

# SCIENTIFIC REPORTS



OPEN

## Modeling and fitting protein-protein complexes to predict change of binding energy

Daniel F.A.R. Dourado & Samuel Coulbourn Flores

Received: 04 January 2016

Accepted: 18 April 2016

Published: 13 May 2016

It is possible to accurately and economically predict change in protein-protein interaction energy upon mutation ( $\Delta\Delta G$ ), when a high-resolution structure of the complex is available. This is of growing usefulness for design of high-affinity or otherwise modified binding proteins for therapeutic, diagnostic, industrial, and basic science applications. Recently the field has begun to pursue  $\Delta\Delta G$  prediction for homology modeled complexes, but so far this has worked mostly for cases of high sequence identity. If the interacting proteins have been crystallized in free (uncomplexed) form, in a majority of cases it is possible to find a structurally similar complex which can be used as the basis for template-based modeling. We describe how to use MMB to create such models, and then use them to predict  $\Delta\Delta G$ , using a dataset consisting of free target structures, co-crystallized template complexes with sequence identity with respect to the targets as low as 44%, and experimental  $\Delta\Delta G$  measurements. We obtain similar results by fitting to a low-resolution Cryo-EM density map. Results suggest that other structural constraints may lead to a similar outcome, making the method even more broadly applicable.

Modeling Protein-Protein Interactions (PPIs)<sup>1</sup> is fundamentally important in biology as it probes normal as well as diseased protein function. For example, such models explain the role of Parkinson's-disease associated mutations in Parkin<sup>1-3</sup>. PPIs are also important in the development of therapeutic and diagnostic biologics (monoclonal antibodies, or mAbs, and alternative scaffolds)<sup>4</sup>.

Biologics have a growing and economically substantial field of application. However raising antibodies or finding an alternative scaffold to bind a given target is difficult and time consuming. Even when starting with a scaffold that binds reasonably, affinity maturation requires a substantial experimental effort, and maintaining specificity can be a challenge<sup>5</sup>. Likewise protein engineering often creates many simultaneous mutations, with possible immunogenicity and solubility issues, and no insight as to which substitutions are responsible for the main effect<sup>6</sup>. Thus there is demand for an economical computational method which will suggest a relatively small number of substitutions which have high likelihood of improving binding.

Computational methods have made significant progress for cases where a crystallographic complex is available of the potential biologic bound to its target (we will refer to these as bound structures). Some are Molecular Dynamics (MD) based methods<sup>7-10</sup>, which typically are associated with a high computational cost. So, the applicability of such methods to large complexes or to a substantial number of mutations, which is required the case for protein-protein affinity maturation protocols, can be quite limited. On the other hand, Knowledge Based (KB) methods, which empirically combine several energetic terms including implicit solvent<sup>11-16</sup>, are fast but most perform little or no structural optimization and cannot model the backbone rearrangements induced by mutation. KB methods have also been combined with sequence analysis<sup>17</sup>, and interface structure alignments<sup>18</sup> but this requires evolutionary information which is not available for all complexes (e.g. many biologics), and further has only been demonstrated for homology models based on high sequence identity (only 4% of their dataset had sequence identity below 50%)<sup>17</sup>. Recently, we described Zone Equilibration of Mutants (ZEMu)<sup>1</sup>, validated with 1254 mutants (1-15 simultaneous mutations) of 65 different complexes, which offers both accuracy and economy. ZEMu is implemented in MacroMoleculeBuilder (MMB)<sup>19,20</sup>, a multiscale internal-coordinate modeling code in which flexibility and an all-atom force field can be limited to regions of interest<sup>1,21</sup>. The method significantly improves the existing FoldX potential<sup>13</sup>, and arguably shows promise to improve others<sup>17,18</sup> which perform limited structural minimization.

