Optimization of protein-protein docking for predicting Fc-protein interactions

Mark Agostino[1,2,3], Ricardo L. Mancera[1], Paul A. Ramsland[3,4,5,6*], and Juan Fernández-Recio[2*]

[1]School of Biomedical Sciences, Curtin Health Innovation Research Institute and Curtin Institute for Computation, Curtin University, Perth WA 6102, Australia

[2]Joint BSC-CRG-IRB Research Program in Computational Biology, Life Sciences Department, Barcelona Supercomputing Center, Barcelona 08034, Spain

[3]Centre for Biomedical Research, Burnet Institute, Melbourne, VIC 3004, Australia.

[4]School of Science, RMIT University, Bundoora, VIC 3083, Australia

[5]Department of Surgery Austin Health, University of Melbourne, Heidelberg, VIC 3084, Australia

[6]Department of Immunology, Monash University, Alfred Medical Research and Education Precinct, Melbourne, VIC 3004, Australia

*Correspondence should be addressed to: paul.ramsland@rmit.edu.au, juanf@bsc.es

Abstract

The antibody crystallizable fragment (Fc) is recognized by effector proteins as part of the immune system. Pathogens produce proteins that bind Fc in order to subvert or evade the immune response. The structural characterization of the determinants of Fc-protein association is essential to improve our understanding of the immune system at the molecular level and to develop new therapeutic agents. Furthermore, Fc-binding peptides and proteins are frequently used to purify therapeutic antibodies. Although several structures of Fc-protein complexes are available, numerous others have not yet been determined. Protein-protein docking could be used to investigate Fc-protein complexes; however, improved approaches are necessary to efficiently model such cases. In this study, a docking-based structural bioinformatics approach is developed for predicting the structures of Fc-protein complexes. Based on the available set of X-ray structures of Fc-protein complexes, three regions of the Fc, loosely corresponding to three turns within the structure, were defined as containing the essential features for protein recognition and used as restraints to filter the initial docking search. Rescoring the filtered poses with an optimal scoring strategy provided a success rate of approximately 80% of the test cases examined within the top ranked 20 poses, compared to approximately 20% by the initial unrestrained docking. The developed docking protocol provides a significant improvement over the initial unrestrained docking and will be valuable for predicting the structures of currently undetermined Fc-protein complexes, as well as in the design of peptides and proteins that target Fc.

Keywords: structural immunology, fragment crystallizable, infectious disease, immune response, structure-based design, structural bioinformatics

Background

Antibodies, or immunoglobulins, are the key components of the immune system mediating host defense and protection. Five immunoglobulin (Ig) isotypes have been characterized in humans, these being IgG, IgE, IgA, IgM and IgD. IgG is the most abundant circulating antibody and the principal antibody associated with the booster/secondary response (Cruse and Lewis 2010). IgE is typically associated with the allergic response (Platts-Mills and Woodfolk 2011). IgA is the most abundant antibody in mucosal interfaces and secretions, and typically exists as a dimer (Macpherson, et al. 2008). IgM is the largest of the five classes, with a molecular weight of around 1 MDa, and is the first isotype to appear upon initial antigen exposure (Cruse and Lewis 2010). The function and interactions of IgD are a subject of ongoing investigation (Chen and Cerutti 2011). Despite the varying functions of the five isotypes, they are all structurally related, comprising two key regions: the antigen-binding fragment (Fab) and the crystallizable fragment (Fc).

Protein recognition of the antibody Fc is necessary for mediating antibody-dependent immune effector functions. Furthermore, several human pathogens, notably *Streptococcus, Staphylococcus aureus* and *Plasmodium falciparum*, produce proteins that are capable of binding to the antibody Fc region (Wines, et al. 2012). The pathogen proteins typically bind to the same region of the Fc as host effector proteins, which may allow the pathogens to subvert or evade attack by the immune system. Such proteins are frequently used to purify monoclonal antibodies of specific isotypes and serve as valuable leads for the discovery of synthetic peptides for this purpose (Gautam and Loh 2013; Hatanaka, et al. 2012; Hober, et al. 2007; Jerlström, et al. 1996; Sandin, et al. 2002; Wang, et al. 2014; Zhao, et al. 2014a; Zhao, et al. 2014b). Understanding Fc-protein interactions at the molecular level would improve our understanding of host-pathogen interactions and foster new therapeutic applications.

Protein binding to Fc typically occurs to one of two binding sites (Figure 1). The first site,

3

which occurs at the hinge-CH2 interface of IgG and the CH2-CH3 interface of IgE, is the binding site for Fcγ receptors (FcγRs) and the high affinity Fcε receptor, FcεRI. The majority of interactions between these receptors and Fc have been structurally characterized by X-ray crystallography (Holdom, et al. 2011; Mimoto, et al. 2013; Mizushima, et al. 2011; Ramsland, et al. 2011; Sondermann, et al. 2000). The FcγRs and FcεRI all exhibit high sequence similarity and conserved protein folds. No similar proteins capable of binding to IgA or IgM are currently known, nor are any other proteins known to binding to this region of IgA or IgM, suggesting that this site may only be relevant for protein binding in IgG and IgE. The second site occurs at the CH2-CH3 interface of IgG and IgA, and the structurally equivalent CH3-CH4 interface of IgE and IgM. A variety of antibody receptors and pathogen-derived proteins, exhibiting a diverse range of folds, have been structurally characterized at this site in IgA, IgG and IgE (Table 1). Although there are no IgM-protein complexes currently available, biochemical evidence suggests that both pathogen proteins and Fc receptors bind to IgM-Fc at the CH3-CH4 interface (Ghumra, et al. 2008; Ghumra, et al. 2009). This site, unlike the first site, is therefore of relevance for protein binding to Fc across the range of antibody isotypes. Furthermore, there are numerous proteins that have been demonstrated biochemically to interact with this site of Fc, however, the complexes for these have not yet been structurally determined. These include the interaction of Fcα/μR and pIgR with both IgA and IgM (Coyne, et al. 1994; Røe, et al. 1999; Yang, et al. 2013), the interaction of streptococcal M and β proteins with IgA (Jerlström, et al. 1996; Johnsson, et al. 1999), and the interaction of Duffy binding-like (DBL) domains of certain variants of the *Plasmodium falciparum* erythrocyte membrane antigen 1 (PfEMP1) with IgG and IgM (Rasti, et al. 2006). The interaction of tripartite motif-containing protein 21 (TRIM21) with IgG has been structurally characterized by X-ray crystallography (James, et al. 2007) and interactions of TRIM21 with IgM (Mallery, et al. 2010) and IgA (Bidgood, et al. 2014) have been demonstrated; however, experimental structures of the latter interactions have so far not been determined. The large size of

IgM-Fc (Czajkowsky and Shao 2009; Müller, et al. 2013), as well as the high flexibility associated with some regions of the DBL domains of PfEMP1 (Khunrae, et al. 2009), present particular challenges for the application of experimental methods for structure determination. Manual docking of the DBL4ζ domain of PfEMP1 to IgM-Fc (Czajkowsky, et al. 2010) was subsequently shown to be incorrect (Semblat, et al. 2015), highlighting the need for automated computational methods with demonstrated predictive power.

Protein-protein docking is a computational approach whereby the structure of a protein-protein complex is predicted starting from the structures of the unbound proteins. In general, a sampling step to generate poses is performed, typically a fast Fourier transform (FFT) based grid search (Chen, et al. 2003; Gabb, et al. 1997), combined with one or more scoring steps, wherein the poses are ranked. Despite considerable research into protein-protein docking, existing approaches struggle to rank highly the correct poses for the variety of protein-protein interactions, and exhibit target-dependent performance (Huang 2015). Of particular relevance to the current study is that accurate structures for the two Fc-protein complexes in the protein-protein docking benchmark (Hwang, et al. 2010), the complexes of IgG-Fc with Protein G (PDB 1FCC) (Sauer-Eriksson, et al. 1995) and Protein A (PDB 1FC2) (Deisenhofer 1981), are generally challenging to predict or rank highly for the majority of docking programs (Huang 2015). Restraints could be used to improve docking outcomes, and since many proteins target a common site of the Fc region across immunoglobulin isotypes (Wines, et al. 2012), it could be possible to define docking restraints that are generally applicable to studying protein binding to this region. These would then form a key part of a docking protocol specifically for studying Fc-protein interactions.

In this study, we have analyzed a series of structurally characterized Fc-protein complexes to define common Fc regions involved in protein recognition. The regions defined are likely to be generally applicable for studying protein binding to the CH2-CH3 interface of IgA, IgG and IgD, as

well as the CH3-CH4 interface of IgE and IgM. Utilizing the regions as restraints significantly improves the performance of rigid body protein-protein docking for predicting Fc-protein complexes. Further improvements are achieved through utilizing an optimized scoring strategy to reduce the number of poses to a sufficiently small number for manual examination and refinement while maintaining a reasonable level of accuracy. The protocol was optimized using a set of docking runs where one or both of the Fc and the protein structures were generated by homology modelling, representing both the most realistic and challenging modelling scenario. Application to a set of cross and unbound docking cases not included in the optimization yielded similar success rates to the optimization, indicating the utility of the protocol in studying novel cases. The developed protocol will be useful for determining structurally uncharacterized Fc-protein interactions, as well as in the design of Fc-binding peptides and proteins.

