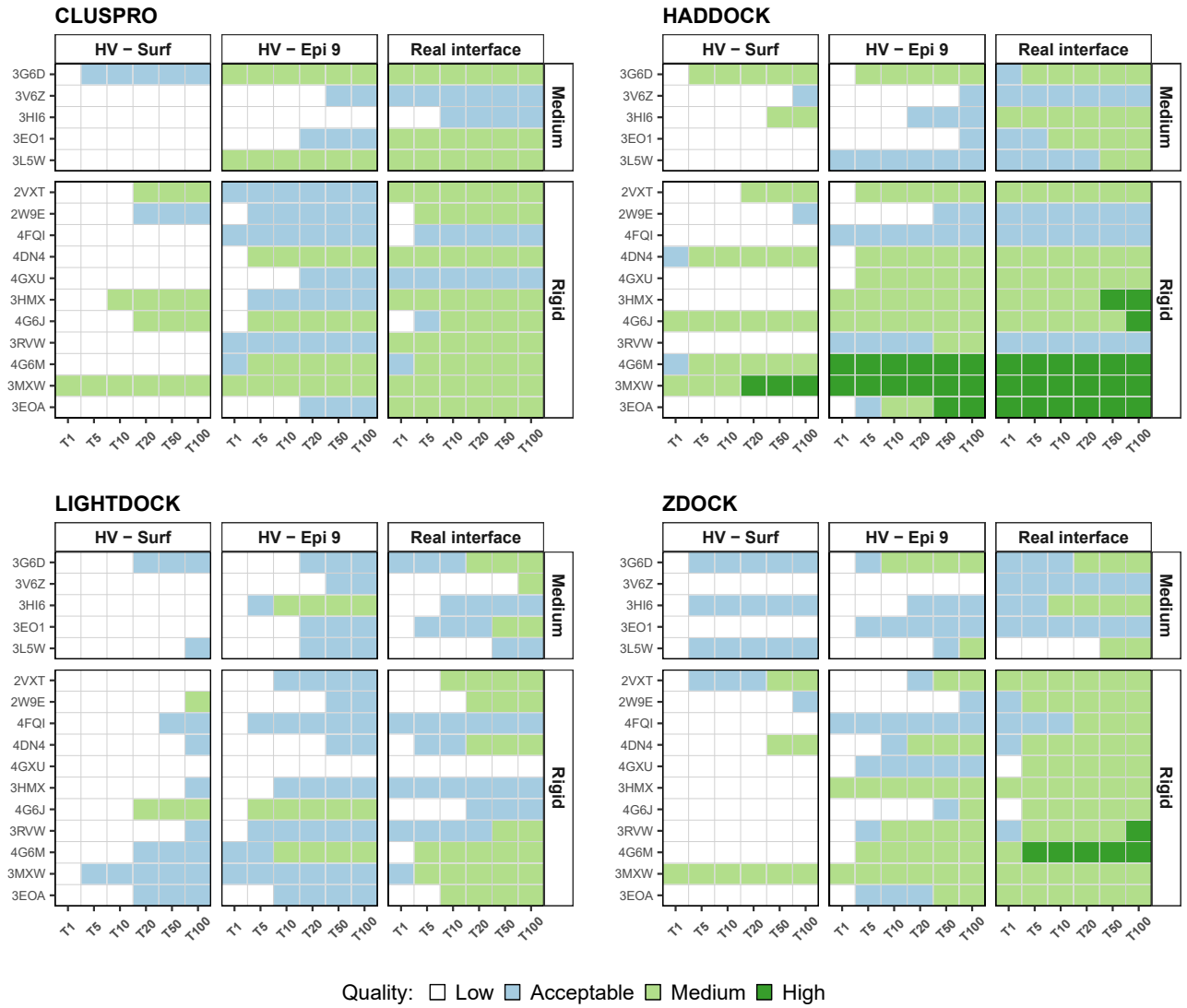


Figure 3. Performance of ClusPro, HADDOCK, and ZDOCK as a Function of the Number of Ranked Models per Complex and per Scenario

The dataset is split into rigid and medium complexes according to the docking benchmark 5 (BM5) definition. For each class, the complexes are sorted from bottom to top in increasing interface Cα-RMSD between the reference and the superimposed unbound structures. The color coding indicates the quality of the models according to CAPRI criteria (see STAR Methods).

0.83 ± 0.10, 0.93 ± 0.04, and 0.99 ± 0.003, respectively, for the HV-Surf, HV-Epi 9, and 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 (or not) of epitope information.