Department of Cell and Molecular Biology, Computational and Systems Biology, Uppsala University, Biomedical Center Box 596, 751 24, Uppsala, Sweden. Correspondence and requests for materials should be addressed to S.F. (email: sam@xray.bmc.uu.se)

There are limited options for computing  $\Delta\Delta G$  for the case in which the interacting proteins have only been crystallized in the free form. For many such structures, low resolution density maps of their complex are available<sup>22</sup>. The recent explosion in Cryo Electron Microscopy, brought about by the direct electron detector<sup>23</sup>, promises a rich source of new structural data, notably of complexes which are hard to crystallize. In addition, solution scattering produces many low-resolution density maps<sup>24</sup>, and Free-Electron Lasers promise to eventually reach comparable single-molecule resolution<sup>25</sup>.

Alternatively, for most structures available in the free form (referred to as targets)<sup>26</sup>, it is possible to find a structurally related template which can be used to build a template-based model of the complex<sup>27</sup>. Template modeling uses a structural alignment, which can be done accurately even at low sequence identity<sup>27,28</sup>.

This realization has led to considerable interest in template-based docking<sup>29</sup>. Specific cases in which the free structure is related to one of the proteins in the template complex include antibody-bound IGF-I (complex exists of the related IGF-II bound to an antibody)<sup>30</sup>, human Chorionic Somatomammotropin (hCS)<sup>31</sup> vs. human Growth Hormone (GH) Receptor (complex exists of GH vs. GH Receptor), Fc $\gamma$ RI vs. IgG1 (complexes have long existed of Fc $\gamma$ R2 and Fc $\gamma$ R3 vs. IgG, while for Fc $\gamma$ RI vs. IgG1 a mutated complex was recently solved)<sup>32–34</sup>. In this study we implement a fast and simple to use internal-coordinate template-based docking protocol in MMB<sup>19,20</sup>, that works even in the range of ~40% sequence identity for homologous proteins (quite near the twilight zone)<sup>27,35</sup> and extend ZEMu<sup>1</sup> to predict  $\Delta\Delta G$  for thus-modeled and fitted complexes.

## Results

The template-based docking protocol introduced here results in good  $\Delta\Delta G$  precision for homologous templates, those which (in this work) have sequence identities (vs. the targets) in the range of 44% to 51%. It is naturally more precise for self-templates, meaning those which have high sequence identity (>93%) to their targets. In this work when we provide RMSD (Root Mean Square Deviation) we refer in all cases to the discrepancy in backbone 3D atomic structure of modeled vs. template complexes. When we provide RMSE (Root Mean Square Error) and correlation, we are comparing experimental to computed  $\Delta\Delta G$ .

Double-free models are made by docking two targets onto the template complex, while single-free models are made by docking one target onto the template (see Methods). For the double-free models based on self-templates the Root Mean Square Deviations (RMSDs) range from 0.71 to 3.42 Å (Table 1). The exception is the TGF- $\beta$  Type II Receptor/TGF- $\beta$ 3 complex for which the RMSD is 22.84 Å (Table 1). If we analyze the structures of the model and template in detail (Fig. S1) we can observe that distal region of the co-crystal chain A (TGF- $\beta$  3) is poorly resolved. In fact, if we omit template chain A residues 40 to 80 the RMSD decreases from 21.7826 Å to just 0.8819 Å, which is in line with RMSD found for the other complexes (Table 1). Since these poorly resolved residues are placed on a distal region of chain A, far from the interface, the  $\Delta\Delta G$  prediction precision with this model is in line with that of the rest of the dataset. We describe this complex in more detail later. For the single- and double-free models based on homologous templates the RMSD ranges from 2.76 to 4.86 Å. As expected the RMSDs of this sub-group are higher than the ones observed for the double-free models based on self-templates. However the differences are relatively small, which is in line with idea that structure is more conserved than sequence<sup>27</sup>. We discuss this point in more detail below.