Methods

Selection and preparation of structures

The X-ray crystal structures of a series of Fc-protein complexes were identified using the IMGT 3D-StructureDB (Ehrenmann, et al. 2010) (imgt.org/3Dstructure-DB) and obtained from the Protein Data Bank (PDB) (Berman, et al. 2000) (www.rcsb.org) (Table 1). The complexes included IgA-Fc, IgG-Fc and IgE-Fc crystallized with a variety of proteins from pathogens as well as human effector proteins. Complexes were selected where the pathogen/effector proteins were bound to the CH2-CH3 interface of IgA and IgG, or the CH3-CH4 interface of IgE. Structures were checked for missing residues prior to docking using the Protein Preparation Wizard in Maestro 9.3 (Schrödinger). Hydrogen atoms were not considered in docking. The receptor and ligand structures were separated and the ligand subjected to random rotation and translation, to shift it from its starting coordinates; this was performed manually using the Local Transformation tool in Maestro.

Identification of common Fc regions involved in interactions

The DIMPLOT module of LigPlot+ v.1.4 was used to identify hydrogen bonding and van der Waals interactions in the crystal structures (Laskowski and Swindells 2011). The strands, turns and loops of the Fc domains that the interacting residues were part of, according to their IMGT definitions (Lefranc, et al. 2005), were identified. Guided by both the IMGT definitions of Fc secondary structure regions and the residues involved in interactions in the Fc-protein crystal structures, new Fc regions were defined representing the portions of the Fc most involved in Fc-protein interactions. These were then used as restraints to filter the docking results.

Rigid body protein-protein docking

Rigid body protein-protein docking was performed using ZDOCK 3.0.1 (Pierce, et al. 2011). No regions of the ligand or receptor were blocked in the docking simulation. Dense ligand sampling was performed and the entire set of 54000 poses was retained. In all cases, the Fc was treated as the "receptor" (fixed in location) and the protein binding to the Fc treated as the "ligand" (mobile). Although flexibility in Fc has been reported and investigated, particularly for IgE (Wurzburg and Jardetzky 2009), the rigid body approximation was deemed an appropriate treatment for Fc, since a survey of available Fc structures indicates antibody isotypes largely adopt a conserved angle between the CH2-CH3/CH3-CH4 domains (Figure S1); variations to the angle are typically associated with alterations to Fc glycosylation, mutations to Fc residues, and protein binding to the hinge-CH2/CH2-CH3 site (i.e., binding of Fcγ receptors and FcεRI to IgG and IgE respectively). To evaluate pose quality, the ligand RMSD (l_RMSD) and the interface RMSD (i_RMSD) were calculated for each ligand pose compared to the reference structure of the ligand in complex with Fc. The l_RMSD was calculated on all backbone atoms of the ligand (after superimposing the receptors). The i_RMSD was

calculated on all backbone atoms for ligand residues within 10 Å of the Fc in the reference structure. Poses with an i_RMSD of less than or equal to 4.0 Å were considered to represent good quality solutions. This cutoff captures the majority of higher quality solutions considered at least "acceptable" according to the CAPRI criteria (Lensink and Wodak 2009), and was used as the primary quality metric to indicate docking success. Poses with an i_RMSD greater than 4.0 Å and an l_RMSD of less than or equal to 10.0 Å were considered to represent approximate solutions. This cutoff captures the majority of lower quality solutions still considered "acceptable" in CAPRI. Although upon inspection such solutions may not appear to accurately reproduce a given complex, they may be improved to a higher quality solution following the application of a refinement procedure (Schindler, et al. 2015). Thus, this cutoff was used as a secondary quality metric of docking success, applied only to docking runs where no poses satisfying the primary quality metric of docking success were identified by the ZDOCK sampling.

Cognate bound docking, cross-docking and docking with homology models were carried out. Cognate bound docking represents the most trivial scenario, where it is attempted to reproduce a given complex using both the ligand and receptor structures from that complex. Cross-docking is less trivial; in this scenario, it is attempted to reproduce a given complex using ligand and receptor structures obtained from different experiments. The most challenging scenario is docking with homology models, due to the difficulty in producing the binding conformation of a given protein with a homology model. In this scenario, homology models of either or both the ligand and receptor were prepared against their nearest suitable templates and docking was carried out to reproduce the complex. For this part of the study, three complexes were focused on, those of IgG-Fc with TRIM21 (PDB 2IWG), IgE-Fc with FcεRII (PDB 4EZM) and IgA-Fc with SSL7 (PDB 2QEJ). These were selected to have equal representation of the various Ig subtypes, but also because it was possible to identify suitable templates for homology modelling of the ligands. Various combinations of ligand and receptor structures gave

rise to a total of 21 docking combinations aiming to reproduce these complexes.

Homology modelling procedure

Model building was performed using the Structure Prediction Wizard in Prime. To model the ligands, the most appropriate template with a sequence identity of no greater than 50% was selected. SSL7 was modelled against SSL11 (PDB 2RDH) (Chung, et al. 2007). FcεRII was modelled against the structure of DC-SIGN complexed to LNFP-III (PDB 1SL5) (Guo, et al. 2004). TRIM21 was modelled against the PRYSPRY domain of the human pyrin b30.2 (PDB 2WL1) (Weinert, et al. 2009). Minor manual adjustments to alignments generated were made to ensure that gaps were not present in regions predicted to have helix/sheet secondary structure. Some structural gaps were not closed during model building; these were closed by inserting bonds and optimizing the region using the Structure Sculpting tool in Maestro. IgA Fc, IgE Fc and IgG Fc models were prepared using each of the other Ig-Fc structures as templates. Fc dimers were generated by overlaying the modeled monomers to the template used, then checking the dimer interface for atomic overlaps and clashes. All intermolecular steric clashes below 2.0 Å in length (as given by the Protein Reports tool) were fixed using the Structure Sculpting tool, followed by Prime Minimization of the entire structure. Modeled proteins were numbered to be consistent with the relevant reference structures and then used for docking. Quality metrics for the models are provided in Supplementary Table S2.

Optimizing application of Fc-based distance restraints

The results obtained from docking with homology models were used to optimize the application of Fc-based distance restraints. This modelling scenario represents what would be undertaken to predict the majority of the unknown Fc-protein interactions. The pyDockRST module of pyDock was used to identify the number of Fc-based distance restraints satisfied by each pose in each of the docking runs

(Chelliah, et al. 2006). Although this procedure was designed to be used in conjunction with the pyDock scoring function (Cheng, et al. 2007), it is convenient to use for this purpose as it is capable of generating protein-ligand complexes "on-the-fly" from the list of rotations and translations converted from ZDOCK. For pyDockRST, the user provides a list of restraining residues, that is, residues from either or both of the ligand and receptor that are expected to be at the protein-protein interface. Distance restraints are deemed to be satisfied if any part of the ligand/receptor occurs within 6.0 Å of the restraining residue. In this study, only distance restraints from the Fc region are used, which, by definition (see Methods section *Protein-protein docking*), is the receptor; thus, restraints are satisfied when any part of the protein binding to Fc (which, by definition, is a ligand) comes within 6.0 Å of a restraining residue.

Since it is not possible for poses to satisfy all of the available restraints at once, nor do the poses satisfying the greatest number of the restraints strictly correspond to the correct poses, the number of restraints applied to filter the poses was optimized. This was done by selecting poses from the initial set generated by ZDOCK that satisfied at least a given number of restraints, starting from one and incrementally increasing the number of restraints applied. For each increment in the number of restraints applied, it was determined for what percentage of cases a successful pose (i.e., the success rate) was found within specific ranges of top ranking poses; the top 1, 2, 5, 10, 20, 50, 100, 200, 500 and 1000 poses were considered.

Optimization of scoring

The results obtained from docking with homology models and filtered by the optimized number of restraints were re-ranked using a wide range of scoring functions suitable for investigating protein-protein interactions (Moal, et al. 2013). The naming of the functions follows that previously published in their evaluation against the general protein-protein docking benchmark available at the time (Moal,

et al. 2013) and as used in the CCharPPI server (Moal, et al. 2015). The performance of the scoring functions was assessed by determining the success rates within specific ranges of top ranking poses, in a similar fashion to the optimization of the restraints. Additionally, the area under the success rate plots up to the top 100 ranked poses was computed as a secondary evaluation of the scoring functions. This area provides a single-value metric in assessing pose prediction performance, and is useful for both identifying scoring functions that are particularly good at consistently ranking successful poses more highly and for providing greater discrimination among scoring functions that give identical success rates within given numbers of ranked poses. The area calculated was expressed as a percentage of the maximum possible area, for simpler comparison.

To further optimize the ranking of successful solutions, top ranked poses from multiple scoring functions were pursued. The top 5, 10, 15, 20 and 25 poses from all possible combinations of three different scoring functions were assessed to find combinations of scoring functions that could identify successful poses in the majority of the cases. Combined sets of poses providing the highest success rate using the smallest number of poses were desired. The set of poses selected by the optimal combination of scoring functions was then re-ranked according to all of the other functions. This was performed to identify a single scoring function to use for ranking the selected poses, as well as to further improve the ranking performance. The best function for the re-ranking was selected by determining the area under the success rate plots up to the top 50 poses. The re-ranked poses were then clustered using a 3.0 Å l_RMSD threshold, with the best ranking representative of the cluster (i.e., lowest energy according to the final scoring function used for re-ranking) reported.