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 includes the information about the binding interface at the scoring stage, but HADDOCK is the only system that uses this information to drive the sampling (the information is encoded into an additional energy term that generates forces to drive the minimization and molecular dynamics steps). The effect of these 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 suite in Figure 5 (see also Tables S5–S8). One can clearly see 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). There is, however, the danger that no single acceptable model might be generated in the case of bad information. The other software, ClusPro, LightDock, and ZDOCK, use the interface information only at the scoring level (except for LightDock, which 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

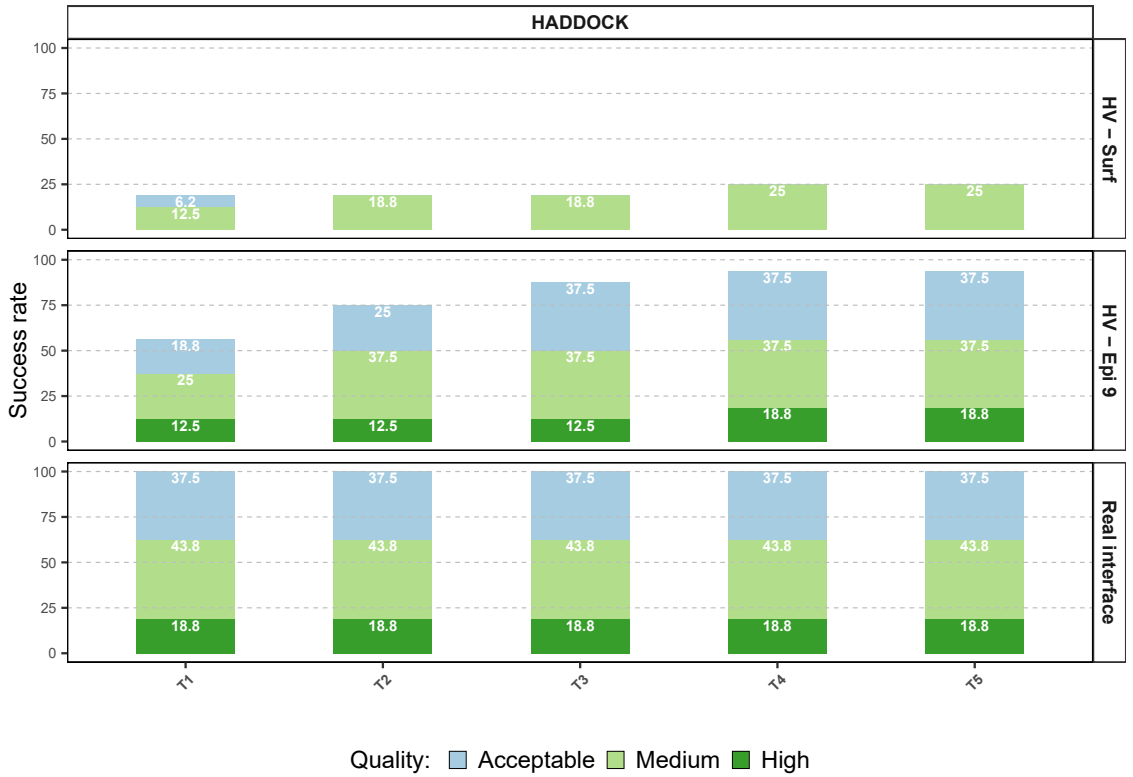


Figure 4. HADDOCK Cluster-Based Success Rate for the Three Docking Scenarios as a Function of the Top 1, 2, 3, 4, and 5 Ranked Clusters
The color coding indicates the quality of the models according to CAPRI criteria (see STAR Methods).

number of near-native models generated. In this case, the scoring becomes crucial in identifying the acceptable models.

H3 Loop Modeling Performance

As already mentioned, the H3 loop of antibodies is the most important loop involved in antigen recognition. Its accurate modeling is still a challenge due to its high structural and sequence variability. Of the four docking software applications used in this work, two allow for conformational changes during the docking, namely HADDOCK and LightDock. We analyzed their capability of inducing the appropriate conformational changes of the loop upon binding with the antigen. For this, we superimposed the antibody framework residues of the bound and unbound structure and calculated the RMSD of H3 ($H3_{unbound}$). We then repeated the same procedure for each docking model compared with the native complex ($H3_{model}$). For both HADDOCK and LightDock, models produced from the different scenarios were merged and split into correct (i-RMSD $\leq$ 4Å) and wrong models (i-RMSD > 4Å). Figure 6 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 6A) 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 1.25 Å. In the case of LightDock (Figure 6B), the final selected H3 loop conformation from normal modes remains very close to the unbound form, with no remarkable changes in terms of RMSD.

