Computer-aided Binding Mode Prediction and Affinity Maturation of LRR Protein Binder without Structural Determination

Yoonjoo Choi[1,†], Sukyo Jeong[1,†], Jung-Min Choi[1], Christian Ndong[2], Chris Bailey-Kellogg[3], Karl E. Griswold[2,4,5], and Hak-Sung Kim[1,*]

[1]Department of Biological Sciences, Korea Advanced Institute of Science and Technology, Daejeon, 34141, Korea.

[2]Thayer School of Engineering, Dartmouth College, Hanover, NH 03755, USA.

[3]Department of Computer Science, Dartmouth College, Hanover, NH 03755, USA.

[4]Norris Cotton Cancer Center at Dartmouth, Lebanon, NH 03766, USA.

[5]Department of Biological Sciences, Dartmouth College, Hanover, NH 03755, USA.

† These authors contributed equally

Correspondence

Hak-Sung Kim

Department of Biological Sciences, Korea Advanced Institute of Science and Technology, Daejeon, 34141, Korea.

Tel: +82-42-350-2616

Email: hskim76@kaist.ac.kr

Keywords

Protein binding mode; Computational protein design; LRR protein; Epitope mapping; Binding affinity

Abstract

Binding affinity maturation without structure determination remains a difficult challenge in the computer-aided protein engineering. Precise binding mode identification is a vital prerequisite for the affinity maturation. However, pure computational methods have been unreliable in practice so far and experimental structural biology techniques are generally too costly. Herein, we show that computational epitope localization followed by the full-atom energy minimization with intermediate experimental validation can yield precisely bound complex model structure, which ultimately enables effective affinity maturation and redesign of binding specificity. As a proof-of-concept, we targeted a leucine-rich repeat (LRR) protein binder which specifically binds to the human IgG_1 ($hIgG_1$). Based on the computationally predicted binding mode of the LRR protein binder to $hIgG_1$, the binding affinity of the protein binder was significantly increased and its specificity was redesigned toward multiple IgGs from other species. Experimental determination of the complex structure showed that the predicted model closely matched the X-ray crystal structure. Through the benchmark of therapeutically relevant existing LRR protein complexes, we demonstrate that the present approach can be broadly applicable to other proteins which undergo small conformational changes upon binding.

Significance Statement

Computer-aided design of binding affinity and specificity of two interacting proteins without structural determination is a challenging problem in protein engineering. Despite recent advances in the computational biology techniques, however, *in silico* evaluation of binding energies has proven to be extremely difficult. We show that, in the cases of protein-protein interactions where only small structural changes upon binding occur, partial experimental validation of binding can greatly complement the computational energy. Using an LRR (leucine-rich repeat) protein binder as a model system, the binding orientation of two binding partners was precisely dentified by the hybrid approach. Based on the predicted model, we could successfully redesign the binding affinity and specificity of the protein binder by the full-atom energy calculation.

Introduction

Protein-protein interactions are central to all biological processes, including signal transduction, cellular regulation, and immune response (1). Determining the binding mode of such interactions thus provides fundamental insight into the underlying mechanisms for the control and regulation of biological processes. In the development of both therapeutic proteins and vaccines, because the precise identification of the binding mode is particularly important, because the efficacy varies depending on their epitopes (2-4) and the effectiveness of vaccines are affected by their epitopes (5-7). Although the experimental determination of the crystal structures has been used as the gold standard for binding mode identification (8), it is extremely labor-intensive and time-consuming. Therefore, considerable effort has been made to develop computational methods as alternatives over the past decades (9, 10). One of the practical applications of the binding mode prediction is an affinity maturation of a protein binder from scratch without structure determination, which has been regarded as one of the "Holy Grails" in the computationally-aided protein engineering (11). However, computational approaches have been insufficiently reliable in practice so far for predicting the binding mode (12, 13) and the association energies (11). Last rounds of the Critical Assessment of Predicted Interactions (CAPRI) have also shown that current computational methods are not successful in identifying the actual binding mode which is necessary for affinity maturation of a protein binder from scratch (14-16).

Leucine-rich repeat (LRR) proteins have a rigid horseshoe-like structural feature and play key roles in many biological processes (17) including the immune system (18-21) and cellular processes (22-24). Leucine-rich repeat (LRR) proteins constitute one of the most common protein families found in a wide range of species, and more than 2,000 LRR proteins have been identified (25, 26). Typical examples include mammalian toll-like receptors (TLR) and the variable

79 lymphocyte receptors (VLR) of jawless vertebrates in innate and adaptive immune systems,
80 respectively (19, 21). Considering the importance and abundance of LRR proteins in nature, a
81 broadly enabling strategy for the control of LRR binding is of great significance for accelerating
82 the understanding of their functions as well as advancing the potential applicability to other
83 therapeutic protein binders.

84 Herein, we demonstrate computer-driven binding mode prodeiction and affinity maturation
85 of a leucine-rich repeat (LRR) protein binder without structural determination. As a proof of
86 concept, we focued on a LRR protein binder which specifically binds to an immunoglobulin G
87 (IgG). IgG is a challenging model system to show the redesign of binding affinity and specificity,
88 since it is structurally highly conserved, but diverse in sequence among species. We show that a
89 computationally-guided approach can effectively narrow down where the binding interface is, and
90 the full-atom energy minimization can identify a native-like complex model which closely
91 matched the X-ray crystal structure. Further energy calculations on the predicted model complex
92 enabled the redesign of the binding affinity and specificity of the LRR protein binder.

93

94 ## Results

95 **Binding mode prediction of an LRR protein binder**

96 In an effort to exploit the structural and functional features of LRR proteins for biotechnological
97 and medical applications, we previously developed an LRR protein binder, called "repebody (Rb)"
98 (27). As a proof of concept, in this work, a human IgG_1 ($hIgG_1$)-binding repebody (RbF4) (28)
99 was targeted for the binding mode prediction and affinity maturation without the structure
100 determination. There is indirect evidence that RbF4 binds to the constant region of $hIgG_1$ (hFc)
101 (28, 29), but actual epitopes and the binding mode were completely unknown. RbF4 has a typical

102 LRR protein sequence motif, whose structural scaffold consists of three major parts: 1) an N
103 terminal cap (LRRNT), 2) LRR modules (LRRVs), and 3) a C terminal cap (LRRCT) with an
104 additional loop (**Fig. S1**). Unlike the complementary determining regions of antibodies, the target
105 binding sites of a repebody comprise parallel beta strands (LRRVs), which are assumed to remain
106 unchanged upon target-binding. During the development of RbF4, three variable residues on each
107 module of LRRV2–LRRV5 (LRRV4 excluded) were randomized and subject to a phage display
108 selection(28).

109 To predict the binding mode of RbF4-hFc, we first localized the RbF4 epitope on hFc using
110 EpiScope, a computational method that provides optimal sets of disruptive (triplet) mutations
111 against possible docking models (12). For antibody-antigen pairs, on average of three mutation
112 sets are sufficient to cover all of the docking models, and at least one set of mutations is expected
113 to overlap the true epitopes (epitope localization). It should be noted that epitope localization only
114 provides a rough idea of where the protein would bind and does not provide the binding mode.
115 ClusPro webserver (30) was employed to dock the RbF4-hFc pair, as demonstrated in the previous
116 study (12). We used a crystal structure of the unbound form of hFc (PDB code: 3AVE) and a
117 homology model of RbF4 for the docking (**Table 1**), assigning attractions at the concave residues
118 of LRRV modules from 2 to 5, as known during the phage display selection. We assumed that the
119 major driving force of RbF4 for target binding should be no different from antibodies or high
120 affinity protein binders (Antibody Mode) (31). The antibody docking mode was enabled to
121 generate target-bound complex structures, yielding 29 models in total (**Table 1**). Variants with the
122 triplet mutation sets were then experimentally verified by measuring their binding affinity using
123 isothermal titration calorimetry (ITC).