The optimized docking and scoring protocol was verified by application to the cross-docking and unbound docking cases, which were not included in the optimization. The full set of IgA and IgE cases was examined here, however, a subset of IgG cases were examined (described in Supplementary Material).

Results

Identification of common Fc regions involved in interactions

The Fc residues involved in protein recognition in each of the Fc-protein complexes were highlighted on a sequence alignment, revealing regions of general importance for protein recognition by the variety of Fc isotypes (Figure 2). Following the IMGT definition of strands, loops and turns within C domains (Lefranc, et al. 2005), the locations involved in recognition loosely correspond to the CH2 AB turn, the CH2 E strand, the CH2 F strand and the CH3 FG loop. However, for the purpose of defining key Fc regions involved in protein recognition, the IMGT regions are not perfect. There are residues in some complexes that fall just outside the definitions of the CH2 AB turn and the CH3 FG loop, suggesting that some adjustments to these regions could improve their ability to describe Fc-protein recognition. The involvement of residues from the end of the CH2 E strand and the beginning of the CH2 F strand suggests that a new region that better describes protein recognition could be defined with residues from both of these. Based on these observations, three new regions, designated Region I, Region II and Region III, were defined (Figure 2). Region I includes the residues from the CH2 AB turn, as well as those two residues before and three residues after the turn (IMGT numbering 14-15, 15.1-15.3, 16-18). Region II covers a continuous sequence of residues in the CH2 E and CH2 F strands involved in recognition in at least one of the crystal structures (IMGT numbering 90-101), with the exception of Trp96. This residue is conserved across Fc isotypes in both the CH2 and CH3 domains and is occluded from protein binding, suggesting it is of importance in folding these domains. Region III includes all residues from the CH3 FG loop, excluding the first two residues, as well as the first residue of the CH3 G strand (IMGT numbering 107-118). There are some residues involved in recognition in some complexes just outside the definition of Region III. However, this definition was chosen as positions 106 and 119 are never involved in recognition, making them convenient limits for the region.

12

The three regions identified typically include a total of approximately 30 residues. Since these regions capture the majority of contacts in the known Fc-protein complexes, these regions provide a general set of restraints to be applied when studying binding of proteins to the CH2-CH3/CH3-CH4 interface of Fc. The regions are highlighted on the structure of the Fc in Figure 3.

Sampling accuracy in rigid body Fc-protein docking

The sampling accuracy for rigid body docking by ZDOCK for each of the docking scenarios is summarized in Figure 4a. In cognate bound docking, ZDOCK could sample poses with very high accuracy, with an i_RMSD of ~1.0 Å observed for all cases (Table S3).

Solutions satisfying the primary quality metric are sampled in the majority of cross-docking and unbound docking cases (Tables S4-S6). The most challenging cases in this scenario occur when attempting to reproduce the complex of Protein G with IgG-Fc. In particular, utilizing the native structure of Protein G (Frericks Schmidt, et al. 2007) often resulted in the failure to sample any poses satisfying the primary quality metric, although poses satisfying the secondary quality metric could be identified. The difficulty in sampling Protein G on the IgG-Fc structures is most likely a result of the small contact area between Protein G and IgG-Fc, as well as side-chain variations between alternative structural forms of Protein G and IgG-Fc. Utilizing IgG-Fc structures bearing alternative glycoforms (Crispin, et al. 2009), as well as those featuring proteins bound at alternative sites (Duquerroy, et al. 2007; Mizushima, et al. 2011), both of which would be expected to influence the global IgG-Fc conformation, generally does not appear to affect ligand sampling accuracy.

Solutions satisfying the primary quality metric are also sampled in the majority of docking simulations where one or both of the protein structures have been produced by homology modelling (Tables S7-S9). Across all of the cases, sampling was generally more successful when an experimentally determined structure for the ligand was used in the docking. This is expected, since a

ligand structure generated by homology modelling will not be identical to an experimentally determined structure. The most challenging cases were those attempting to reproduce the complex of FcεRII with IgE-Fc. In addition to the effect on sampling quality introduced by using a ligand structure obtained from homology modelling, the quality of the best sampled FcεRII pose also decreased as the angle between the CH3 and CH4 domains of the Fc increased; that is, moving from the relatively acute bend of IgE (on average, approximately 80°) to the more obtuse bends of IgA (~100°) and IgG (~105°) resulted in decreased sampling quality. Altering the Fc bend does not appear to affect the sampling quality in reproducing either of the IgA Fc-SSL7 complex or the IgG-Fc-TRIM21 complex.

Although the sampling in ZDOCK generally performs quite well (Figure 4a), the scoring generally performs quite poorly (Figure 4b), with a success rate of around 20% within the top 20 ranked poses obtained for both the cross/unbound docking and the docking with homology models scenarios. This indicates the need for alternative/additional scoring procedures in identifying the likely binding mode.

Distance restraints based on known binding regions improve modeling of Fc-protein complexes
The effect of applying a gradually increasing number of restraints to filter the poses was examined. The optimization was only applied to the most challenging and arguably most relevant case of docking with homology models. The results are summarised in Figure 5a, and detailed in Table S10.

Using the restraining residues as filters, significant improvements are made in the success rates within the top 1000 ranked poses. Filtering to poses satisfying at least one of the restraints improves the success rate within the top 1000 ranked poses from nine (~40%) to thirteen (~60%) out of the 21 cases. As the number of restraints applied increases to five, the success rate within the top 1000 ranked poses further increases to 18 out of the 21 cases (~85%). When the number of restraints applied is increased to eleven, just over one third of the total number of restraints provided, solutions satisfying either the

14

primary or secondary quality metrics are identified for all cases within the top 1000 ranked poses. In the majority of cases, poses satisfying the primary quality metric are retained. Additionally, some modest improvement in the earlier rankings is observed, with the success rate of the top 50 ranked poses now equaling that observed within the top 1000 ranked poses without filtering (~40%). Increasing the number of restraints applied beyond eleven eliminates all solutions satisfying both the primary and secondary quality metrics in some cases (not shown). In general, between 1000-4000 poses are retained following filtering. Lower numbers of poses are retained for proteins where the interaction is less strictly focused on the Fc regions defined (most notably here, FcεRII).

Scoring optimization

The top 1000 poses from the homology model-based docking experiments satisfying eleven or more of the Fc-based restraints were rescored using a large selection of scoring functions. Compared to the initial ZDOCK results, the majority of scoring functions provide at least some improvement in the success rate (Table S11). This is expected as the poses initially generated by ZDOCK have not been filtered, whereas the poses analyzed by the scoring functions have been filtered to a very specific set. No single scoring function is able to improve the ranking of poses satisfying either of primary or secondary quality metrics beyond the top 500 poses in all cases. However, several scoring functions are able to improve the docking success rate quite significantly within the top 20 ranked poses. The best performing individual scoring functions are AP_T2, AP_T1, CP_ES3DC_MIN and CP_ZS3DC_MIN (Figure 5b), affording success rates within the top 20 ranked poses of 40-60%.

Since different scoring functions may be more or less suitable for studying particular cases, selections of top ranked poses were made from multiple functions in order to improve the success rates further. Several combinations of poses were identified that could provide a success rate of 85% within a selection of 45 or more poses (Table S12). Within the range of poses examined, three combinations

could achieve success rates of 90%; these were combinations of AP_T2 (Tobi 2010) and NSC (Mitra and Pal 2010) with either CP_ES3DC_MIN, CP_EPAIR_MIN or CP_ZS3DC_MIN (Feliu, at al. 2011). The AP_T2/NSC/CP_ZS3DC_MIN combination requires a minimum of 70 poses (25 from AP_T2, 20 from NSC and 25 from CP_ZS3DC_MIN) in order to achieve a 90% success rate, whereas the combination of AP_T2 and NSC with either CP_ES3DC_MIN or CP_EPAIR_MIN requires a minimum of 60 poses (20 from AP_T2, 20 from NSC and 20 from either CP_ES3DC_MIN or CP_EPAIR_MIN). As CP_EPAIR_MIN has lower rates of success within the higher rankings compared to CP_ES3DC_MIN (see Table S11), the combination of AP_T2, NSC and CP_ES3DC_MIN was preferred. Within the range of poses selected by each scoring function, AP_T2 identifies successful solutions for 12 out of the 21 cases (~60%), NSC for a further four cases (~20%) and CP_ES3DC_MIN for a further three cases (~15%). Thus, successful solutions are able to be identified for nearly all of the cases – 19 out of 21 – by selecting a relatively small number of poses from multiple scoring functions. The two cases for which no poses satisfying either of the quality metrics are able to be identified are generally challenging for all of the scoring functions examined. There are also some examples where successful solutions are identified by two of the scoring functions. Successful solutions are identified by both AP_T2 and CP_ES3DC_MIN in six cases (~30%), while successful solutions are identified by both AP_T2 and NSC in just one case. There are no cases for which successful solutions are identified by all three scoring functions, nor is there any overlap in solutions identified by CP_ES3DC_MIN and NSC.

For each of the test cases, the 60 poses selected by the three scoring functions were reranked according to all of the other functions, in order to identify a single function by which to rank the selected poses. Five scoring functions provide standardized areas under their success rate plot greater than 60%, and all of these functions perform similarly over the range examined. The CP_ES3DC_CB function was chosen as it afforded the highest success rate within the top 10 poses (Table S13). At a

practical level, this scoring function was also chosen as it is computed at the same time as CP_ES3DC_MIN, thus reducing the number of computations needed to be undertaken.