Based on the validation dataset described in Table 1, we compared the performance of ZEMu to FoldX-only (meaning FoldX with no prior MMB equilibration) in predicting  $\Delta\Delta G$ . For the entire dataset the correlation between FoldX-only and experimental  $\Delta\Delta G$  ( $\Delta\Delta G_{\text{FoldX-only}}$  and  $\Delta\Delta G_{\text{exp}}$ , respectively) is 0.12, while the Root Mean Square Error (RMSE) is 1.83 kcal/mol. To our knowledge this is the first time FoldX is evaluated with modeled complexes; this also serves as an external (non-ZEMu) validation of our modeling protocol. For ZEMu the correlation improves to 0.34 (p-value < 0.00001); the RMSE also improves, to 1.54 kcal/mol (Table 2, Fig. 1, Tables S1 and S2).

The improvement over FoldX-only is reflected in the complete dataset as well as in the single and multiple mutant sub-groups (Table 2). For the single mutants, FoldX-only shows an RMSE of 1.89 kcal/mol and correlation of 0.06, while for ZEMu we obtain an RMSE of just 1.54 kcal/mol and a higher correlation of 0.24 (p-value = 0.00004) (Table 2, Fig. 1, Tables S1 and S2). In the case of the multiple mutants FoldX-only achieves an RMSE of 1.79 kcal/mol and correlation of 0.15, and ZEMu an RMSE of 1.54 kcal/mol and a correlation of 0.37 (p-value < 0.00001) (Table 2, Fig. 1, Tables S1 and S2).

From the main dataset we also created a sub-group comprising mutants for which self-template structures are available (Table 3). Based on this sub-group we compared the performance of ZEMu for modeled vs. crystallographic complexes. As expected, performance was better for crystallographic than for modelled complexes, but only moderately (RMSE of 1.58 vs. 1.76, Correlation of 0.61 vs. 0.38, respectively). This further highlights the quality of the modeling protocol. In the particular case of the TGF- $\beta$  Type II Receptor / TGF- $\beta$ 3 complex model, the RMSD vs. its self-template is 22.84 Å, when computed based on all resolved residues. As explained above the huge RMSD value found is due to poorly resolved residues in a distal region of chain A co-crystal (Table 1, Fig. S1) and so does not affect the quality and performance of the model at the interface. The RMSEs are 1.68 kcal/mol and 1.19 kcal/mol, and the correlations are 0.28 and 0.83, for the model and co-crystal, respectively. The performance is thus in line with the other models.

We then tested a ZEMu variant with an additional flexibility radius (0–14 Å) (Fig. 2) centered on the mutation site, which can include residues from the binding partner. When a radius of 4 Å is used an RMSE very slightly higher than that of regular ZEMu is obtained, also an added computational cost is incurred, so there is no reason to use the added flexibility. Nonetheless, ZEMu variants with an extra flexibility radius of 4 to 10 Å still show better performance than FoldX-only.

The application to the Fc $\gamma$ RI/IgG1 complex illustrates some of the advantages of ZEMu. We created single-free and double-free template models, and a fitted model, of this complex (Table 4, Table S2), based on an Fc $\gamma$ R3/IgG1 template (sequence identity of Fc $\gamma$ RI with Fc $\gamma$ R3 is 45%)<sup>33</sup>. When we performed this work no Fc $\gamma$ RI/