124 Any mutations are symmetrically present due to the symmetric nature of the Fc domain
125 structure. Considering the symmetry, three mutation sets (thus, six in total) in contact with all of
126 the generated docking models were identified (Var 1, 2, and 3 in **Fig. 1A**). Among the three designs
127 (with triple mutations each), Var 3 (H75A, N80K, and H200K) was shown to be approximately 3-
128 times more disruptive to RbF4 compared to the wild-type hFc, resulting in a decrease in K_d from
129 128 nM to 427 nM (**Fig. 1B** and **Fig. S2**). This result confirms that RbF4 indeed binds to hFc.
130 Although the hFc interface region in contact with RbF4 was roughly identified through the triplet,
131 it is probable that not all of them are involved in the binding. We therefore measured the binding
132 affinities of the single variants. Among the triplet, two single mutations (H75A and N80A) led to
133 meaningful reductions (two-fold decreases) in the binding affinities (194 and 187 nM respectively),
134 whereas the binding affinity of H200K remained similar (or stronger) to that of the wild-type hFc.
135 From the binding tests, it is likely that RbF4 establishes contacts with H75 and N80, but unlikely
136 with H200. There were seven RbF4-hFc docking models in contact with the two residues (H75
137 and N80), but not with H200 or the six other positions in Var 1 and 2 (**Fig. 1C**).

138 We hypothesized that the all-atom force field energy could largely capture the binding free
139 energy landscape, and consequently, the model with the lowest force field energy should be near-
140 native if any such models exist. Based on this hypothesis, the seven docking models were ranked
141 in terms of the AMBER99sb force field parameter (32) after full-atom minimization using the
142 Tinker molecular dynamics package (33). The determination of the X-ray crystal structure (PDB
143 ID: 6KA7) confirms that the binding mode of the docking model with the lowest energy (**Fig. 1C**
144 and **D**, blue model) and the crystal structure are extremely similar (f_{nat}: 0.43 and I-RMSD: 2.32 Å,
145 see **Fig. 2**). These results demonstrate that the computational prediction with intermediate
146 experimental validation enabled the binding mode prediction for LRR proteins. The computational

147 method suggests an optimal set of mutations to test based on the docking models. The experimental
148 verification using binding assays then effectively narrows down the plausible computational
149 docking model candidates, and the full-atom minimization assists in the ranking of the filtered
150 docking models, leading to the identification of a native-like docking model. It is noteworthy that
151 the prior epitope localization is an essential step for precise binding mode prediction. A ranking
152 assessment using the force field energy alone may be insufficient for finding a native-like docking
153 model (**Fig. 2B** and **C**).

154

155 **Redesign of binding specificity based on the predicted binding mode**

156 Based on the predicted binding mode, we redesigned RbF4 to mitigate its binding specificity. RbF4
157 was identified as highly specific for human IgG_1, while showing weak and negligible cross-
158 reactivities against mouse IgG_1 and rabbit IgG (28, 29). The predicted binding mode reveals that
159 the loop (**Fig. S1**) may be largely responsible for the binding specificity of RbF4 toward hFc (**Fig.**
160 **1D**). To obtain insight into the specificity, we investigated the Fc sequences of IgGs from three
161 species: human (hFc), mouse (mFc), and rabbit (rFc). Calculation of the inter-molecular atomic
162 distances between RbF4 and hFc based on the predicted binding mode showed that the RbF4 loop
163 forms a tight contact with the positions where hFc, mFc, and rFc share different amino acids from
164 each other (**Fig. 1D**), strongly implying its crucial role in the binding specificity of RbF4. We thus
165 reasoned that the engineering of the loop would lead RbF4 to a cross-species Fc binder. To prove
166 our hypothesis, we truncated and replaced the loop sequence of RbF4 with flexible amino acids:
167 RNSAGSVA starting at 239 was changed into GG. We measured the binding affinities of the
168 resulting loop-truncated RbF4 variant (RbF4-LT) for the IgGs from the three species through ITC
169 assays: $hIgG_1$ (Trastuzumab), mouse IgG_1 ($mIgG_1$), and rabbit IgG (rIgG).

170 As shown in our previous work (29), the original RbF4 strongly binds to $hIgG_1$ with a
171 binding affinity of 128.7 nM, and weakly to $mIgG_1$ with an 8-fold weaker affinity than $hIgG_1$ (1
172 μM), whereas it has a negligible binding affinity for rIgG (**Fig. 3** and **Fig. S3**). The loop-truncated
173 variant (RbF4-LT) gained a binding affinity for rIgG (1.2 μM), indicating that the loop is indeed
174 involved in the binding specificity of RbF4 for $hIgG_1$. The binding affinity of the variant for $mIgG_1$
175 was also improved to 775 nM, whereas that for $hIgG_1$ decreased to a similar level to $mIgG_1$ (598
176 nM). It bears noting that the variant was entirely designed based on the predicted binding mode
177 prior to the determination of the X-ray co-crystal structure. The results indicate that the predicted
178 binding mode is sufficiently accurate for further engineering of the binding specificity.

179

180 **Improvement of binding affinity based on the model binding complex**

181 Our final goal is to improve the binding affinity of the LRR binder through a computer-aided
182 design. As shown in the binding mode prediction, although the full-atom force field energy alone
183 does not discriminate the native complex, it can largely capture the binding energy landscape.
184 Herein, we used the AMBER force-field energy to select mutations that can simultaneously
185 improve the binding affinities of RbF4 against multiple targets. To reduce the search space, FoldX
186 (34) was employed to scan possible mutations in the loop (**Fig. S4**). The predicted binding energy
187 at G243 indicates that the inclusion of the loop may not enhance a binding affinity for rIgG. Thus
188 we aimed to design an RbF4 variant which can bind to both $hIgG_1$ and $mIgG_1$ with high binding
189 affinities. The FoldX scan suggests that S241M may substantially enhance the binding affinities
190 for $hIgG_1$ and $mIgG_1$. The mutation was also observed during the phage display affinity maturation
191 for the $mIgG_1$-specific repebody (29). We then fixed S241M and introduced all other amino acids
192 *in silico* at S244. The AMBER99sb force field energy was used to minimize the variants. The

193 binding energy prediction indicates that S241M with S244R (RbF4-MR) may significantly
194 improves the binding affinities for both IgGs (**Fig. S5**). The ITC binding assy results show that
195 RbF4-MR indeed strongly binds to the two IgGs as predicted, and the binding affinity for $mIgG_1$
196 was markedly increased (1 μM to 168.1 nM, **Fig. 3**).

197

198 **General applicability of the binding mode prediction**

199 To assess the general applicability of the binding mode prediction, we tested three known Rb-
200 targets whose co-crystal structures are available (**Table 1**): Interleukin-6 (IL-6: 4J4L) (27),
201 epidermal growth factor receptor (EGFR: 4UIP) (35) and complement component 5a (C5a: 5B4P)
202 (36). We firstly investigated the detailed settings of the protein docking. Docking without the
203 antibody mode option using ClusPro resulted in a larger number of docking models, but the overall
204 accuracy of such docking models proved to be worse than those generated with the option enabled.

205 As demonstrated in RbF4-hFc, only a small number of variants were sufficient to localize the
206 epitopes (two for C5a and EGFR, and four for IL-6). The filtering process resulted in a subset of
207 "filtered" docking models which include native-like binding modes (5 to 12 models; solid circles
208 in **Fig. 4**). In order to identify the most native-like docking model, the ClusPro score was initially
209 considered to rank the filtered docking models. However, the ClusPro score was found to be
210 unreliable for ranking (**Fig. S6**). **Figure 4** shows that ranking based on the AMBER force field
211 energy successfully discriminates high-quality docking models for IL-6 and EGFR. It bears noting
212 that the prior epitope localization is again critical. Ranking by the force field energy alone is not
213 sufficient to find native-like docking models. For example, in the case of IL-6, there are two

214 docking models with lower energies than the most native-like model, but both of them are not in
215 contact with true epitopes.