To further investigate the impact of the HADDOCK flexible refinement stage on the H3 loop conformation, we analyzed the F_{nat} that H3 makes at the rigid-body docking stage ($H3_{it0}$) and after flexible refinement ($H3_{water}$). Figure 7 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 7 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 improving the H3 conformation during the simulation. A deeper analysis of the impact of HADDOCK flexible refinement on the quality of the models is reported in terms of distribution of i-RMSD and F_{nat} differences in Figures S5 and S6.

DISCUSSION

This work aimed at comparing four well-established protein-protein docking software suites (ClusPro, HADDOCK, LightDock, and ZDOCK) in their accuracy for predicting antibody-antigen

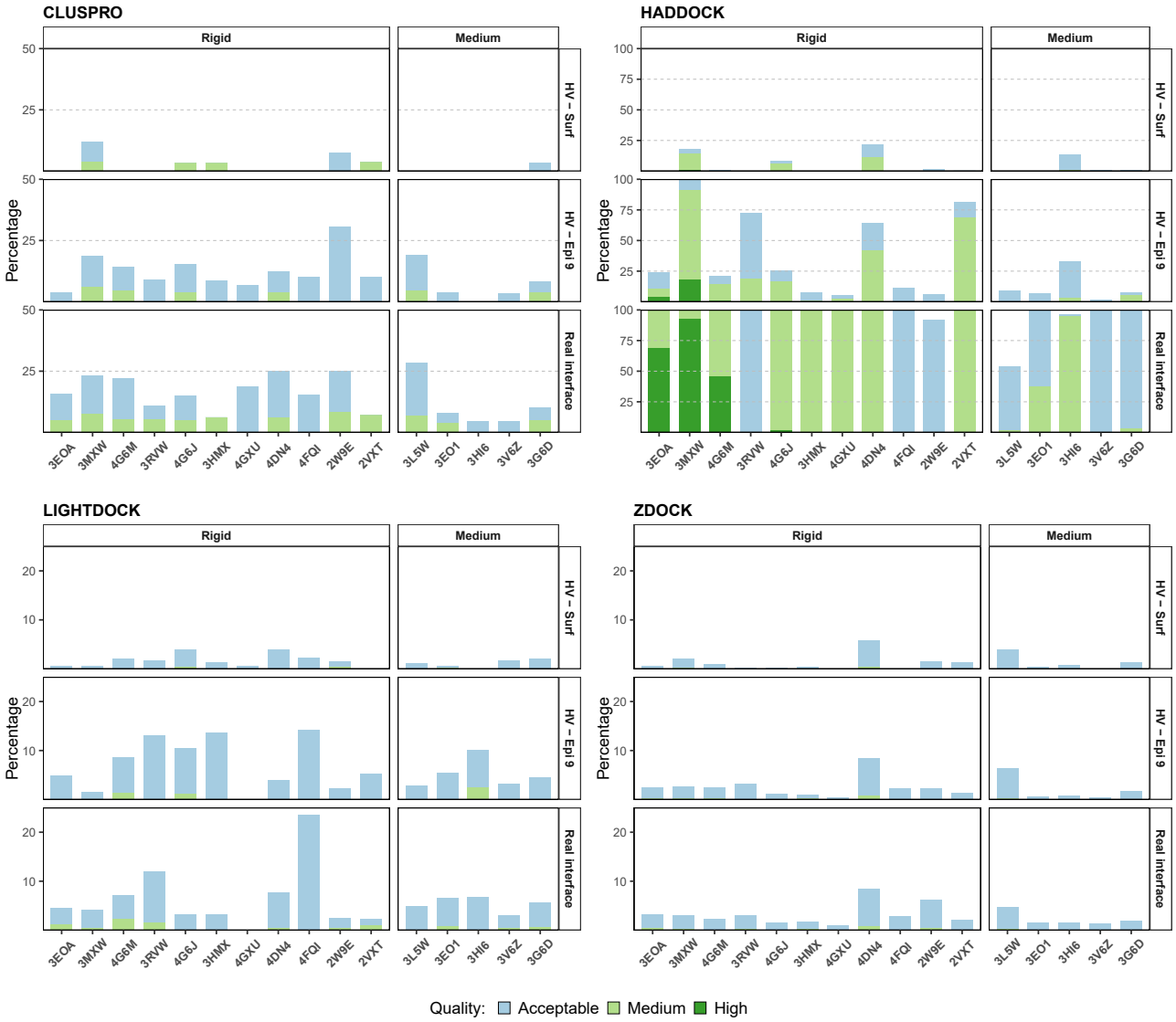


Figure 5. Percentages of Acceptable, Medium and High-Quality Models Generated by Each Software per Complex and per Scenario