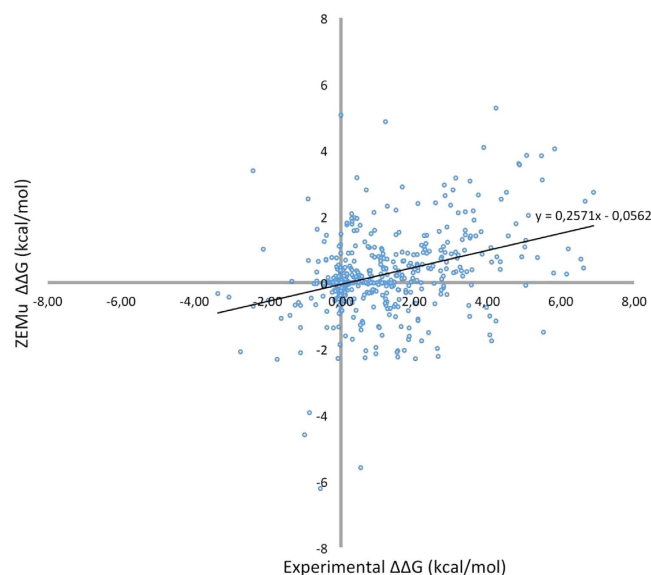
| Protein 1 (PDB)                               | Protein 2 (PDB)                 | Complex                      | MODEL   | Sequence identity                           | Model vs template RMSD (Å)  | Number of mutants (substitutions) |
|---|---------------------------------|------------------------------|---|---|---|-----------------------------------|
| IgG1-K D44.1 FAB (1MLB)                       | Hen egg-white lysozyme (1DPX)   | 1MLC                         | IgG1-K D44.1 FAB/Hen egg-white lysozyme                         | 1MLC:A/B/1MLB:A=100%<br>1MLC:E/1DPX:A=100%  | global = 0.98<br>chain A = 0.96<br>chain B = 1.10<br>chain E = 0.75   | 16(26)                            |
| IgG1-K D1.3 FV (1VFA)                         | Hen egg-white lysozyme (1DPX)   | 1VFB                         | IGG1-KAPPA D1.3 FV/Hen egg-white lysozyme                       | 1VFB:A/1VFA:A=100%<br>1VFB:C/1DPX:A=100%    | global = 0.80<br>chain A = 0.33<br>chain B = 0.68<br>chain C = 1.19   | 42(56)                            |
| HyHEL-63 FAB (1DQM)                           | Hen egg-white lysozyme (1DPX)   | 1DQJ                         | HyHEL-63 FAB/Hen egg-white lysozyme                             | 1DQJ:A/1DQM:A=100%<br>1DQJ:C/1DPX:A=100%    | global = 0.80<br>chain A = 0.69<br>chain B = 0.91<br>chain C = 0.78   | 34(47)                            |
| TGF- $\beta$ Type II Receptor (1M9Z)          | TGF- $\beta$ 3 (1TGJ)           | 1KTZ                         | TGF- $\beta$ Type II Receptor/TGF- $\beta$ 3                    | 1KTZ:A/1TGJ:A=100%<br>1KTZ:A/1M9Z:A=98.2%   | global = 22.84<br>global (without 40-80) = 0.88<br>chain A = 21.79<br>chain A(without 40-80) = 1.21<br>chain B = 0.63 | 27(27)                            |
| $\beta$ -lactamase inhibitor protein-I (3GMU) | TEM-1 $\beta$ -lactamase (1ZG4) | 1JTG                         | $\beta$ -lactamase inhibitor protein-I/TEM-1 $\beta$ -lactamase | 1JTG:A/3GMU:A=100%<br>1JTG:B/1ZG4:A=98.2%   | global = 3.42<br>chain A = 1.94<br>chain B = 0.64   | 143(307)                          |
| IgG1 (4DZ8)                                   | Fragment B of protein A (2JWD)  | 1FC2                         | IgG1/Fragment B of protein A                                    | 1FC2:C/2JWD:A=93.1%<br>1FC2:D/4DZ8:A=96.4%  | global = 1.67<br>chain C = 0.27<br>chain D = 1.35   | 9(9)                              |
| Iso-1-cytochrome C (1NMI)                     | Cytochrome C peroxidase (3R99)  | 2PCC                         | Iso-1-Cytochrome C/Cytochrome C peroxidase                      | 2PCC:A/3R99:A=99.3%<br>2PCC:B/1NMI:A=99.1%  | global = 1.21<br>chain A = 0.40<br>chain B = 2.18   | 12(18)                            |