216 The C5a case offers another insight into precise binding mode prediction (**Fig. 4**, bottom row).
217 While the binding interfaces were mostly correct (15 out of 18 correct interface residues: 83 %),
218 the predicted binding mode was completely inverted (N terminal to C, and vice versa). During the
219 affinity maturation of the C5a-specific Rb, it was observed that some LRRV modules gave rise to
220 a negligible increase in the binding affinity (36), which suggests that only LRRV1 and 2 are
221 responsible for interacting with C5a. We thus hypothesized that the accuracy of structural
222 modeling and an incorporation of the paratope information may also enhance the docking quality.
223 The four possible combinations of hypotheses (incorporation of the phage display information
224 versus no incorporation, and a homology-model target versus a crystal structure target) reveal that
225 the use of a high-quality structure (here the crystal structure) is not sufficient for accurate binding
226 mode prediction (**Fig. 5**). The imprecise definition of Rb paratopes with the crystal structure
227 resulted in worse correlations than the model with precisely annotated paratopes (Spearman ρ for
228 f_{nat} and I-RMSD of crystal structures: -0.1 and 0.1, and those of models with precise paratope
229 definition: -0.7 and 0.19, respectively). The ideal case (crystal structures with precise paratopes)
230 led to extremely accurate results (I-RMSD: 0.62Å and f_{nat}: 0.76) with high correlations (Spearman
231 ρ for f_{nat} and I-RMSD: -0.79 and 0.71).

232 This study demonstrated critical criteria to be satisfied for accurate binding mode prediction.
233 The full-atom force field energy is a good indicator to discriminate the most native-like docking
234 model, but it alone may not be sufficient if the initial epitope localization process with
235 experimental verification is omitted. As also shown in the previous study with antibodies (12),
236 high-quality antigen structures are not particularly required for epitope localization in Rb-target

237 pairs. However, they become necessary to generate and predict native-like docking modes. Though
238 an Rb has predefined target-binding sites, not all of them are involved in target-binding. The
239 inclusion of paratope information (e.g. phage display selection) was shown to improve the quality
240 of docking models and binding mode prediction.

241

242 **Discussion**

243 Prediction of accurate protein binding mode is an essential prerequisite for protein binding affinity
244 maturation without structural determination. It can certainly accelerlate our understanding of
245 protein functions, advancing the potential utility of relevant proteins in basic and translational
246 research. We chose LRR proteins for the binding mode prediction since, as a promising therapeutic
247 protein binder, they have an advantageous property in engineering persepctive because they
248 undergo small structural changes upon binding. To our best knowledge, prediction of protein-
249 protein binding modes has not been consistently achieved yet. As a proof of concept, we began
250 with the binding mode prediction of the LRR protein binder which targets the constant domain of
251 human IgG_1 (hFc). The AMBER force field energy with full-atom minimization after epitope
252 localization enabled the successful selection of the most native-like docking model. The X-ray
253 crystal structure of the LRR protein complex with hFc is indeed extremely close to the predicted
254 binding model with the lowest energy.

255 Unlike epitope localization or binding site identification, which do not particularly require
256 extremely native-like binding modes in the sampling step (12), the inclusion of native-like samples
257 (i.e., docking models) is absolutely necessary for affinity improvement from scratch. In general,
258 the sampling quality of the antibody-antigen docking often depends on the targets; however, in our

259 case, native-like models were nearly always included in the initial samples. It is likely that the
260 rigid nature of LRR domains enables the successful sampling of high-quality docking models.
261 Therefore, we anticipate that this approach may be in principle applicable not only to LRR domains,
262 but also to any other proteins with rigid binding sites. As also proven in the success of interface-
263 guided docking methods (37-39), the exact definition of binding interfaces including paratope
264 mutational data from phage display selection is also critical. As repeatedly shown in our study,
265 selection only based on the force-field energy value may be misled to wrong complex binding
266 modes, perhaps due to the inaccuracy of the energy functions (40). The mutagenesis-based
267 verification and filtering were shown to be a compulsory process in order to complement imperfect
268 energy scores. The computational method used in this study, EpiScope (12), provided the optimal
269 choice of mutations. For the binding mode prediction, six sets of *in vitro* experiments were
270 performed, which were shown to be sufficient to effectively localize the epitope. With the epitope
271 localization, the full-atom minimization enabled ranking of filtered docking models, resulting in
272 identification of a native-like docking model.

273 The success of the binding mode prediction directly leads to the design of binding
274 specificities and affinities. From the predicted model, we were able to identify which region
275 assigns the binding specificity of RbF4 to the IgGs. The simple manipulation of the binding region
276 entirely changed its binding specificity. The force field energy is indicative of the binding energy.
277 Mutations selected based on the energy dramatically improves binding affinities as predicted.
278 However, it bears noting that design only based on the energy may result in wrong predictions.
279 For example, the force field energy prediction tells that RbF4-MR would bind to $hIgG_1$ with a
280 strong binding affinity as well (**Fig. S6**) though the actual binding affinity stays the same as before.
281 From the two studies of the binding mode prediction and affinity improvement, the force field

energy prediction becomes informative only if actual binding is known, i.e. supportive data from experiments are critical in practice.

Acknowledgements

This work was supported by the Global Research Laboratory (NRF-2015K1A1A2033346) and Mid-Career Researcher Program (NRF-2017R1A2A1A05001091) through the National Research Foundation (NRF) of Korea, funded by the Ministry of Science and ICT. Y.C. was supported by the Korea Research Fellowship Program (NRF-2016H1D3A1938246) through the National Research Foundation of Korea, funded by the Ministry of Science and ICT. J.M. was supported by Basic Science Research Program (NRF-2017R1A6A3A04012313) of the National Research Foundation funded by the Ministry of Science and ICT, and the Ministry of Education. The production of the Fc and Fc variant was supported by the National Institute of General Medical Sciences of the U.S. National Institutes of Health under award numbers P20-GM113132 and 2R01GM098977.

Contributions

Y.C. designed the computational methods, and implemented and applied them to generate all of the computational results. S.J. conducted all in vitro experiments. J-M.C. determined the X-ray crystal structure. C.N. generated the Fc variants. Y.C., S.J., J-M.C., C.N., C.B-K., K.E.G., and H-S.K. designed and analyzed all of the prospective and retrospective tests and contributed toward the preparation of the manuscript.

304 **Conflict of Interest**: Chris Bailey-Kellogg and Karl E. Griswold are Dartmouth faculties and co-
305 members of Stealth Biologics, LLC, a Delaware biotechnology company. These authors
306 acknowledge that there may be a potential financial conflict of interest related to their associations
307 with this company, and they hereby affirm that the data presented in this paper are free of any bias.
308 This work has been reviewed and approved as specified in the conflict of interest management
309 plans of Dartmouth.

310

311 ## Methods

312 ### Computational Methods

313 The computational method initially starts with epitope localization. We used EpiScope for
314 localization of the repebody binding sites with the default settings (12): triple mutations per
315 binding patch, and one pareto-optimal curve and two suboptimal ones. The docking models were
316 generated using the ClusPro webserver (30). The webserver provides "Antibody mode" based on
317 statistics between the Ab epitopes and paratopes (31). The non-CDR masking option was disabled,
318 but the binding sites and residues in the convex (**Fig. S1A**) were masked for attraction and
319 repulsion, respectively. All model structures were minimized using the Tinker molecular dynamics
320 package (33) with the AMBER99sb parameter set (32) and the GB/SA implicit solvent model (41).