Following the reranking, the poses were clustered, which typically reduced the set of poses selected by 25-50% and provided a slight improvement in the success rate (data not shown). The resulting protocol provides a successful solution for 16 out of the 21 cases (~75%) within the top 20 selected poses (Figure 6a), representing a significant improvement over the initial unrestrained docking, for which the success rate within the top 20 poses was approximately 20%. In the majority of cases, there is at least one pose identified that satisfies the primary quality metric (Table 2). The final protocol is summarized in Figure 6b, and examples of poses selected by the procedure for the optimization cases are shown in Figure 7.

Application of protocol to cross-docking and unbound docking cases

The protocol was then applied to a set of cases against which it had not been optimized – the cross-docking and unbound docking cases. This set included complexes for which no similar analogues were present in the training set. The results are summarized in Figure 8, and detailed in Table 3. In unrestrained rigid body ZDOCK docking, a similar level of performance is observed for these cases compared to the docking with homology models scenario. Filtering the ZDOCK docking runs using the Fc-based restraints markedly improves the ranking of successful poses for the majority of cases, particularly within the top 20 ranked poses, where successful poses are identified in 11 of the 19 cases (~60%). Rescoring using the optimized scoring strategy affords a further improvement; successful poses are identified within the top 20 ranked poses in 15 of the 19 cases (~80%). Of particular note are large improvements in the ranking of successful solutions in each of the IgE-FcεRII cases. The rescoring procedure is able to bring successful solutions for all three of these cases to the top 10 ranked poses, whereas prior to rescoring, most of these fall well outside the top 20 ranked poses. In both

docking predictions involving the native FcαRI to IgA-Fc, as well as the docking of Protein G to the native IgG-Fc (PDB 1FC1), successful solutions were unable to be selected following rescoring. The problems with these cases reiterate those observed during the optimization for complexes with small contact areas, as well as those featuring structural rearrangements of protein loops upon binding. Nonetheless, the ability of the docking protocol to find poses for the majority of the IgG-Protein G pairs evaluated is particularly encouraging, and suggestive of the likely success in unknown cases. Selected structures are shown in Figure 9.

Discussion

This study has identified several points of consideration for docking unknown cases to Fc. The first is that there appears to be a marked impact of the bend between the CH2-CH3/CH3-CH4 interface on the sampling accuracy, but only when moving from a more acute bend to a more obtuse bend. This is of particular importance for docking to IgM-Fc, where the structural data relating to this is of low resolution (furthermore, no structures of monomeric IgM-Fc are currently available). The observation that sampling accuracy is adversely affected when moving from an acute to an obtuse bend (e.g., to dock FcεRII to an IgE-Fc modelled in an IgG-like conformation) but not the other way around (e.g., to dock TRIM21 to an IgG-Fc modelled in an IgE-like conformation) suggests that, for docking, it may be preferable to model the bend of IgM-Fc as acutely as allowed by the experimental data. It is likely that a more acute bend permits greater exposure of amino acids at the interface, thus facilitating good quality outcomes in docking. The second is that it appears to be more difficult to identify good quality poses for smaller proteins compared to larger proteins. This is presumably related to the greater contact area in the interface that a larger protein will make with Fc compared to a smaller protein. Furthermore, inaccurate conformations of interface side-chains are likely to have a greater impact on the binding mode predictions of small proteins compared to larger ones, since in the latter case, a small number of

incorrectly modelled specific interactions might be overcome by other interactions in an expectedly larger interface. However, larger proteins are not without their own challenges, and as observed in docking the two Fc receptors, FcεRII and FcαRI, the conformation of the loops involved in binding can significantly affect the quality of the binding modes predicted. The affinity of larger proteins may also have an impact on the success of the docking. The structures of proteins exhibiting high affinity for Fc, SSL7 and TRIM21 (with affinities in the low nanomolar range), were generally well predicted, while the complexes of the lower affinity Fc receptors (with affinities in the low micromolar range) were more challenging to predict. For lower affinity interactions, it may be desirable to incorporate ligand-based knowledge into the docking protocol, if available, by applying a ligand-based filter prior to rescoring, in addition to the receptor-based filter already employed. For instance, knowing that the binding domain of Fcα/μR is likely to adopt an immunoglobulin V-like fold (Shibuya, et al. 2000), recognition of Fc by this receptor is likely to involve the CDR-like regions of this domain, which could be used as a ligand-based filter in docking; mutagenesis studies have also verified the importance of these regions for Fc recognition (Yang, et al. 2013).

This study has reiterated one of the key observations in the docking literature, in that there is as yet no scoring function that can accurately reproduce the free energy of binding (Kastritis and Bonvin 2012; Yuriev, et al. 2011). This has several implications. One is that it is pertinent to consider poses obtained from multiple scoring functions, particularly those exhibiting complementary strengths. The three scoring functions selected here include an atomic contact potential (AP_T2), a coarse-grained statistical potential (CP_ES3DC_MIN) and a measure of surface complementarity (NSC). An atomic contact potential considers details at the atomic level, and is likely to have the greatest success when applied to interactions involving minimal conformational change in the interacting partners, as well as accurate homology models. A coarse-grained potential is likely to be less sensitive to structural variations of the interacting partners, and may therefore be more suitable to predict more flexible

19

interactions, as well as those involving less accurately modelled proteins. Incorporating poses from both of these types of potentials is likely to afford at least one accurate pose for the majority of docking situations, as we have observed here. Another implication of the lack of a perfect scoring function is that the final decision as to which pose within the set of potential poses represents the true or closest binding mode remains non-trivial; some knowledge of the binding interface of the interacting proteins is also necessary to make this decision. Previous docking studies on the interaction of proteins and peptides with Fc have typically utilized knowledge of specific residues, either inferred from crystallographic structures or derived from biochemical studies, in predicting the complexes with Fc (Karaca and Bonvin 2011; Wang, et al. 2014; Yang, et al. 2010; Zacharias and Schneider 2012). We have taken a similar approach in this study, identifying restraints by considering the Fc regions most frequently contacted in known Fc-protein complexes. However, rather than consider specific residues from the regions, we have used entire segments from the Fc domains, selecting poses in proximity to a certain number of the residues from the regions – but not any *specific* residues – in the initial round of filtering. This type of filter has an advantage in that knowledge of specific residues involved in recognition is not strictly required. We have shown that this type of filter can significantly reduce the set of poses for further processing with minimal loss of accuracy, while being adequately non-prescriptive that it can be applied to study the full range of Fc-protein recognition scenarios. Ligand-based knowledge may be valuable in determining the true binding mode from the set of possibilities.

It is important to emphasise that in this study that we have only considered rigid body placement, and have not performed further refinement of solutions. Since refinement procedures typically take into account some degree of flexibility of the interacting partners, they can introduce considerable computational expense, and thus, their use is generally restricted to a limited number of poses, rather than in the initial search for potential solutions as has been the primary focus of this work. The optimized docking and scoring protocol demonstrated here provides a high success rate within the

top 20 ranked poses, which would be a sufficiently small number to consider for refinement. Techniques such as rigid body optimization (Mirzaei, et al. 2012), normal modes analysis (Carrington and Mancera 2004; Moal and Bates 2010; Tama, et al. 2000), as well as loop and side-chain resampling (Krivov, et al. 2009; Zhu, et al. 2007), could be applied to refine the poses selected by the protocol described here, as well as more computationally intensive scoring strategies such as MM-GB(PB)/SA (Kollman, et al. 2000; Ryde and Genheden 2015; Tajne, et al. 2012; Xu, et al. 2013), and ideally, would further improve the ranking of good quality solutions.

Conclusions

We have developed a framework utilizing rigid body docking and rescoring for predicting protein binding to the CH2-CH3/CH3-CH4 interface of Fc based on consideration of the range of available moderate to high resolution structures. The most notable performance is observed within the top 20 poses, for which the success rate was 75-80% for the systems evaluated, representing a significant improvement over the initial unrestrained docking, for which a success rate of only 20% within the top 20 poses was obtained. The protocol will be valuable for predicting protein binding to this site on all immunoglobulin isotypes, and may be useful in the design of Fc-binding peptides.

List of abbreviations

Fc – crystallizable fragment; IgG – immunoglobulin G; IgA – immunoglobulin A; IgM – immunoglobulin M; IgE – immunoglobulin E; TRIM21 – tripartite motif-containing protein 21; PDB – Protein Data Bank; RMSD – root-mean-squared deviation; i_RMSD – interface RMSD; l_RMSD; ligand RMSD

Authors' contributions

MA designed and carried out the experiments and prepared the manuscript. RLM, PAR and JFR

assisted in drafting and critically revising the manuscript. All authors read and approved the final

manuscript.

Declarations

None of the authors have any competing interests.