| IgG1 (3DNK)                                   | Fc $\gamma$ R II (3RY4)         | 3RY6                         | IgG1/Fc $\gamma$ R II   | 3RY6:C/3RY4:A=97.1%<br>3RY6:A/3DNK:A=98.8%  | global = 2.72<br>chain A = 1.90<br>chain B = 3.62<br>chain C = 2.22   | 65(138)                           |
| IgG1 (3DNK)                                   | Fc $\gamma$ R III (1E4J)        | 1E4K                         | IgG1/Fc $\gamma$ R III  | 1E4K:C/1E4J:A = 100%<br>1E4K:A/3DNK:A=97.2% | global = 1.99<br>chain A = 1.59<br>chain B = 2.55<br>chain C = 1.59   | 95(155)                           |
| IgG1 (3DNK)                                   | Fc $\gamma$ R N (4N0F)          | 4N0U                         | IgG1/Fc $\gamma$ R N  | 4N0U:A/4N0F:A=100%<br>4N0U:A/3DNK:A=97.1%   | global = 0.71<br>chain A = 0.31<br>chain E = 1.01   | 53(53)                            |
|   |                                 |                              |   |   |   |                                   |
| IgG1 (1E4K)                                   | Fc $\gamma$ R I (3RJD)          | 1E4K(IgG1/Fc $\gamma$ R III) | IgG1/Fc $\gamma$ R I  | 1E4K:C/3RJD:A=45.2%                         | global = 2.76<br>chain A = 2.80<br>chain B = 2.34<br>chain C = 2.80   | 66(146)                           |
| IgG1 (3DNK)                                   | Fc $\gamma$ R II (3RY4)         | 1E4K(IgG1/Fc $\gamma$ R III) | Ig1/Fc $\gamma$ R II  | 1E4K:C/3RY4:A=44.0%<br>1E4K:A/3DNK:A=97.2%  | global = 3.61<br>chain A = 3.10<br>chain B = 4.23<br>chain C = 3.36   | 65(138)                           |
| IgG1 (3DNK)                                   | Fc $\gamma$ R III (1E4J)        | 3RY6(IgG1/Fc $\gamma$ R II)  | IgG1/Fc $\gamma$ R III  | 3RY6:C/1E4J:A=50.9%<br>3RY6:A/3DNK:A=98.8%  | global = 4.86<br>chain A = 4.79<br>chain B = 4.61<br>chain C = 4.88   | 95(155)                           |
| IgG1 (3DNK)                                   | Fc $\gamma$ R I (3RJD)          | 1E4K(IgG1/Fc $\gamma$ R III) | IgG1/Fc $\gamma$ R I  | 1E4K:C/3RJD:A=45.2%<br>1E4K:A/3DNK:A=97.2%  | global = 3.68<br>chain A = 3.03<br>chain B = 2.52<br>chain C = 4.80   | 62(128)                           |
| IgG1 (3DNK)                                   | Fc $\gamma$ R I (3RJD)          | Density map from 1E4K        | IgG1/Fc $\gamma$ R I  | 1E4K:C/3RJD:A=45.2%<br>1E4K:A/3DNK:A=97.2%  | global = 3.37<br>chain A = 2.71<br>chain B = 2.56<br>chain C = 4.30   | 62(128)                           |
|   |                                 |                              |   |   | TOTAL   | 846(1531)                         |