321 There are three Rb-target complex structures in the PDB up to date: IL-6 (PDB code 4J4L),
322 EGFR (4UIP), and C5a (5B4P) binders. Unbound forms of target structures were used for docking
323 (**Table 1**). There is a missing loop in the structure of IL-6 (1ALU chain A: 52 SSKEALAEN). The
324 loop was filled using MODELLER (42) and all the backbone atoms of the loop were minimized
325 using Tinker as described above. While Rb has a very rigid predefined structure, a single mutation

326 at the 11th LRRV position (**Fig. S1B**) to proline significantly changes structural conformations.
327 Two of the Rbs (4J4L and 5B4P) have such proline residues (at LRRV1 and LRRV2 respectively).
328 The two Rbs were modeled using each other as templates. One LRRV unit of the EGFR Rb (4UIP)
329 was omitted. The trimmed Rb structure was reconstructed by splitting the free Rb at LRRV3 and
330 superimposing the LRRV3 on LRRV2 of the complete Rb using PyMol.

331

332 **Preparation of Fc Model and Variants**

333 The sequence of the human Fc binder repebody (RbF4) was obtained from a previously
334 published study (28). The Rb structure was modeled using MODELLER with a free form (3RFS:A)
335 as a template structure. A free form of the hFc domain (3AVE) was used for docking. A total of
336 29 docking models were generated using ClusPro (**Table 1**).

337 Trastuzumab (trade name, Herceptin) Fc sequence available from the literature (wild-type)
338 and all subsequent variants were reverse translated, codon optimized for expression in mammalian
339 cells, and synthesized by Integrated DNA technologies (IDT), Inc. (Redwood City, CA). CMVR
340 VRC01 expression vector (NIH AIDS reagent program, Germantown, MD) harboring the wild-
341 type Fc or the Fc variant sequences was transfected into suspension HEK 293 cells using
342 polyethylenimine (PEI) (Polysciences, Warrington, PA). Briefly, 500 µg of the wild-type Fc or Fc
343 variant DNA was combined with 1 ml of PEI and incubated at room temperature for 10 minutes.
344 The mixture was then added to HEK cells in the suspension and incubated in a humidified chamber
345 at 37 °C with 8% CO_2 for at least 5 to 6 days. The secreted wild-type Fc or Fc variants were
346 clarified through centrifugation at 8000 rpm at 4 °C for 15 minutes on a Beckman Avanti-J25

347 centrifuge (Brea, CA). The resulting supernatant was filtered through a 0.45 μm filter to remove
348 any residual cell debris and other large particles before loading onto a FPLC column.

349 Affinity purification was conducted on a pre-packed 5 ml Protein A column (for wild-type
350 Fc, and Fc variants 1 and 2) or pre-packed 5 ml Protein G column (for Fc variant 3, and single
351 mutations of the variant) from GE Healthcare (Pittsburgh, PA) as suggested by the manufacturer.
352 The final sample was eluted with 100 mM Glycine at pH 3 in 2 ml Eppendorf tubes prefilled with
353 50 μl of 1 M Tris and 5 mM EDTA. The purification process was automated on an AKTA FPLC
354 system (GE Healthcare, Pittsburgh, PA). The purified protein was subjected to a second buffer
355 exchange step using a hitrap desalting column (GE Healthcare, Pittsburgh, PA). The final product
356 was eluted in phosphate buffer saline and stored at -20 °C until further use. The purified proteins
357 were analyzed under reduced SDS-PAGE conditions and stained with coomassie blue.

358 **Expression and Purification of Repebodies**

359 The gene-encoding repebodies were inserted into NdeI and XhoI restriction sites of a
360 pET21a vector (Invitrogen, Carlsbad, CA). Plasmids were cloned into competent *E.coli* DH5α
361 cells using a heat shock method (at 42 °C for 90 seconds). The recombinant plasmids were
362 transformed into *E.coli* Origami-B cells (Merck, Kenilworth, NJ). Single colonies were inoculated
363 into 5 mL of a Luria-Bertani (LB) medium containing 50 μg/mL carbenicillin and grown overnight
364 at 37 °C in a shaking incubator (200 rpm). A total of 250 mL of LB containing 50 μg/mL
365 carbenicillin was inoculated with an OD_{600} 0.05 volume of the overnight-saturated culture and
366 grown at 37 °C with shaking at 200 rpm until the OD_{600} reached 0.5–0.8. The cells were induced
367 using 0.5 mM IPTG and incubated at 18 °C with shaking at 200 rpm for 16 hours. The cells were
368 harvested through centrifugation at 8000 rpm for 20 mins and suspended in a lysis buffer (50 mM
369 NaH_2PO_4, 300 mM NaCl, and 10 mM imidazole, at pH 8.0). After cell lysis by sonication, the cell

370 debris was removed through centrifugation at 16,000 rpm for 1 hour at 4 °C. Cell lysates were
371 loaded into a Ni-NTA column (Qiagen, Hilden, Germany) and washed using a wash buffer solution
372 (50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole, at pH 8.0). The repebodies were eluted using
373 an elution buffer (50 mM NaH_2PO_4, 300 mM NaCl, 250 mM imidazole, at pH 8.0), and further
374 purified using gel permeation chromatography (Superdex 75, GE Healthcare). The buffer of the
375 purified repebodies changed into PBS, and the concentrate was developed in an AmiconUltra
376 centrifugal filter (10 kDa cutoff, Millipore).

377 **Determination of crystal structure**

378 The Fc domain of human IgG (hFc) was purified after digestion of the purchased human
379 IgG (Sigma, St. Louis, MO) with papain, as described elsewhere (43). The RbF4-hFc complex was
380 purified through gel-filtration with a buffer containing 5 mM Tris·HCl and 0.1 M NaCl (pH 7.4)
381 after reconstitution of the complex by incubating RbF4 and hFc at a 2:1 molar ratio on ice. Crystals
382 of the complex were grown using a hanging drop vapor diffusion method against a crystallization
383 buffer containing 0.1 M sodium acetate, 12% (w/v) polyethylene glycol 6000, and 0.1 M
384 magnesium chloride (pH 4.6) at 20°C. The crystals formed in the space group $P2_12_12_1$ with a =
385 59.9 Å, b = 107.4 Å, and c = 171.4 Å, and contained one complex molecule in an asymmetric unit.
386 The diffraction data were collected at 100 K, with crystals flash-frozen in a crystallization buffer
387 containing 30% glycerol. A single-wavelength (1 Å) dataset was collected using a native crystal
388 on beam line 5C (Pohang Accelerator laboratory, Korea). Integration, scaling, and merging of the
389 diffraction data were conducted using the HKL2000 program suite (44). The initial phases were
390 determined through molecular replacement using the Phenix AutoMR program (45) and human Fc
391 of IgG (PDB accession 1H3X) and repebody (PDB accession 3RFJ) as models. Successive rounds

392 of model building using Coot (46), refinement using the Phenix program (47), and phase
393 combination allowed the building of the complete structure (**Table S1**).

394 **Isothermal Titration Calorimetry (ITC)**

395 A binding affinity experiment was conducted using MicroCal-iTC200 (Malvern
396 Instruments, Malvern, UK). Fc variants, mouse IgG_1, and rabbit IgG (Sigma) were diluted in a
397 PBS buffer to a final concentration of 0.02 mM. RbF4 or RbF4-LT (RbF4 with the loop truncation)
398 was diluted using the same buffer to a final concentration of 0.2 mM. The ITC experiments were
399 performed for a total 20 injections and stirred at 1000 rpm. The initial injection of 1 μL was
400 excluded for data analysis. Titration curves were fitted with a one-site binding model. The value
401 of K_d was determined using Origin (OriginLab).