References

Berman HM, Westbrook J, Feng Z, Gililand G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. 2000. The Protein Data Bank. *Nucleic Acids Res.*, **28**: 235-242.
Bidgood SR, Tam JCH, McEwan WA, Mallery DL, James LC. 2014. Translocalized IgA mediates neutralization and stimulates innate immunity inside infected cells. *Proc. Natl. Acad. Sci. U. S. A.*, **111**: 13463-13468.
Carrington BJ, Mancera RL. 2004. Comparative estimate of vibrational entropy changes in proteins through normal mode analysis. *J. Mol. Graph. Model.*, **23**: 167-174.
Chelliah V, Blundell TL, Fernández-Recio J. 2006. Efficient restraints for protein-protein docking by comparison of observed amino acid substitution patterns with those predicted from local environment. *J. Mol. Biol.*, **357**: 1669-1682.
Chen K, Cerutti A. 2011. The function and regulation of immunoglobulin D. *Curr. Opin. Immunol.*, **23**: 345-352.
Chen R, Li L, Weng Z. 2003. ZDOCK: an initial-stage protein docking algorithm. *Proteins*, **52**: 80-87.
Cheng TM-K, Blundell TL, Fernandez-Recio J. 2007. pyDock: electrostatics and desolvation for effective scoring of rigid-body protein-protein docking. *Proteins*, **68**: 503-515.
Chung MC, Wines BD, Baker H, Langley RJ, Baker EN, Fraser JD. 2007. The crystal structure of staphylococcal superantigen-like protein 11 in complex with sialyl Lewis X reveals the mechanism for cell binding and immune inhibition. *Mol. Microbiol.*, **66**: 1342-1355.
Coyne RS, Siebrecht M, Peitsch MC, Casanova JE. 1994. Mutational analysis of polymeric

immunoglobulin receptor/ligand interactions. Evidence for the involvement of multiple complementarity determining region (CDR)-like loops in receptor domain I. *J. Biol. Chem.*, **269**: 31620-31625.

Crispin M, Bowden TA, Coles CH, Harlos K, Aricescu AR, Harvey DJ, Stuart DI, Jones EY. 2009. Carbohydrate and domain architecture of an immature antibody glycoform exhibiting enhanced effector functions. *J. Mol. Biol.*, **387**: 1061-1066.

Cruse JM, Lewis RE. 2010. *Atlas of Immunology, 3rd ed.* CRC Press.

Czajkowsky DM, Salanti A, Ditlev SB, Shao Z, Ghumra A, Rowe JA, Pleass RJ. 2010. IgM, Fc mu Rs, and malarial immune evasion. *J. Immunol.*, **184**: 4597-4603.

Czajkowsky DM, Shao Z. 2009. The human IgM pentamer is a mushroom-shaped molecule with a flexural bias. *Proc. Natl. Acad. Sci. U. S. A.*, **106**: 14960-14965.

Deisenhofer J. 1981. Crystallographic refinement and atomic models of a human Fc fragment and its complex with fragment B of protein A from Staphylococcus aureus at 2.9- and 2.8-A resolution. *Biochemistry*, **20**: 2361-2370.

Dhaliwal B, Yuan D, Pang MOY, Henry AJ, Cain K, Oxbrow A, Fabiane SM, Beavil AJ, McDonnell JM, Gould HJ, Sutton BJ. 2012. Crystal structure of IgE bound to its B-cell receptor CD23 reveals a mechanism of reciprocal allosteric inhibition with high affinity receptor FcεRI. *Proc. Natl. Acad. Sci. U. S. A.*, **109**: 12686-12691.

Duquerroy S, Stura EA, Bressanelli S, Fabiane SM, Vaney MC, Beale D, Hamon M, Casali P, Rey FA, Sutton BJ, Taussig MJ. 2007. Crystal structure of a human autoimmune complex between IgM rheumatoid factor RF61 and IgG1 Fc reveals a novel epitope and evidence for affinity maturation. *J. Mol. Biol.*, **368**: 1321-1331.

Ehrenmann F, Kaas Q, Lefranc M-P. 2010. IMGT/3Dstructure-DB and IMGT/DomainGapAlign: a database and a tool for immunoglobulins or antibodies, T cell receptors, MHC, IgSF and MhcSF. *Nucleic Acids Res.*, **38**: D301-307.

Frericks Schmidt HL, Sperling LJ, Gao YG, Wylie BJ, Boettcher JM, Wilson SR, Rienstra CM. 2007. Crystal polymorphism of protein GB1 examined by solid-state NMR spectroscopy and X-ray diffraction. *J. Phys. Chem. B*, **111**: 14362-14369.

Gabb HA, Jackson RM, Sternberg MJ. 1997. Modelling protein docking using shape complementarity, electrostatics and biochemical information. *J. Mol. Biol.*, **272**: 106-120.

Gautam S, Loh K-C. 2013. Human pIgR mimetic peptidic ligand for affinity purification of IgM. Part II: Ligand binding characteristics. *Sep. Purif. Technol.*, **102**: 43-49.

Ghumra A, Semblat J-P, McIntosh RS, Raza A, Rasmussen IB, Braathen R, Johansen F-E, Sandlie I, Mongini PK, Rowe JA, Pleass RJ. 2008. Identification of residues in the Cμ4 domain of polymeric IgM essential for interaction with Plasmodium falciparum erythrocyte membrane protein 1 (PfEMP1). *J. Immunol.*, **181**: 1988-2000.

Ghumra A, Shi J, Mcintosh RS, Rasmussen IB, Braathen R, Johansen F-E, Sandlie I, Mongini PK, Areschoug T, Lindahl G, Lewis MJ, Woof JM, Pleass RJ. 2009. Structural requirements for the interaction of human IgM and IgA with the human Fcα/μ receptor. *Eur. J. Immunol.*, **39**: 1147-1156.

Guo Y, Feinberg H, Conroy E, Mitchell DA, Alvarez R, Blixt O, Taylor ME, Weis WI, Drickamer K. 2004. Structural basis for distinct ligand-binding and targeting properties of the receptors DC-SIGN and DC-SIGNR. *Nat. Struct. Mol. Biol.*, **11**: 591-598.

Hatanaka T, Ohzono S, Park M, Sakamoto K, Tsukamoto S, Sugita R, Ishitobi H, Mori T, Ito O, Sorajo K, Sugimura K, Ham S, Ito Y. 2012. Human IgA-binding peptides selected from random peptide-libraries: affinity maturation and application in IgA purification. *J. Biol. Chem.*, **287**: 43126-43136.

Herr AB, Ballister ER, Bjorkman PJ. 2003. Insights into IgA-mediated immune responses from the crystal structures of human FcalphaRI and its complex with IgA1-Fc. *Nature*, **423**: 614-620.

Hober S, Nord K, Linhult M. 2007. Protein A chromatography for antibody purification. *J. Chromatogr.*

B, **848**: 40-47.

Holdom MD, Davies AM, Nettleship JE, Bagby SC, Dhaliwal B, Girardi E, Hunt J, Gould HJ, Beavil AJ, McDonnell JM, Owens RJ, Sutton BJ. 2011. Conformational changes in IgE contribute to its uniquely slow dissociation rate from receptor FcεRI. *Nat. Struct. Mol. Biol.*, **18**: 571-576.

Huang S-Y. 2015. Exploring the potential of global protein-protein docking: an overview and critical assessment of current programs for automatic *ab initio* docking. *Drug Discov. Today*, **20**: 969-977.

Hwang H, Vreven T, Janin J, Weng Z. 2010. Protein-protein docking benchmark version 4.0. *Proteins*, **78**: 3111-3114.

Idusogie EE, Presta LG, Gazzano-Santoro H, Totpal K, Wong PY, Ultsch M, Meng YG, Mulkerrin MG. 2000. Mapping of the C1q binding site on rituxan, a chimeric antibody with a human IgG1 Fc. *J. Immunol.*, **164**: 4178-4184.

James LC, Keeble AH, Khan Z, Rhodes DA, Trowsdale J. 2007. Structural basis for PRYSPRY-mediated tripartite motif (TRIM) protein function. *Proc. Natl. Acad. Sci. U. S. A.*, **104**: 6200-6205.

Jerlström PG, Talay SR, Valentin-Weigand P, Timmis KN, Chhatwal GS. 1996. Identification of an immunoglobulin A binding motif located in the beta-antigen of the c protein complex of group B streptococci. *Infect. Immun.*, **64**: 2787-2793.

Johnsson E, Areschoug T, Mestecky J, Lindahl G. 1999. An IgA-binding peptide derived from a streptococcal surface protein. *J. Biol. Chem.*, **274**: 14521-14524.

Karaca E, Bonvin AMJJ. 2011. A multidomain flexible docking approach to deal with large conformational changes in the modeling of biomolecular complexes. *Structure*, **19**: 555-565.

Kastritis PL, Bonvin AMJJ. 2012. On the binding affinity of macromolecular interactions: daring to ask why proteins interact. *J. R. Soc. Interface*, **10**: 20120835.

Khunrae P, Philip JMD, Bull DR, Higgins MK. 2009. Structural comparison of two CSPG-binding DBL domains from the VAR2CSA protein important in malaria during pregnancy. *J. Mol. Biol.*, **393**: 202-213.

Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, Lee M, Lee T, Duan Y, Wang W, Donini O, Cieplak P, Srinivasan J, Case DA, Cheatham TE. 2000. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.*, **33**: 889-897.

Krivov GG, Shapovalov MV, Dunbrack RL. 2009. Improved prediction of protein side-chain conformations with SCWRL4. *Proteins*, **77**: 778-795.

Laskowski RA, Swindells MB. 2011. LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. *J. Chem. Inf. Model.*, **51**: 2778-2786.

Lefranc M-P, Pommié C, Kaas Q, Duprat E, Bosc N, Giraudou D, Jean C, Ruiz M, Da Piédade I, Rouard M, Foulquier E, Thouvenin V, Lefranc G. 2005. IMGT unique numbering for immunoglobulin and T cell receptor constant domains and Ig superfamily C-like domains. *Dev. Comp. Immunol.*, **29**: 185-203.