**Table 1. Validation dataset.** The dataset is divided in two groups. The first is composed of double-free template models, based on self-templates. The second group (bottom of table, separated by a blank row) includes : 1) a single-free template model of IgG1/Fc $\gamma$ R I, based on IgG1/Fc $\gamma$ R III crystal, where the structure of IgG1 from the crystallographic complex is kept rather than being replaced; 2) double-free template models of IgG1/Fc $\gamma$ R I, IgG1/Fc $\gamma$ R II and IgG1/Fc $\gamma$ R III, which were modeled from based on homologous templates; 3) A model of IgG1/Fc $\gamma$ R I built by fitting to a low-resolution density map synthesized from an IgG1/Fc $\gamma$ R III crystallographic complex<sup>21</sup> Sequence identity and backbone RMSD of targets vs. templates are shown.

IgG1 co-crystal was available, but one was reported more recently<sup>34</sup>. With the Fc $\gamma$ RI/IgG1 single-free template model, FoldX-only yields an RMSE of 2.33 and a correlation of 0.12 while ZEMu RMSE is just 0.92 and correlation is 0.42. The difference in performance between the two protocols is overwhelmingly due to three mutations involving N-terminus residue G236. When the three mutants are left out, FoldX-only RMSE drops substantially

| Dataset          | Number of mutants | Models          |             |                 |             |
|------------------|-------------------|-----------------|-------------|-----------------|-------------|
|                  |                   | FoldX-only      |             | ZEMu            |             |
|                  |                   | RMSE (kcal/mol) | Correlation | RMSE (kcal/mol) | Correlation |
| All mutants      | 846               | 1.83            | 0.12        | 1.54            | 0.34        |
| Multiple mutants | 584               | 1.79            | 0.15        | 1.54            | 0.37        |
| Single mutants   | 262               | 1.89            | 0.06        | 1.54            | 0.24        |

**Table 2. Comparison between Foldx-only and ZEMu performance.**



**Figure 1. ZEMu versus experimental  $\Delta\Delta G$  over the full dataset (846 mutants).**

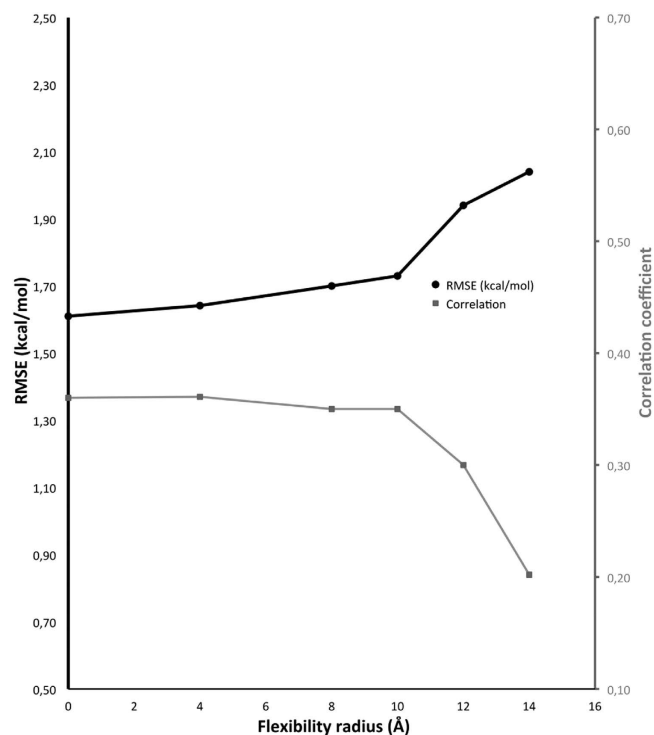
| Number of mutants | Double-free Models |             |                 |             | Co-crystals     |             |
|-------------------|--------------------|-------------|-----------------|-------------|-----------------|-------------|
|                   | FoldX-only         |             | ZEMu            |             | ZEMu            |             |
|                   | RMSE (kcal/mol)    | Correlation | RMSE (kcal/mol) | Correlation | RMSE (kcal/mol) | Correlation |
| 558               | 2.00               | 0.17        | 1.76            | 0.38        | 1.58            | 0.61        |

**Table 3. Comparing the performance of ZEMu on modeled vs. crystallographic complexes.** Note that ZEMu performance decreased only moderately (1.76 vs. 1.58 RMSE) for modeled vs. crystallographic complexes.

to 0.32 and the correlation increases to 0.32. Though this is a small sample, it illustrates even more strikingly that FoldX's rigid backbone is particularly unsuitable for modeling the terminal region, which has characteristically high mobility.