Complexes are split into rigid and medium categories according to the Docking Benchmark5 definition. Note that the y axis scales are different for each docking method for better readability. The color coding indicates the quality of the models according to CAPRI criteria (see STAR Methods). See also Tables S5–S8.

complexes assuming different amounts of available experimental information. Alternative docking methods do exist, such as SwarmDock and SnugDock, but these were not considered in this work for various reasons. The SwarmDock server does not allow the user to automate the runs for high-throughput docking, requiring manual selection of each residue in the HV loops and epitope. Furthermore, the option in SwarmDock to specify specific interface residues is still experimental and not yet properly tested (Dr. Paul Bates, personal communication). For these reasons, we did not include it in this work. With regard to SnugDock, it requires the two molecules to be preoriented before optimizing them, which prevents making a fair comparison with the docking scenarios considered in this work.

While for the antibody a proxy of the binding interface can be extracted from the sequence and in particular the HV loops (Al-Lazikani et al., 1997; Novotný et al., 1983; Sela-Culang et al., 2013), predicting the epitope on the antigen is a more challenging problem. Despite many efforts to develop reliable methods to predict the antibody-specific epitopes, most approaches published to date are still very limited in their accuracy. Driving a docking process or filtering docking poses using predicted epitopes can thus be detrimental to the accuracy of the complex structure prediction. There are, however, experimental methods such as H/D exchange detected by MS that can provide more accurate information. In this work, we focus on defining the best way of including available epitope information to guide the modeling process. To this end, we investigated three different scenarios that mimic different levels of knowledge about the epitope: In the first scenario, no information is available and the sampling/scoring must involve the entire surface of the antigen; in the second case, a loose definition of the epitope is assumed (a larger surface than the real interface); finally, the third scenario