Lensink MF, Wodak SJ. 2009. Docking and scoring protein interactions: CAPRI 2009. *Proteins*, **78**: 3073-3084.

Macpherson AJ, Mccoy KD, Johansen F-E, Brandtzaeg P. 2008. The immune geography of IgA induction and function. *Mucosal Immunol.*, **1**: 11-22.

Mallery DL, McEwan WA, Bidgood SR, Towers GJ, Johnson CM, James LC. 2010. Antibodies mediate intracellular immunity through tripartite motif-containing 21 (TRIM21). *Proc. Natl. Acad. Sci. U. S. A.*, **107**: 19985-19990.

Martin WL, West AP, Gan L, Bjorkman PJ. 2001. Crystal structure at 2.8 A of an FcRn/heterodimeric Fc complex: mechanism of pH-dependent binding. *Mol. Cell.*, **7**: 867-877.

Mimoto F, Katada H, Kadono S, Igawa T, Kuramochi T, Muraoka M, Wada Y, Haraya K, Miyazaki T, Hattori K. 2013. Engineered antibody Fc variant with selectively enhanced FcγRIIb binding over both

FcγRIIaR131 and FcγRIIaH131. *Protein Eng. Des. Sel.*, **26**: 589-598.

Mirzaei H, Beglov D, Paschalidis IC, Vajda S, Vakili P, Kozakov D. 2012. Rigid body energy minimization on manifolds for molecular docking. *J Chem Theory Comput*, **8**: 4374-4380.

Mizushima T, Yagi H, Takemoto E, Shibata-Koyama M, Isoda Y, Iida S, Masuda K, Satoh M, Kato K. 2011. Structural basis for improved efficacy of therapeutic antibodies on defucosylation of their Fc glycans. *Genes Cells*, **16**: 1071-1080.

Moal IH, Bates PA. 2010. SwarmDock and the use of normal modes in protein-protein docking. *Int. J. Mol. Sci.*, **11**: 3623-3648.

Moal IH, Jiménez-García B, Fernandez-Recio J. 2015. CCharPPI web server: computational characterisation of protein-protein interactions from structure. *Bioinformatics*, **31**: 123-125.

Moal IH, Torchala M, Bates PA, Fernández-Recio J. 2013. The scoring of poses in protein-protein docking: current capabilities and future directions. *BMC Bioinformatics*, **14**: 286.

Müller R, Gräwert MA, Kern T, Madl T, Peschek J, Sattler M, Groll M, Buchner J. 2013. High-resolution structures of the IgM Fc domains reveal principles of its hexamer formation. *Proc. Natl. Acad. Sci. U. S. A.*, **110**: 10183-10188.

Pierce BG, Hourai Y, Weng Z. 2011. Accelerating protein docking in ZDOCK using an advanced 3D convolution library. *PLoS One*, **6**: e24657.

Platts-Mills TAE, Woodfolk JA. 2011. Allergens and their role in the allergic immune response. *Immunol. Rev.*, **242**: 51-68.

Ramsland PA, Farrugia W, Bradford TM, Sardjono CT, Esparon S, Trist HM, Powell MS, Tan PS, Cendron AC, Wines BD, Scott AM, Hogarth PM. 2011. Structural basis for FcγRIIa recognition of human IgG and formation of inflammatory signaling complexes. *J. Immunol.*, **187**: 3208-3217.

Ramsland PA, Willoughby N, Trist HM, Farrugia W, Hogarth PM, Fraser JD, Wines BD. 2007. Structural basis for evasion of IgA immunity by Staphylococcus aureus revealed in the complex of SSL7 with Fc of human IgA1. *Proc. Natl. Acad. Sci. U. S. A.*, **104**: 15051-15056.

Rasti N, Namusoke F, Chene A, Chen Q, Staalsoe T, Klinkert MQ, Mirembe F, Kironde F, Wahlgren M. 2006. Nonimmune immunoglobulin binding and multiple adhesion characterize *Plasmodium falciparum*-infected erythrocytes of placental origin. *Proc. Natl. Acad. Sci. U. S. A.*, **103**: 13795-13800.

Røe M, Norderhaug IN, Brandtzaeg P, Johansen F-E. 1999. Fine specificity of ligand-binding domain 1 in the polymeric Ig receptor: importance of the CDR2-containing region for IgM interaction. *J. Immunol.*, **162**: 6046-6052.

Ryde U, Genheden S. 2015. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin. Drug Discov.*, **10**: 449-461.

Sandin C, Linse S, Areschoug T, Woof JM, Reinholdt J, Lindahl G. 2002. Isolation and detection of human IgA using a streptococcal IgA-binding peptide. *J. Immunol.*, **169**: 1357-1364.

Sauer-Eriksson AE, Kleywegt GJ, Uhlén M, Jones TA. 1995. Crystal structure of the C2 fragment of streptococcal protein G in complex with the Fc domain of human IgG. *Structure*, **3**: 265-278.

Schindler CE, de Vries SJ, Zacharias M. 2015. iATTRACT: simultaneous global and local interface optimization for protein-protein docking refinement. *Proteins*, **83**: 248-258.

Semblat J-P, Ghumra A, Czajkowsky DM, Wallis R, Mitchell DA, Raza A, Rowe JA. 2015. Identification of the minimal binding region of a *Plasmodium falciparum* IgM binding PfEMP1 domain. *Mol. Biochem. Parasitol.*, **201**: 76-82.

Shibuya A, Sakamoto N, Shimizu Y, Shibuya K, Osawa M, Hiroyama T, Eyre HJ, Sutherland GR, Endo Y, Fujita T, Miyabayashi T, Sakano S, Tsuji T, Nakayama E, Phillips JH, Lanier LL, Nakauchi H. 2000. Fc α/μ receptor mediates endocytosis of IgM-coated microbes. *Nat. Immunol.*, **1**: 441-446.

Sondermann P, Huber R, Oosthuizen V, Jacob U. 2000. The 3.2-Å crystal structure of the human IgG1 Fc fragment-FcγRIII complex. *Nature*, **406**: 267-273.

Tajne S, Sanam R, Gundla R, Gandhi NS, Mancera RL, Boddupally D, Vudem DR, Khareedu VR.

2012. Molecular modeling of Bt Cry1Ac (DI-DII)-ASAL (Allium sativum lectin)-fusion protein and its interaction with aminopeptidase N (APN) receptor of Manduca sexta. *J. Mol. Graph. Model.*, **33**: 61-76.

Tama F, Gadea FX, Marques O, Sanejouand YH. 2000. Building-block approach for determining low-frequency normal modes of macromolecules. *Proteins*, **41**: 1-7.

Wang R-Z, Lin D-Q, Tong H-F, Yao S-J. 2014. Molecular insights into the binding selectivity of a synthetic ligand DAAG to Fc fragment of IgG. *J. Mol. Recognit.*, **27**: 250-259.

Weinert C, Grütter C, Roschitzki-Voser H, Mittl PRE, Grütter MG. 2009. The crystal structure of human pyrin b30.2 domain: implications for mutations associated with familial Mediterranean fever. *J. Mol. Biol.*, **394**: 226-236.

Wines BD, Trist HM, Farrugia W, Ngo C, Trowsdale J, Areschoug T, Lindahl G, Fraser JD, Ramsland PA. 2012. A conserved host and pathogen recognition site on immunoglobulins: structural and functional aspects. *Adv. Exp. Med. Biol.*, **946**: 87-112.

Wurzburg BA, Jardetzky TS. 2009. Conformation flexibility in immunoglobulin E-Fc 3-4 revealed in multiple crystal forms. *J. Mol. Biol.*, **393**: 176-190.

Xu L, Li Y, Sun H, Li D, Hou T. 2013. Structural basis of the interactions between CXCR4 and CXCL12/SDF-1 revealed by theoretical approaches. *Mol. BioSys.*, **9**: 2107-2117.

Yang H, Gurgel PV, Williams DKJ, Bobay BG, Cavanagh J, Muddiman DC, Carbonell RG. 2010. Binding site on human immunoglobulin G for the affinity ligand HWRGWV. *J. Mol. Recognit.*, **23**: 271-282.

Yang X, Zhao Q, Zhu L, Zhang W. 2013. The three complementarity-determining region-like loops in the second extracellular domain of human Fc alpha/mu receptor contribute to its binding of IgA and IgM. *Immunobiology*, **218**: 798-809.

Yuriev E, Agostino M, Ramsland PA. 2011. Challenges and advances in computational docking: 2009 in review. *J. Mol. Recognit.*, **24**: 149-164.

Zacharias M, Schneider S. 2012. Atomic resolution model of the antibody Fc interaction with the complement C1q component. *Mol. Immunol.*, **51**: 66-72.

Zhao WW, Liu FF, Shi QH, Sun Y. 2014a. Octapeptide-based affinity chromatography of human immunoglobulin G: comparisons of three different ligands. *J Chromatogr A*, **1359**: 100-111.

Zhao WW, Shi QH, Sun Y. 2014b. Dual-ligand affinity systems with octapeptide ligands for affinity chromatography of hIgG and monoclonal antibody. *J Chromatogr A*, **1369**: 64-72.