For the double-free and fitted  $Fc\gamma RI/IgG1$  models the three mutations involving N-terminus residue G236 are immediately adjacent to unresolved residues. The free  $IgG1$  structure (3DNK) was missing several residues from the N-terminus (residues 229 to 235, part of the lower hinge connecting Fc to Fab in a full-length  $IgG1$ ), which were resolved in the template complex (1E4K). These three mutants plus one mutation in the missing region therefore could not be modeled in the double-free and fitted models. ZEMu still outperforms FoldX-only but by a smaller margin (Table 4). Directly comparing against the double-free and fitted models, we can conclude that for both FoldX-only and ZEMu the best performance was obtained for the fitted model (Table 4). If we analyze the RMSD of both models with respect to the co-crystal structure we can observe that the fitted model has a lower RMSD by  $\sim 0.32$  Å. In the fitted model the hinges between the D1 and D2 domains on  $Fc\gamma RI$ , and between the CH2 and CH3 domains on Fc, were made flexible, to allow domain motions, explaining the better RMSD of chains A and C. This highlights the efficacy of the fitting protocol.

ZEMu and MMB performance for models of  $Fc\gamma RIII/IgG1$  and  $Fc\gamma RII/IgG1$  based on homologous templates further demonstrates the efficacy of the protocol in cases where the self-template is not available. For instance, for the  $Fc\gamma RIII/IgG1$  model based on its self-template the RMSE is 0.85 kcal/mol and the RMSD is 1.98 Å. In the case of the  $Fc\gamma RIII/IgG1$  model based on the  $Fc\gamma RII/IgG1$  template (PDB:3RY6)<sup>32</sup> (sequence identity of  $Fc\gamma RIII$  with  $Fc\gamma RII$  is 51%) the RMSE increases to 1.17 kcal/mol (0.32 kcal/mol higher). The RMSD (vs. the



**Figure 2.** Results of flexibilizing a spatial neighbourhood of the mutation sites for a dataset composed of 687 mutants (not includes the template models of Fc $\gamma$ R1II/IgG1 and Fc $\gamma$ R1I/IgG1 based on a homologous co-crystal complex). Ordinary ZEMu has only five flexible residues about each mutation site flexibilized. Here we also flexibilize all residues within a radius (0–14 Å) of the mutation sites. Radius of 0 Å corresponds to ordinary ZEMu.

| MODEL                            | Number of mutants | FoldX-only      |             | ZEMu            |             |
|----------------------------------|-------------------|-----------------|-------------|-----------------|-------------|
|                                  |                   | RMSE (kcal/mol) | Correlation | RMSE (kcal/mol) | Correlation |
| Single-free template-based model | 66                | 2.33            | 0.12        | 0.92            | 0.42        |
| Double-free template-based model | 62                | 0.64            | 0.41        | 0.57            | 0.49        |
| Fitted model                     | 62                | 0.49            | 0.47        | 0.47            | 0.50        |

**Table 4.** Comparison between Foldx-only and ZEMu performance for Fc $\gamma$ RI/IgG1 single-free and double-free template-based and fitted models.

self-template) is 4.86 Å, which is low but still 2.4 fold higher than the RMSD found for the double-free model based on a self-template. Similarly, for the Fc $\gamma$ R1I/IgG1 model based on a self-template the RMSE is 0.90 kcal/mol while the RMSD (vs. the self-template) is 2.72 Å. For the Fc $\gamma$ R1I/IgG1 model based on Fc $\gamma$ R1II/IgG1 template (PDB:1E4K)<sup>33</sup> (sequence identity of Fc $\gamma$ R1I with Fc $\gamma$ R1II is 44%) the RMSE increases to just 1.13 kcal/mol, possibly because the RMSD increases only 1.3-fold to 3.61 Å. For these two complexes together (150 mutants, including some with single and some with multiple substitutions) the RMSE increased by 0.22 kcal/mol when lower sequence identity (44–51%) homologous templates were used in the modeling instead of high sequence identity (>93%) self-templates.