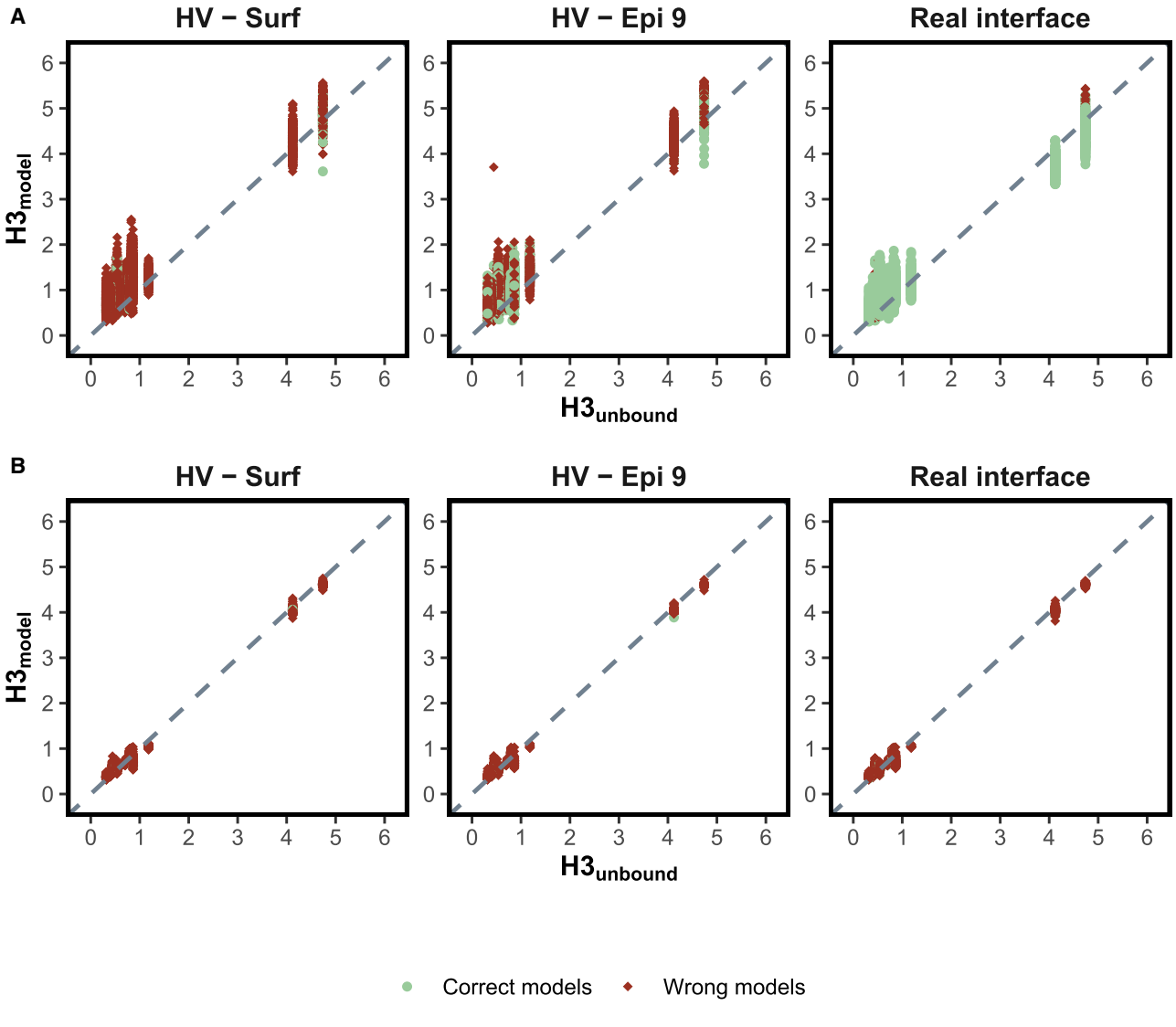


Figure 6. H3 Loop RMSD (Å) from the Bound Conformation for the Docked Models ($H3_{model}$) versus the Starting Unbound Conformation ($H3_{unbound}$)

HADDOCK models (A) and LightDock models (B). Correct and wrong models were defined according to their i-RMSD from the reference structure using a 4 Å cutoff.

represents the ideal case whereby both antibody and antigen interfaces are well characterized. HADDOCK, which has been developed to make use of available information, is performing best in terms of both success rate and quality of the generated models in all of the described scenarios. Analysis of the HADDOCK scoring performance per scenario indicates that clustering is beneficial provided some reasonably accurate information is available for the epitope. If this is not the case, a single structure-based scoring approach might perform better, but still with rather low success rate (31.2% for both top 5 and top 10).

In this work, the unbound forms of antibodies were used to perform the docking, but often in reality only the antibody sequence is known; therefore, we assessed HADDOCK accuracy using antibody models generated with the PIGSpro webserver (https://cassandra.med.uniroma1.it/pigspro/) (Lepore et al., 2017). There are many other tools with which to model antibodies, but it is outside the scope of this work to compare their impact on the docking results. Results and method are reported in Figure S2. The use of antibody models results in a significant decrease of the success rate in all scenarios. For example, the top 10 success rate using antibody models drops from 31.2%, 75%, and 100% to 18.7%, 68.7%, and 75%, respectively, for the first, second, and third scenarios. This will of course depend on the antibody modeling strategy, but it is outside the scope of this work to compare different modeling software/servers.

The sampling used in this work for HADDOCK differs from the default settings. While for the first scenario this is the recommended setting for cases with limited or no information on binding sites (this was also the sampling used in the BM5 benchmarking [Vreven et al., 2015]), the sampling in the two other scenarios