Zhu K, Shirts MR, Friesner RA. 2007. Improved methods for side chain and loop predictions via the Protein Local Optimization Program: variable dielectric model for implicitly improving the treatment of polarization effects. *J Chem Theory Comput*, **3**: 2108-2119.

Tables

Table 1. Systems used to develop model of Fc-protein recognition.[a]

PDB ID	Fc	Interacting protein	Resolution (Å)	Reference
1OW0	IgA1	FcαRI	3.10	(Herr, et al. 2003)
2QEJ	IgA1	*Staphylococcus aureus* SSL7	3.20	(Ramsland, et al. 2007)
1FCC	IgG1	*Streptococcus* Protein G	3.20	(Sauer-Eriksson, et al. 1995)
1L6X	IgG1	*Staphylococcus aureus* Protein A Z34C	1.65	(Idusogie, et al. 2000)
2IWG	IgG1	TRIM21	2.35	(James, et al. 2007)
1I1A	Rat IgG2a	Rat FcRn	2.80	(Martin, et al. 2001)
4EZM	IgE	FcεRII	3.10	(Dhaliwal, et al. 2012)

[a]All proteins are human unless otherwise specified.

Table 2. Best quality solutions selected by optimized protocol in cases involving docking with

homology models (optimization cases).

Fc structure	Protein structure	i_RMSD (Å)	l_RMSD (Å)	Rank
IgA Fc (2QEJ)	SSL7 (modelled)	3.7	6.5	8
IgA Fc (modelled on IgE)	**SSL7 (2QEJ)**	**3.8**	**5.6**	**2**
IgA Fc (modelled on IgE)	**SSL7 (modelled)**	**4.0**	**6.3**	**10**
IgA Fc (modelled on IgE)	**SSL7 (1V1O)**	**2.3**	**2.7**	**3**
IgA Fc (modelled on IgG)	SSL7 (2QEJ)	3.7	7.2	17
IgA Fc (modelled on IgG)	SSL7 (modelled)	4.0	7.3	16
IgA Fc (modelled on IgG)	**SSL7 (1V1O)**	**2.6**	**3.6**	**3**
IgE Fc (2WQR)	*FcεRII (modelled)*	*12.1*	*13.6*	*29*
IgE Fc (4EZM)	FcεRII (modelled)	3.2	8.0	20
IgE Fc (modelled on IgA)	**FcεRII (4EZM)**	**3.8**	**4.3**	**16**
IgE Fc (modelled on IgA)	*FcεRII (modelled)*	*12.0*	*18.0*	*26*
IgE Fc (modelled on IgA)	FcεRII (2H2R)	3.8	6.3	20
IgE Fc (modelled on IgG)	FcεRII (4EZM)	10.5	9.0	10
IgE Fc (modelled on IgG)	FcεRII (modelled)	6.9	8.0	49
IgE Fc (modelled on IgG)	FcεRII (2H2R)	6.1	9.1	6
IgG Fc (2IWG)	TRIM21 (modelled)	3.8	4.5	22
IgG Fc (2WAH)	TRIM21 (modelled)	3.4	4.2	34
IgG Fc (modelled on IgA)	TRIM21 (2IWG)	2.5	2.9	5
IgG Fc (modelled on IgA)	TRIM21 (modelled)	3.7	4.6	3
IgG Fc (modelled on IgE)	**TRIM21 (2IWG)**	**3.9**	**4.7**	**7**
IgG Fc (modelled on IgE)	TRIM21 (modelled)	4.0	5.1	1

[a]Italics denote cases for which solutions satisfying neither of the quality metrics were not identified; the

solution with the lowest i_RMSD is shown instead. Bold indicates cases for which multiple solutions

satisfying the primary quality metric are present in the final set of poses, the best ranking of which is

shown here. In all other cases, a single solution satisfying either of the quality metrics is present in the

final set of poses.

Table 3. Best quality solutions in application of optimized docking protocol to cross and unbound docking cases.[a]

Fc structure	Protein structure	i_RMSD (Å)	l_RMSD (Å)	Rank
IgA Fc (1OW0)	FcαRI (1UCT)	5.1	9.6	20
IgA Fc (1OW0)	SSL7 (1V1O)	1.6	1.9	5
IgA Fc (1OW0)	SSL7 (2QEJ)	2.6	3.6	19
IgA Fc (2QEJ)	FcαRI (1OW0)	1.3	3.2	3
IgA Fc (2QEJ)	*FcαRI (1UCT)*	*6.6*	*16.6*	*17*
IgA Fc (2QEJ)	SSL7 (1V1O)	2.7	3.7	4
IgE Fc (2WQR)	FcεRII (2H2R)	3.7	7.3	4
IgE Fc (2WQR)	FcεRII (4EZM)	3.2	3.0	2
IgE Fc (4EZM)	FcεRII (2H2R)	3.1	7.0	8
IgG Fc (1FC1)	Protein G (1FCC)	7.9	7.8	5
IgG Fc (1FC1)	TRIM21 (2IWG)	2.1	2.5	4
IgG Fc (1FC1)	*Protein G (2QMT)*	*10.1*	*10.1*	*2*
IgG Fc (1FCC)	TRIM21 (2IWG)	3.5	5.7	1
IgG Fc (1FCC)	**Protein G (2QMT)**	**1.9**	**2.0**	**2**
IgG Fc (2IWG)	Protein G (1FCC)	1.7	1.8	1
IgG Fc (2IWG)	**Protein G (2QMT)**	**3.8**	**4.5**	**1**
IgG Fc (2WAH)	**Protein G (1FCC)**	**2.5**	**2.7**	**13**
IgG Fc (2WAH)	TRIM21 (2IWG)	2.3	3.3	7
IgG Fc (2WAH)	**Protein G (2QMT)**	**2.9**	**2.9**	**5**

[a]Italics denote cases for which solutions satisfying neither of the quality metrics were not identified; the solution with the lowest i_RMSD is shown instead. Bold indicates cases for which a single solution satisfying the primary quality metric is present in the final set of poses. In all other cases, multiple solutions satisfying either of the quality metrics are present in the final set of poses, and the best ranking of these is shown here.

Figure legends

Figure 1. Schematic and structural representation protein binding sites on Fc, illustrating the structural and functional similarity of the CH2-CH3 domains of IgG, IgA and IgD with the CH3-CH4 domains of IgE and IgM. A. Schematic representation of a single heavy chain of IgG, IgA and IgD. B. Overlay of the complexes of IgG-Fc with FcRn (PDB 4N0U) and FcγRIIIb (PDB 1E4K). C. Schematic representation of a single heavy chain of IgE and IgM. D. Overlay of the complexes of IgE-Fc with FcεRII (PDB 4EZM) and FcεRII (PDB 2Y7Q). Legend: dark blue – CH3 domain (panels A and B), CH4 domain (panels C and D); light blue – CH2 domain (panels A and B), CH3 domain (panels C and D); light green – hinge-CH2 binding site (panels A and B), CH2-CH3 binding site (panels C and D); red – CH2-CH3 binding site (panels A and B), CH3-CH4 binding site (panels C and D).

Figure 2. Proteins utilize a common set of regions to bind to the CH2-CH3/CH3-CH4 interface of Fc. Sequences for the CH2 (a) and CH3 (b) domains for the Fc-protein complexes. For the complex of FcεRII with IgE-Fc (4EZM), the sequences of the CH3 and CH4 domains are shown in panels a) and b), respectively. Residues involved in interactions from the DIMPLOT analysis are highlighted in bold. Region I (red; CH2 positions 14-15, 15.1-15.3, 16-18), Region II (yellow; CH2 positions 90-101) and Region III (blue; CH3 positions 107-118) are each highlighted with coloured boxes. The following residues, not involved in interactions, are also highlighted: Cys23, Trp41, Leu89, Cys104. Numbering follows the IMGT unique numbering for C domains, generated using the IMGT/DomainGapAlign tool.

Figure 3. Key Fc regions for protein recognition at the CH2-CH3/CH3-CH4 interface. a) IgG-Fc. b) IgA-Fc. c) IgE-Fc. In all panels, Region I is shown in red, Region II in yellow and Region III in blue.

Figure 4. Sampling and scoring performance of ZDOCK. a) Box plots of the sampling performance by ZDOCK in the various docking scenarios. Dashed line indicates the i_RMSD threshold corresponding to the primary quality metric (4.0 Å). Solid squares in each plot indicate the mean i_RMSD for the best quality pose sampled across that docking scenario. Boxes cover the interquartile range, with a band

indicating the median. Whiskers extend to the maximum and minimum values. b) ZDOCK ranking performance for cognate (yellow), cross/unbound docking (red) and docking with homology models (blue).

Figure 5. Optimization of the number of restraints and scoring functions applied to docking. a) Comparison of ranking performance within the top 1000 poses applying no restraints (ZDOCK ranking; blue) to pose ensembles satisfying at least one restraint (red), five restraints (yellow) and eleven restraints (green). Logarithmic scale is used. b) Comparison of ranking performance within the top 100 poses applying no restraints (ZDOCK ranking; dark blue), eleven restraints (green) and the best performing scoring functions identified from the set, applied to the filtered docking results: AP_T2 (red), AP_T1 (yellow), CP_ES3DC_MIN (maroon), CP_ZS3DC_MIN (light blue). Linear scale is used.