402

References

1. Alberts B (1998) The cell as a collection of protein machines: preparing the next generation of molecular biologists. *Cell* 92(3):291-294.
2. Gan HK, Burgess AW, Clayton AH, & Scott AM (2012) Targeting of a conformationally exposed, tumor-specific epitope of EGFR as a strategy for cancer therapy. *Cancer Res.* 72(12):2924-2930.
3. Garrett TP*, et al.* (2009) Antibodies specifically targeting a locally misfolded region of tumor associated EGFR. *Proc. Natl. Acad. Sci. U.S.A.* 106(13):5082-5087.
4. Kim KM*, et al.* (2002) Both the epitope specificity and isotype are important in the antitumor effect of monoclonal antibodies against Her-2/neu antigen. *Int. J. Cancer* 102(4):428-434.
5. Khurana S*, et al.* (2016) Human antibody repertoire after VSV-Ebola vaccination identifies novel targets and virus-neutralizing IgM antibodies. *Nat. Med.* 22(12):1439.
6. Lewis GK (2010) Challenges of antibody-mediated protection against HIV-1. *Expert Rev. Vaccines* 9(7):683-687.
7. Zolla-Pazner S (2004) Identifying epitopes of HIV-1 that induce protective antibodies. *Nat. Rev. Immunol.* 4(3):199.
8. Abbott WM, Damschroder MM, & Lowe DC (2014) Current approaches to fine mapping of antigen–antibody interactions. *Immunology* 142(4):526-535.
9. Gao J & Kurgan L (2014) Computational prediction of B cell epitopes from antigen sequences. *Immunoinformatics*, (Springer), pp 197-215.
10. Zhang W*, et al.* (2011) Prediction of conformational B-cell epitopes from 3D structures by random forests with a distance-based feature. *BMC Bioinform.* 12(1):341.
11. Houk K & Liu F (2017) Holy grails for computational organic chemistry and biochemistry. *Acc. Chem. Res.* 50(3):539-543.
12. Hua CK*, et al.* (2017) Computationally-driven identification of antibody epitopes. *eLife* 6:e29023.
13. Sela-Culang I, Ofran Y, & Peters B (2015) Antibody specific epitope prediction—emergence of a new paradigm. *Cur. Opin. Virol.* 11:98-102.
14. Lensink MF & Wodak SJ (2010) Blind predictions of protein interfaces by docking calculations in CAPRI. *Proteins* 78(15):3085-3095.
15. Mata-Fink J*, et al.* (2013) Rapid conformational epitope mapping of anti-gp120 antibodies with a designed mutant panel displayed on yeast. *J. Mol. Biol.* 425(2):444-456.
16. Rosenfeld L*, et al.* (2015) Combinatorial and computational approaches to identify interactions of macrophage colony-stimulating factor (M-CSF) and its receptor c-FMS. *J. Biol. Chem.* 290(43):26180-26193.
17. Grove TZ, Cortajarena AL, & Regan L (2008) Ligand binding by repeat proteins: natural and designed. *Curr. Opin. Struct. Biol.* 18(4):507-515.
18. Akira S, Takeda K, & Kaisho T (2001) Toll-like receptors: critical proteins linking innate and acquired immunity. *Nat. Immunol.* 2(8):675.
19. Medzhitov R (2001) Toll-like receptors and innate immunity. *Nat. Rev. Immunol.* 1(2):135.
20. Pancer Z*, et al.* (2005) Variable lymphocyte receptors in hagfish. *Proc. Natl. Acad. Sci. U.S.A.* 102(26):9224-9229.
21. Sutoh Y & Kasahara M (2016) Lymphocyte Populations in Jawless Vertebrates: Insights Into the Origin and Evolution of Adaptive Immunity. *The Evolution of the Immune System*, (Elsevier), pp 51-67.

449 22. De Wit J, Hong W, Luo L, & Ghosh A (2011) Role of leucine-rich repeat proteins in the
450 development and function of neural circuits. *Annu. Rev. Cell Dev. Biol.* 27:697-729.
451 23. van der Biezen EA & Jones JD (1998) The NB-ARC domain: a novel signalling motif
452 shared by plant resistance gene products and regulators of cell death in animals. *Curr. Biol.*
453 8(7):R226-R228.
454 24. Deretic V, Saitoh T, & Akira S (2013) Autophagy in infection, inflammation and immunity.
455 *Nat. Rev. Immunol.* 13(10):722.
456 25. Björklund ÅK, Ekman D, & Elofsson A (2006) Expansion of protein domain repeats.
457 *PLOS Comput. Biol.* 2(8):e114.
458 26. Enkhbayar P, Kamiya M, Osaki M, Matsumoto T, & Matsushima N (2004) Structural
459 principles of leucine-rich repeat (LRR) proteins. *Proteins* 54(3):394-403.
460 27. Lee S-C*, et al.* (2012) Design of a binding scaffold based on variable lymphocyte receptors
461 of jawless vertebrates by module engineering. *Proc. Natl. Acad. Sci. U.S.A.* 109(9):3299-
462 3304.
463 28. Heu W, Choi J-M, Lee J-J, Jeong S, & Kim H-S (2014) Protein binder for affinity
464 purification of human immunoglobulin antibodies. *Anal. Chem.* 86(12):6019-6025.
465 29. Jeong S, Heu W, Kim J-w, & Kim H-S (2016) Protein binders specific for immunoglobulin
466 g from different species for immunoassays and multiplex imaging. *Anal. Chem.*
467 88(23):11938-11945.
468 30. Kozakov D*, et al.* (2017) The ClusPro web server for protein–protein docking. *Nat. Protoc.*
469 12(2):255.
470 31. Brenke R*, et al.* (2012) Application of asymmetric statistical potentials to antibody–protein
471 docking. *Bioinformatics* 28(20):2608-2614.
472 32. Hornak V*, et al.* (2006) Comparison of multiple Amber force fields and development of
473 improved protein backbone parameters. *Proteins* 65(3):712-725.
474 33. Ponder JW (2004) TINKER: Software tools for molecular design. *Washington University*
475 *School of Medicine, Saint Louis, MO* 3.
476 34. Schymkowitz J*, et al.* (2005) The FoldX web server: an online force field. *Nucleic acids*
477 *research* 33(suppl_2):W382-W388.
478 35. Lee Jj*, et al.* (2015) Enzymatic prenylation and oxime ligation for the synthesis of stable
479 and homogeneous protein–drug conjugates for targeted therapy. *Angew. Chem. Int. Ed.*
480 *Engl.* 54(41):12020-12024.
481 36. Hwang D-E*, et al.* (2016) Effective suppression of C5a-induced proinflammatory response
482 using anti-human C5a repebody. *Biochem. Biophys. Res. Commun.* 477(4):1072-1077.
483 37. De Vries SJ, Van Dijk M, & Bonvin AM (2010) The HADDOCK web server for data-
484 driven biomolecular docking. *Nat. Protoc.* 5(5):883.
485 38. Dominguez C, Boelens R, & Bonvin AM (2003) HADDOCK: a protein− protein docking
486 approach based on biochemical or biophysical information. *J. Am. Chem. Soc* 125(7):1731-
487 1737.
488 39. Vajda S & Kozakov D (2009) Convergence and combination of methods in protein–protein
489 docking. *Curr. Opin. Struct. Biol.* 19(2):164-170.
490 40. Rosenfeld L, Heyne M, Shifman JM, & Papo N (2016) Protein engineering by combined
491 computational and in vitro evolution approaches. *Trends Biochem. Sci.* 41(5):421-433.
492 41. Still WC, Tempczyk A, Hawley RC, & Hendrickson T (1990) Semianalytical treatment of
493 solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc* 112(16):6127-6129.

494 42. Eswar N, *et al.* (2006) Comparative protein structure modeling using Modeller. *Curr.*
495 *Protoc. Bioinformatics* 15(1):5.6. 1-5.6. 30.
496 43. Coleman L & Mahler SM (2003) Purification of Fab fragments from a monoclonal
497 antibody papain digest by Gradiflow electrophoresis. *Protein Expr. Purif.* 32(2):246-251.
498 44. Otwinowski Z & Minor W (1997) Processing of X-ray diffraction data collected in
499 oscillation mode. *Meth. Enzymol.* 276:307-326.
500 45. Adams PD, *et al.* (2010) PHENIX: a comprehensive Python-based system for
501 macromolecular structure solution. *Acta Crystallogr. D* 66(2):213-221.
502 46. Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta*
503 *Crystallogr. D* 60(12):2126-2132.
504 47. Adams PD, *et al.* (2002) PHENIX: building new software for automated crystallographic
505 structure determination. *Acta Crystallogr. D* 58(11):1948-1954.