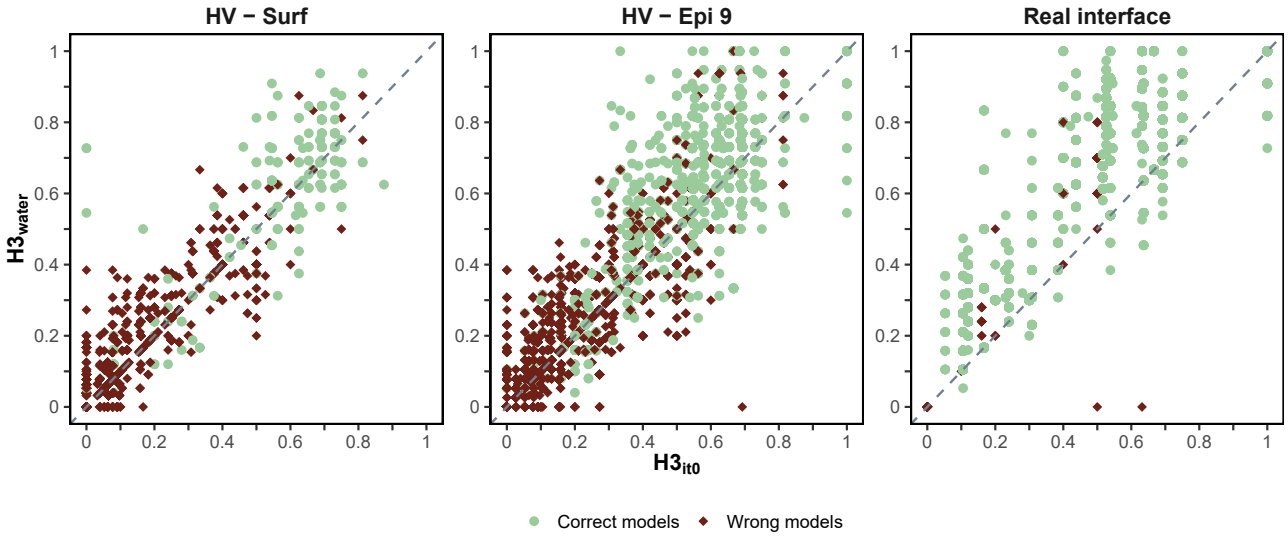


Figure 7. Comparison of the Fraction of Native Contacts (F_{nat}) Made by the H3 Loop after the Rigid-Body Docking Stage ($H3_{it0}$) (horizontal axis) and after Flexible/Water Refinement ($H3_{water}$) (vertical axix) of the HADDOCK Runs

Values above the diagonal correspond to an improvement in F_{nat}. See also Figures S5 and S6. Correct and wrong models were defined according to their i-RMSD from the reference structure using a 4Å cutoff.

(5,000/400/400) differs from the default settings (1,000/200/200). For comparison, we reran the docking for these two scenarios using default sampling settings. The results, presented in Figure S3, show that even with default settings, given a reasonable definition of the epitope (HV-Epi 9 scenario), good results are obtained with default sampling, with 87.5% top 10 success rate versus 75% for the increased sampling. For the third scenario, the reduced sampling does not seem to have an impact on the success rate but rather slightly affects the quality of the generated models.

Another difference to note for HADDOCK between scenarios 2 and 3 is the definition of the interface information on the epitope side: while for the true interface (scenario 3) the interface residues were treated as ''active'' in the definition of interaction restraints, they were defined as passive in scenario 2. To verify whether this might have affected the performance, we repeated scenario 3 defining the epitope residues as passive. The results, presented in Figure S4, show that the passive definition of the antigen interface residues leads to a decrease of the success rate (e.g., from 100% to 87.6% for the top 10), while the quality of the generated models does not change significantly.

We further analyzed the capability of HADDOCK and LightDock, the only two software suites that allow flexibility of the components among the tested ones, in improving the H3 loop conformation. In fact, it has been largely demonstrated that the H3 loop is crucial for antigen recognition, but its modeling represents one of the biggest pitfalls in the antibody modeling field. Only when the H3 loop underwent a large conformational change upon binding did HADDOCK succeed in improving its conformation (measured in terms of RMSDs) toward the bound form, while the LightDock models did not show any significant conformational changes. While the induced conformational changes are rather limited in terms of RMSDs, the flexible refinement stages of HADDOCK do lead to a significant increase in the number of native contacts made by H3, and this effect is more evident when information about the epitope is provided to the system. This is relevant, since it will allow a better identification/prediction of key interactions.

One of the main benefits of this work is to offer researchers a clear overview about the state of the art of antibody-antigen structure prediction (for the software considered) and the various strategies that can be followed depending on the available information. Provided that a vague definition of the epitope can be obtained, reasonably accurate models can be obtained, with HADDOCK performing best among the four software applications compared. Finally, our analysis also indicates that there remain multiple opportunities for improvements, especially in modeling conformational changes, with the H3 loop as a particular challenge, but also in scoring, considering that all software achieved fairly good performance in the top 100, although this significantly dropped in most cases when only the top 10 or lower were considered.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY
- METHOD DETAILS
  - Dataset
  - Docking Scenarios
  - Docking Settings
  - HADDOCK Clustering Parameters
  - Evaluation Criteria
- DATA AND CODE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.str.2019.10.011.

ACKNOWLEDGMENTS

We wish to thank Pier Paolo Olimpieri (Department of Physics, Sapienza University) and Panagiotis Koukos (Bijvoet Center for Biomolecular Research, Utrecht University) for useful comments and discussions. This work has been carried out with the financial support of the European Union Horizon 2020 BioExcel (project #675728 and #823830) and EOSC-hub (project #777536) projects and the Dutch Foundation for Scientific Research (NWO) (TOP-PUNT grant 718.015.001).

The FP7 WeNMR (project #261572), H2020 West-Life (project #675858), and the EOSC-hub (project #777536) European e-Infrastructure projects are acknowledged for the use of the HADDOCK web portal, which makes use of the EGI infrastructure with the dedicated support of CESNET-MetaCloud, INFN-PADOVA, NCG-INGRID-PT, TW-NCHC, SURFsara, and NIKHEF, and the additional support of the national GRID Initiatives of Belgium, France, Italy, Germany, the Netherlands, Poland, Portugal, Spain, UK, Taiwan, and the US Open Science Grid.