Figure 6. Incorporating poses from multiple potentials further improves docking outcomes. a) Comparison of ranking performance within the top 100 poses of the optimized docking strategy (light blue), unfiltered ZDOCK results (dark blue), ZDOCK results filtered to those satisfying eleven or more restraints (green) and the following scoring functions, applied to the filtered docking results: AP_T2 (red), CP_ES3DC_MIN (maroon), NSC (yellow). Linear scale is used. As only up to 60 poses are selected in the optimized docking strategy, the success rate for this is only shown for up to 60 poses. b) Flowchart of the optimized docking protocol. Proteins are docked to Fc using dense sampling in ZDOCK, following by filtering using the Fc-based restraints (Figure 1). The top 1000 poses following the Fc-based filtering are rescored using a variety of scoring functions. The top 20 poses from the AP_T2, CP_ES3DC_MIN and NSC potentials are selected. The 60 poses selected are then reranked using the CP_ES3DC_CB potential, following clustering. This results in a set of typically 30-45 poses, with at least one pose satisfying the primary quality criterion likely to exist in the top 20 poses. Success rates within the top 20 poses for the various scoring, filtering and reranking steps are shown in bold

red.

Figure 7. Examples of the best quality poses selected by the docking procedure in the optimization cases (docking with at least one homology model). In all panels, the reference crystal structure is shown in grey, the Fc structure used in the docking in red and the ligand protein structure used in the docking shown in blue. a) Predicted binding mode of SSL7 (modelled on SSL11) with IgA-Fc (modelled on IgG-Fc) (i_RMSD = 4.0 Å, l_RMSD = 7.3 Å). b) Predicted binding mode of FcεRII (native structure) with IgE-Fc (modelled on IgA-Fc) (i_RMSD = 3.8 Å, l_RMSD = 6.3 Å). c) Predicted binding mode of TRIM21 (modelled on pyrin b30.2) with IgG-Fc (modelled on IgA-Fc) (i_RMSD = 2.5 Å, l_RMSD = 2.9 Å).

Figure 8. The optimized docking strategy provides comparable performance for both the optimization cases (docking with at least one homology model) and the cross/unbound docking cases. Success rate plots comparing the ranking performance of ZDOCK with no restraints (dark blue), ZDOCK applying eleven restraints (green) and following the optimized rescoring strategy (light blue) for both the cross and unbound docking cases (solid lines) and the optimization cases (dashed lines).

Figure 9. Examples of the best quality poses selected by the docking procedure in the cross and unbound docking cases. In all panels, the reference crystal structure is shown in grey, the Fc structure used in the docking in red and the ligand protein structure used in the docking shown in blue. a) Predicted binding mode of FcαRI (unbound structure) with IgA-Fc (from complex with SSL7) (i_RMSD = 5.1 Å, l_RMSD = 9.6 Å). b) Predicted binding mode of FcεRII (native structure) with IgE-Fc (native structure) (i_RMSD = 3.7 Å, l_RMSD = 7.3 Å). c) Predicted binding mode of TRIM21 (from complex with IgG-Fc) with IgG-Fc (native structure) (i_RMSD = 2.1 Å, l_RMSD = 2.5 Å). d) Predicted binding mode of Protein G (native structure) with IgG-Fc (from complex with TRIM21) (i_RMSD = 3.8 Å, l_RMSD = 4.5 Å).
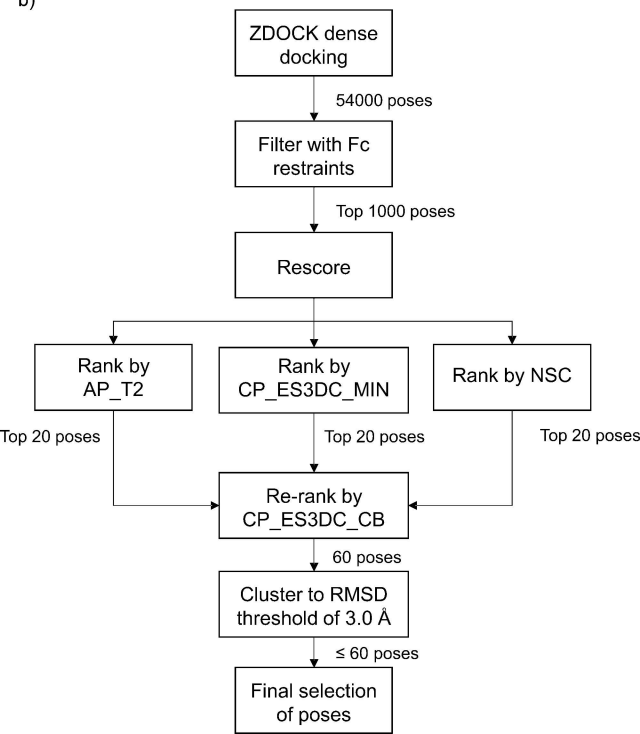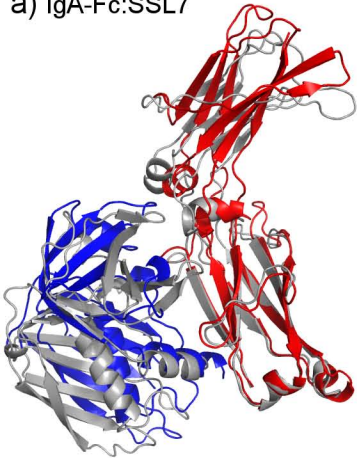
a)

VH
CH1
CH2
CH3

b)

c)

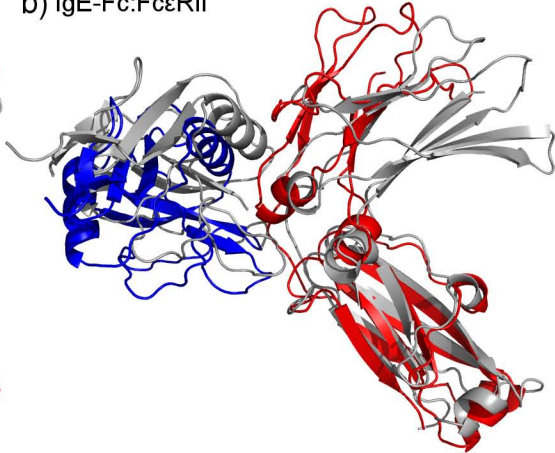VH
CH1
CH2
CH3
CH4

d)

a)

b)
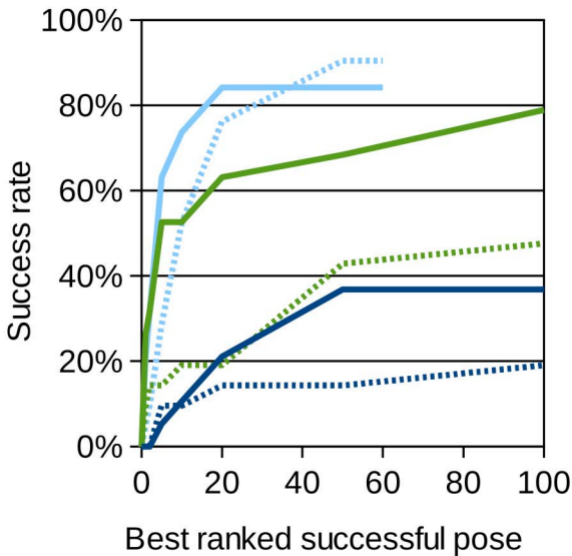
a)

b)

a)

b)

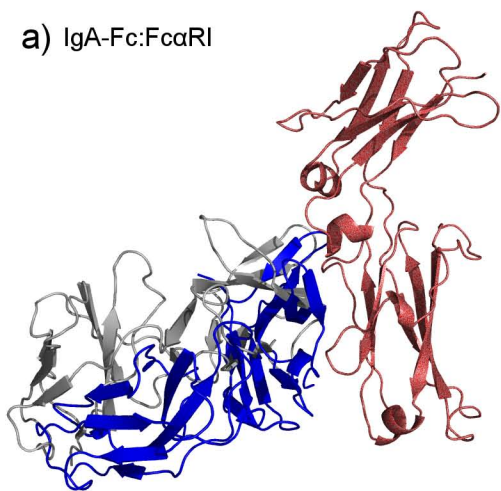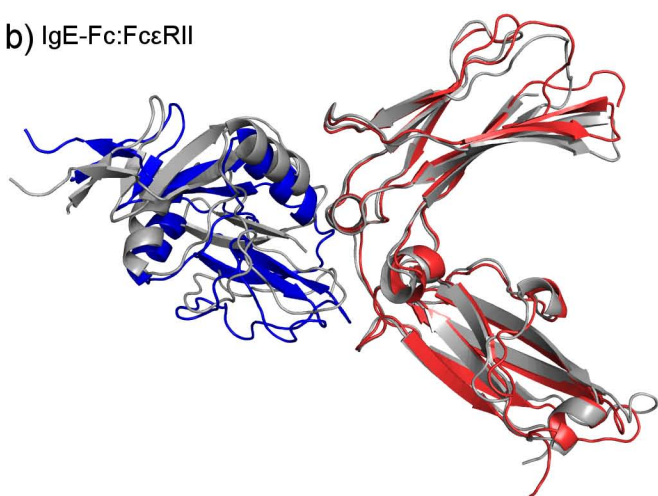a) IgA-Fc:SSL7   b) IgE-Fc:FcεRII   c) IgG-Fc:TRIM21

a) IgA-Fc:FcαRI

b) IgE-Fc:FcεRII

c) IgG-Fc:TRIM21

d) IgG-Fc:Protein G