506

507

508

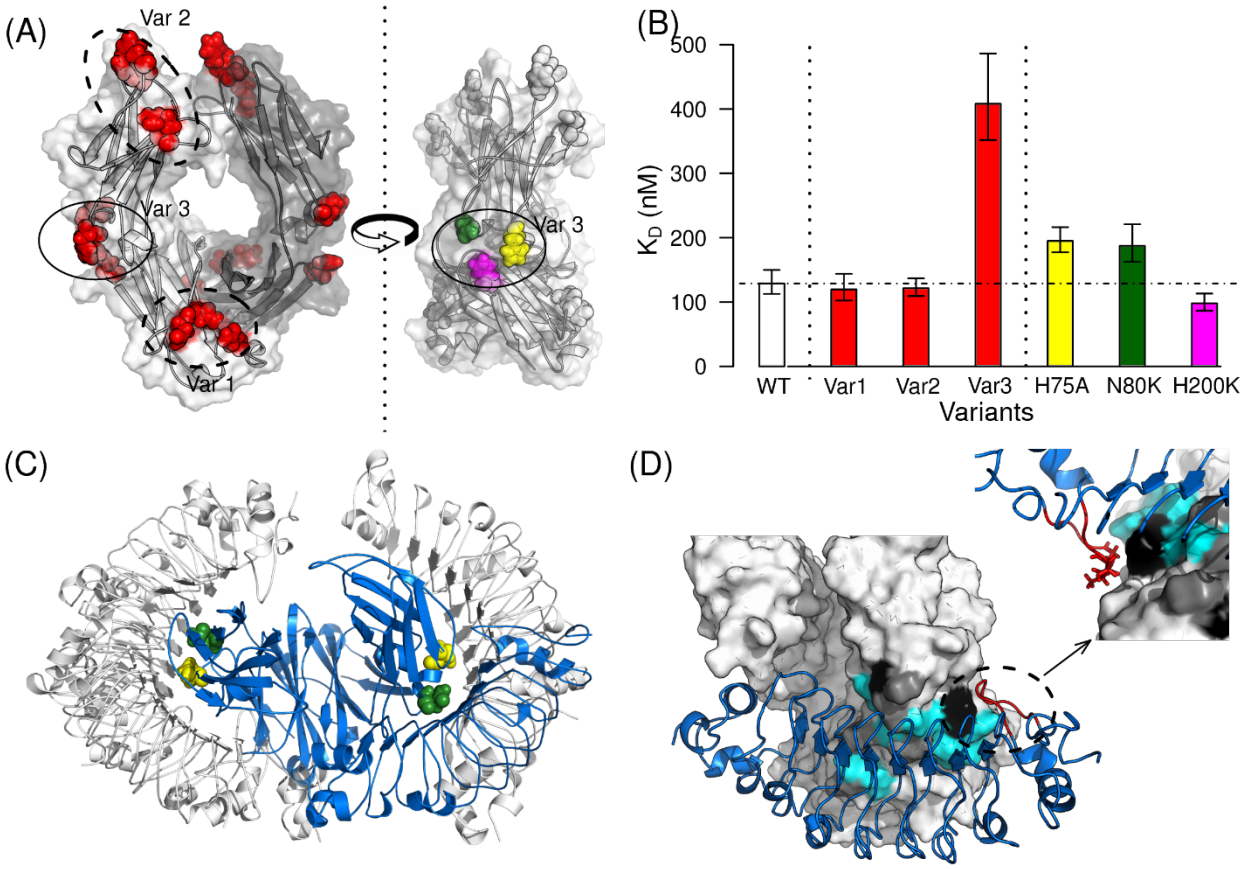
509 **Table 1. Test sets.** The numbers in the brackets indicate results from ClusPro without the antibody
510 mode option. For C5a, the numbers with asterisks (*) are results using the crystal structure for
511 docking with the precise definition of paratopes.

Target		**Human IgG Fc**	**Crystal Structure Complexes**		
			Interleukin 6 (IL-6)	Epidermal Growth Factor Receptor (EGFR)	Complement Component 5a (C5a)
Complex (PDB)		5Z7K	4J4L	4UIP	5B4P
Rb homology template		3RFS	5B4P	3RFS	4J4L
Unbound target		3AVE	1ALU	1NQL	1KJS
Cα RMSD (Å)	Target	1.72	0.92	2.62	1.76
	Repebody	1.39	0.63	1.62	1.17
Number of docking models		29	30 (106)	30 (88)	23 (65) | ***28**
Number of EpiScope Designs		3	4	2	2
Number of localized docking models		7	5 (11)	12 (20)	6 (34) | ***7**
Best I-RMSD (Å)		2.32	1.45 (1.78)	2.23 (6.61)	3.61 (5.53) | ***0.62**
Best f_{nat}		0.43	0.62 (0.49)	0.38 (0.09)	0.27 (0.14) | ***0.76**

512

513

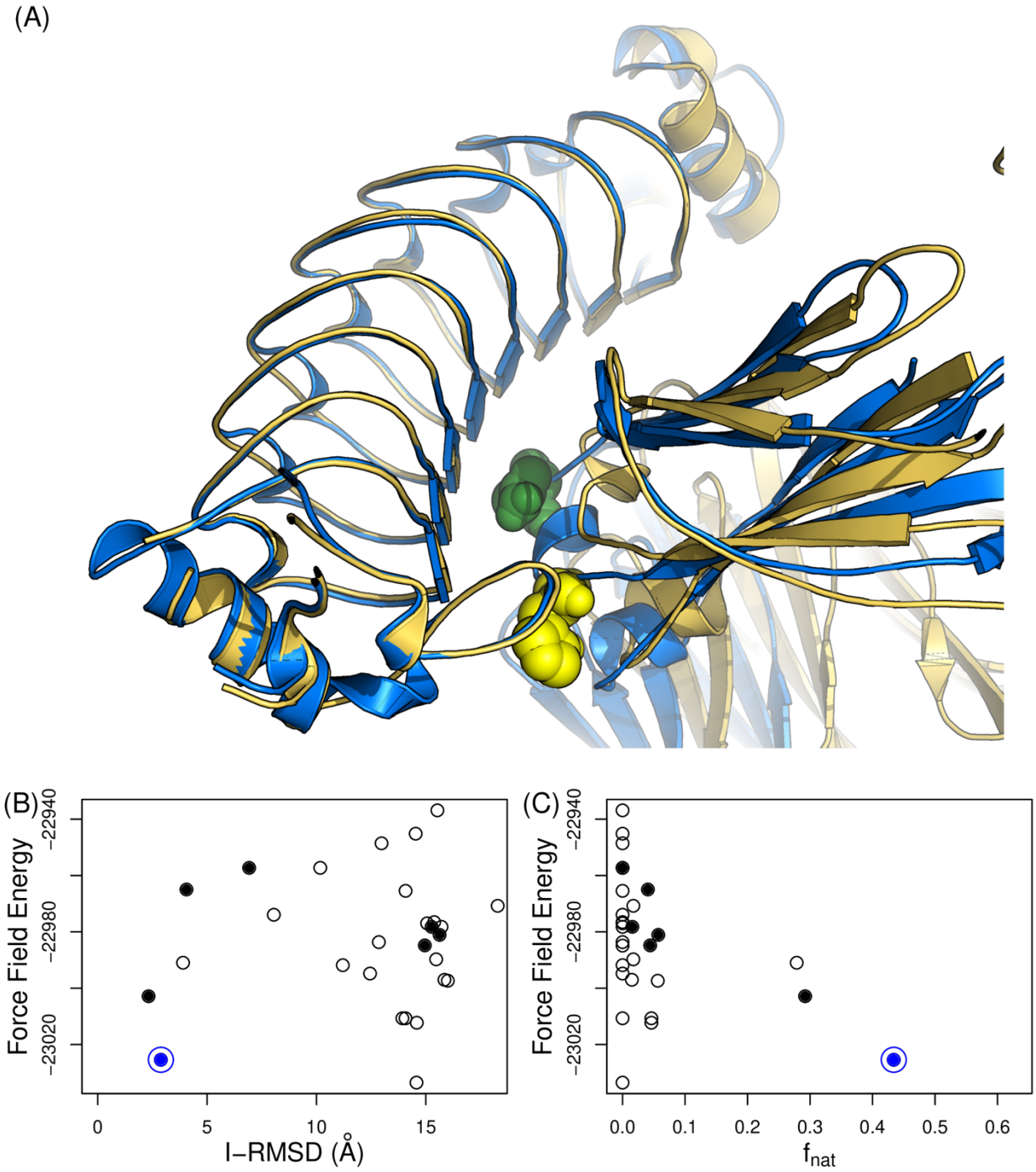
514 **Figure 1. Epitope localization and binding mode prediction of human IgG_1 Fc (hFc)- binding**
515 **repebody (RbF4).** (**A**) EpiScope initially identifies three designs considering the symmetry of the
516 Fc structure. (**B**) A set of the three mutations in Var 3 (H75A, N80K, and H200K) clearly disrupts
517 the binding, and single mutations of Var 3 were further examined. The isothermal titration
518 calorimetry results indicate that H200K may not be involved in the binding. Details are provided
519 in **Fig. S2.** (**C**) There are seven docking models in contact with H75A and N80K. The blue docking
520 complex was found to be at the lowest energy. (**D**) A close look into the model suggests that the
521 repebody loop (**Fig. S1**) may be responsible for the binding specificity of RbF4 repebody to hFc.
522 IgG from the three species are considered (human; mouse, mFc; and rabbit, rFc). The residues that
523 mFc, hFc, and rFc commonly share are colored in cyan, the two common species are in gray, and
524 those not shared in common at all are in black.