AUTHOR CONTRIBUTIONS

F.A., B.J.G., and J.R.T. performed the computational analysis. A.M.J.J.B. and F.A. conceived and designed and interpreted the data. A.M.J.J.B. and F.A. wrote the paper with contributions from the other authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: March 22, 2019
Revised: July 3, 2019
Accepted: October 18, 2019
Published: November 11, 2019

REFERENCES

Al-Lazikani, B., Les, 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.

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.

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

Choi, Y., and Deane, C.M. (2010). FREAD revisited: accurate loop structure prediction using a database search algorithm. Proteins *78*, 1431–1440.

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.

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

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.

Fontayne, A., De Maeyer, B., De Maeyer, M., Yamashita, M., Matsushita, T., and Deckmyn, H. (2007). Paratope and epitope mapping of the antithrombotic antibody 6B4 in complex with platelet glycoprotein Ibα. J. Biol. Chem. *282*, 23517–23524.

Hubbard, S.J., and Thornton, J.M. (1993). NACCESS, Computer Program (University College, London). http://wolf.bms.umist.ac.uk/naccess/.

Janin, J., Henrick, K., Moult, J., Eyck, L.T., Sternberg, M.J.E., Vajda, S., Vakser, I., and Wodak, S.J. (2003). CAPRI: a critical assessment of PRedicted interactions. Proteins *52*, 2–9.

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 (Zenodo). https://doi.org/10.5281/zenodo.2537146.

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. Bioorg. 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–W524.

Leem, J., Dunbar, J., Georges, G., Shi, J., and Deane, C.M. (2016). ABodyBuilder: automated antibody structure prediction with data-driven accuracy estimation. MAbs *8*, 1259–1268.

Lepore, R., Olimpieri, P.P., Messih, M.A., and Tramontano, A. (2017). PIGSPro: prediction of immunoGlobulin structures v2. Nucleic Acids Res. *45*, W17–W23.

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.

Lim, X.X., Chandramohan, A., Lim, X.Y.E., Crowe, J.E., Lok, S.M., and Anand, G.S. (2017). Epitope and paratope mapping reveals temperature-dependent alterations in the dengue-antibody interface. Structure *25*, 1391–1402.e3.

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.

McLachlan, A.D. (1982). Rapid comparison of protein structures. Acta Crystallogr. Sect. A *38*, 871–873.

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 *52*, 51–67.

Messih, M.A., Lepore, R., Marcatili, P., and Tramontano, A. (2014). Improving the accuracy of the structure prediction of the third hypervariable loop of the heavy chains of antibodies. Bioinformatics *30*, 2733–2740.

Meyer, P.A., Socias, S., Key, J., Ransey, E., Tjon, E.C., Buschiazzo, A., Lei, M., Botka, C., Withrow, J., Neau, D., et al. (2016). Data publication with the structural biology data grid supports live analysis. Nat. Commun. *7*, 10882.

Morea, V., Lesk, A.M., and Tramontano, A. (2000). Antibody modeling: implications for engineering and design. Methods *20*, 267–279.

Moreira, I.S., Fernandes, P.A., and Ramos, M.J. (2010). Protein-protein docking dealing with the unknown. J. Comput. Chem. *31*, 317–342.

Morin, A., Eisenbraun, B., Key, J., Sanschagrin, P.C., Timony, M.A., Ottaviano, M., and Sliz, P. (2013). Collaboration gets the most out of software. Elife *2*, e01456.

Narciso, J.E.T., Uy, I.D.C., Cabang, A.B., Chavez, J.F.C., Pablo, J.L.B., Padilla-Concepcion, G.P., and Padlan, E.A. (2011). Analysis of the antibody structure based on high-resolution crystallographic studies. Nat. Biotechnol. *28*, 435–447.

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.

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.

Ritchie, D. (2008). Recent progress and future directions in protein-protein docking. Curr. Protein Pept. Sci. *9*, 1–15.

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.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 *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.

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.

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.

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. Proteins *60*, 176–180.

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.

Weitzner, B.D., Jeliazkov, J.R., Lyskov, S., Marze, N., Kuroda, D., Frick, R., Adolf-Bryfogle, J., Biswas, N., Dunbrack, R.L., and Gray, J.J. (2017). Modeling and docking of antibody structures with Rosetta. Nat. Protoc. *12*, 401–416.

Wu, T.T. (2004). 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.

Yamashita, K., Ikeda, K., Amada, K., Liang, S., Tsuchiya, Y., Nakamura, H., Shirai, H., and Standley, D.M. (2014). Kotai Antibody Builder: automated high-resolution structural modeling of antibodies. Bioinformatics *30*, 3279–3280.