525

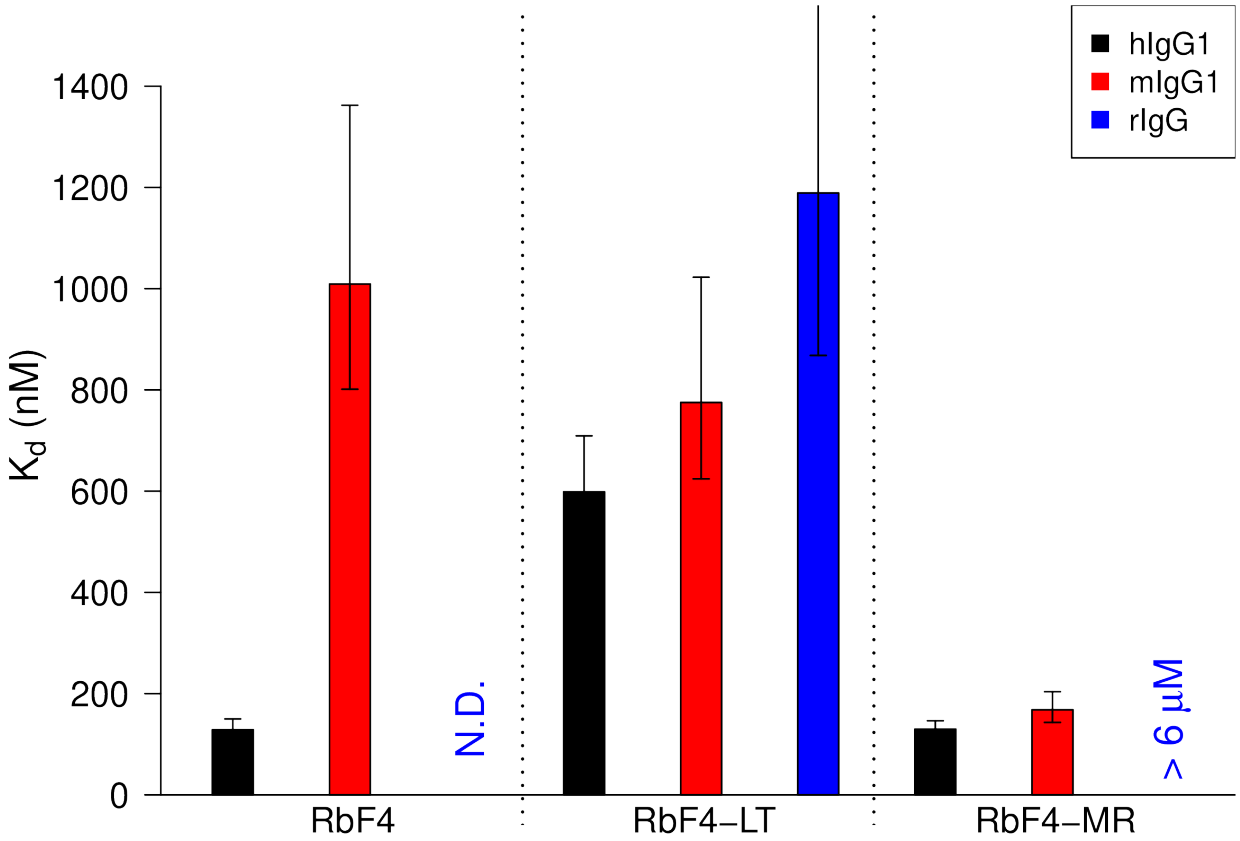
526

527 **Figure 2. Computational binding mode prediction of RbF4-hFc complex.** The crystal structure
528 is in yellow and the predicted model is in blue. Docking models that are in contact with the epitope
529 overlapping residues (localized docking models) are shown with the solid circle. The blue circle
530 is the model with the lowest force field energy (AMBER99sb) score. (**A**) The crystal structure
531 (6KA7), which is in yellow, whereas the docking models with the lowest energies are in blue on
532 the right-hand side. (**B**) I-RMSD and (**C**) f_{nat}.



533

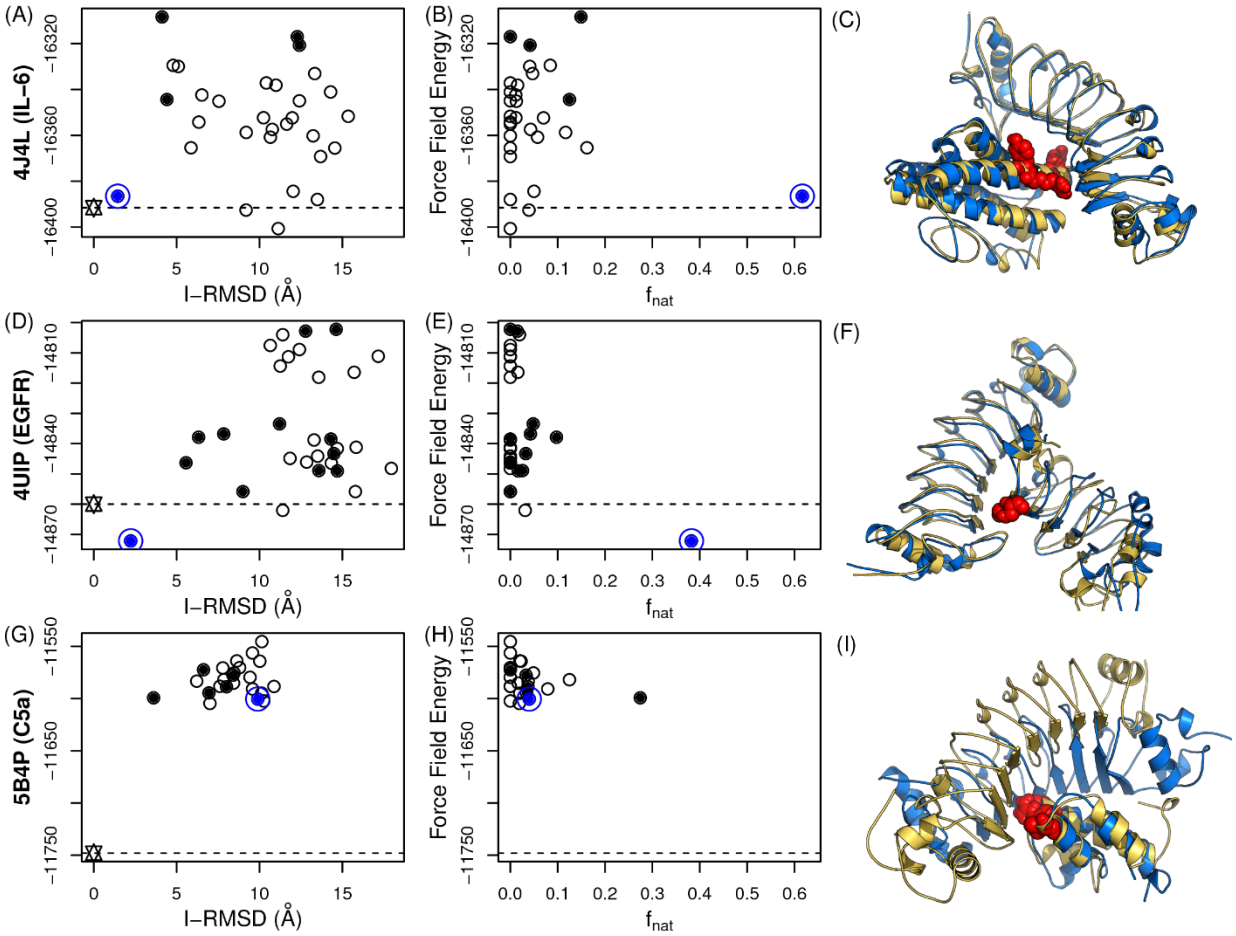
534 **Figure 3. Binding specificity control and affinity improvement based on model complex**
535 **structure.** The predicted binding mode gives insight that the binding specificity of RbF4 comes
536 from the loop (**Fig. 1D**). The IgGs from the three species are bound to the loop truncated RbF4
537 (RbF4-LT). Further computational engineering in the loop on the model complex identified
538 mutations (S241M and S244R, RbF4-MR) that could significantly improve the binding affinity
539 for $mIgG_1$ (1 μM to 168.1 nM) while maintaining that for $hIgG_1$ (**Table S2** for details).



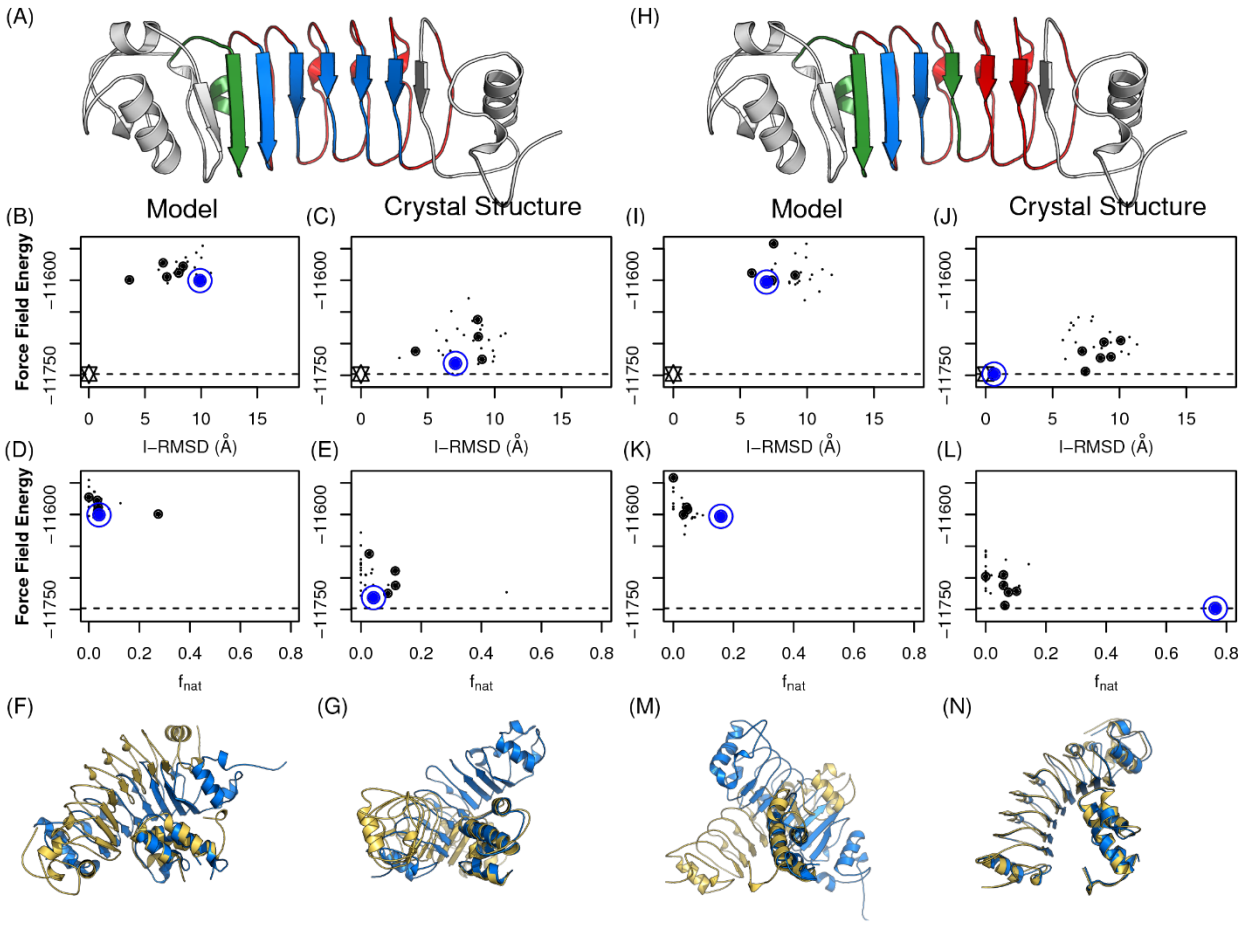
540

541

Figure 4. Binding mode prediction on various targets. Docking models that are in contact with epitope overlapping residues (localized docking models) are in solid circle. The blue circle is the model with the minimum force field energy (AMBER99sb). Wild-type (star) energy levels are depicted as dotted lines. Crystal structures are in yellow and the docking models with the lowest energies are in blue on the right hand side. In the cases of IL-6 (**A**, **B** and **C**) and EGFR (**D**, **E**, and **F**), the filtered models with the lowest energy values are native-like. Binding mode prediction for C5a is presented in **G**, **H** and **I**. As shown in IL-6, some models have lower energy values than the most native-like docking model, indicating that ranking only by the force-field energy is not sufficient for binding mode prediction.



553 **Figure 5. Impacts of high-quality structure and paratope definition.** The incorporation of
554 phage display results and the use of high quality structures (here crystal structures) lead to an
555 extremely accurate prediction of the binding mode. LRRV in blue is assigned for attraction, in red
556 for repulsion and green for neutral (**A** and **H**). Results in the right two columns (**I-N**) are from
557 docking models with precisely assigned paratopes. Docking models that are in contact with epitope
558 overlapping residues (localized docking models) are in solid circle. Models with the minimum
559 force field energy values are in blue circles. Wild-type (star) energy levels are depicted as dotted
560 lines. Crystal structures are in yellow and the docking models with the lowest energies are in blue
561 on the right hand side.



562

563