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.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and Algorithms		
ClusPro	(Kozakov et al., 2017)	https://cluspro.org/
HADDOCK	(De Vries et al., 2010)	https://haddock.science.uu.nl/services/HADDOCK2.2
LightDock	(Jiménez-García et al., 2019)	https://github.com/brianjimenez/lightdock
ZDOCK	(Pierce et al., 2011)	http://zdock.umassmed.edu
PIGSpro	(Lepore et al., 2017)	https://cassandra.med.uniroma1.it/pigspro/
Deposited Data		
Raw and analyzed data	This paper	https://data.sbgrid.org/dataset/686/
Others		
Docking Benchmark v5	(Vreven et al., 2015)	http://zlab.umassmed.edu/benchmark/

LEAD CONTACT AND MATERIALS AVAILABILITY

All data gathered for this work are publicly available and can be find at: https://data.sbgrid.org/dataset/686/.

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Alexandre M.J.J. Bonvin (a.m.j.j.bonvin@uu.nl).

METHOD DETAILS

Dataset

The dataset used in this work includes 16 complexes, 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. Antibody structures were renumbered using an in-house R script. Only the 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).

Docking Scenarios

All methods allow the user to provide information about the binding interface. Nevertheless, only HADDOCK applies a purely data-driven sampling strategy in order to create models in agreement with the provided information. LightDock uses this 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. In this case the methods are able to assign a better score to the models that better satisfy the given restraints.

For each complex three different docking runs were performed in order to represent different scenarios corresponding to different amounts of available experimental information (Figure 1):

(1) **HV - Surf**: No information about the epitope residues are provided to the docking algorithm. Only knowledge of the antibody HV loops, defined according to the Chothia numbering scheme (Al-Lazikani et al., 1997), is used in the docking. For HADDOCK this was complemented by all solvent-exposed antigen residues defined by selecting those with a relative accessible surface area (RSA) $\geq$ 40% (calculated with NACCESS (Hubbard and Thornton, 1993)).
(2) **HV – Epi 9**: This scenario represents the case in which only a vague definition of the epitope region is provided. To mimic this situation the epitope was defined by selecting all antigen residues within 9Å from the antibody. The docking run was performed by providing to the docking algorithms the HV loops residues and the 9Å epitope.
(3) **Real interface**: In this ideal scenario both antibody and the antigen interfaces are well characterized. All interface residues selected using a distance cutoff of 4.5Å were given to the docking software.

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 webserver (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 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 we used release 0.5.6 (Jiménez-García et al., 2019) of the software which provides a mechanism for including residue restraints. At the receptor level, the surface swarms used in the simulation 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 (DFIRE (Zhou and Zhou, 2009) in this work) 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: 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 run and to 10000 for the HV – Surf scenario. The flexible (it1) and water refinement sampling were set to 400 models for all scenarios (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 (Van Zundert et al., 2016) (https://haddock.science.uu.nl). 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 first two scenarios. 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.

HADDOCK Clustering Parameters

By default, HADDOCK performs a cluster analysis. In this work clustering was based on the Fraction of Common Contacts (FCC) (Rodrigues et al., 2012) using 0.6 as cutoff and 4 as minimum cluster size. The clusters were sorted according to the average HADDOCK score of the best 4 model of each cluster, from the lowest HADDOCK score to the highest.

Evaluation Criteria

Docking models were classified as high (***), medium (**) of low (*) quality according to the CAPRI criteria (Janin et al., 2003; Méndez et al., 2003) (see Table 1) based on their similarities with the native structure by calculating the interface root mean square deviation (i-RMSD), the ligand root mean square deviation (L-RMSD) and the fraction of native contacts (F_{nat}). Briefly, the i-RMSD is calculated on the backbone atoms of all interface residues of the native complex defined using a 10Å cutoff and the L-RMSD is calculated by superimposing on the backbone atoms of the antibody and calculating the RMSD of the antigen backbone atoms. Finally, F_{nat} is calculated as number of native contacts in a docking model divided by the total number of contacts in the reference structure. F_{nat} has been calculated using is house scripts while fitting and RMSD calculations were performed using the McLachlan algorithm (McLachlan, 1982) as implemented in the program ProFit (http://www.bioinf.org.uk/software/profit/) from the SBGrid distribution (Morin et al., 2013).

DATA AND CODE AVAILABILITY

All models generated using the four software and the different scenarios, together with their quality statistics and scores have been deposited into the SBGrid data repository (Meyer et al., 2016) and can be find at: https://data.sbgrid.org/dataset/686/. Links to the various software used in this work are provided in the Key Resources Table.