Positive Predictive Value (PPV) is defined as TP/(TP+FP) (see supporting information). In order to compute statistical quantities like this we would need a random sample of possible mutations. However many of the available experimental  $\Delta\Delta G$ 's are the result of alanine scanning mutagenesis experiments and although 36% of the dataset have experimental  $\Delta\Delta G < 0$  kcal/mol, only 5% have experimental  $\Delta\Delta G < -0.5$  kcal/mol. Also given the interest in affinity maturation<sup>36</sup> it is likely that there are more affinity-improving mutations in the peer-reviewed literature and in SKEMPI than would occur through random mutagenesis. It is thus doubtful that SKEMPI, or our dataset, contains a representative sample of substituted residue types or a representative ratio of affinity increasing vs. affinity reducing mutations. We nonetheless computed the PPV which gives the odds of obtaining  $\Delta\Delta G_{\text{exp}} \leq c_{\text{exp}}$  for  $c_{\text{exp}} = 0, -0.5$  and  $-1.0$  kcal/mol, given  $\Delta\Delta G_{\text{ZEMu}} \leq c_{\text{ZEMu}}$ , for several  $c_{\text{ZEMu}}$ 's. Considering the entire dataset, for  $\Delta\Delta G_{\text{ZEMu}} \leq -0.5$  kcal/mol, the probability of satisfying  $\Delta\Delta G_{\text{exp}} \leq -0.5$  kcal/mol is 0.43. We emphasize strongly that this PPV is not applicable to the case that  $\Delta\Delta G_{\text{ZEMu}}$  is computed for all possible point mutations at an interface, as would be done in a likely practical application.

## Discussion

In prior work<sup>1</sup> we described a means to predict  $\Delta\Delta G$  in crystallographically observed PPIs. A key goal in Structural Bioinformatics is the ability to compute  $\Delta\Delta G$  even for the very common case of proteins whose structure is known crystallographically only in the free form. In many cases evolutionary information<sup>17</sup> does not exist or is not applicable. We formed template models based on an available cocrystallized complex (which in principle could be e.g. isoforms or homologous of the free structures), with a new template modelling protocol that we describe. We here demonstrate that the protocol does not require a high sequence identity for building significant template models based on a homologous template (recall, sequence identity for homologous proteins ranged between 44% and 51%), whereas existing methods work only with high sequence identity templates<sup>17,18</sup>. For low sequence identity, our method is only moderately less precise, again related to the idea that structure is more conserved than sequence<sup>27</sup>. We also assembled protein complexes by flexibly fitting to a 10 Å resolution density map of an isoform protein<sup>21</sup>.

We emphasize that all results labelled “FoldX-only” were obtained on complexes modelled by us (albeit without subsequent ZEMu processing of the mutation sites), providing independent (non-ZEMu) validation of our template-docking protocol.

Globally, our success using template-docked and fitted models suggests that other means of docking under constraints, e.g. using biochemical or bioinformatical data, may lead to comparable results. We further suggest that if experimental  $\Delta\Delta G$  data is available, ZEMu could be used to validate and/or refine the docked, modeled, or fitted structures. This could significantly improve docking results<sup>37</sup> but is important even for template-based modeling or fitting when the constraints come from an isoform or homolog, since the binding mode may not be conserved.

Our main validation dataset consists of template modeled complexes. We used several variations of ZEMu on these complexes and evaluated the accuracy of  $\Delta\Delta G$  prediction. More-sophisticated variants of ZEMu, which flexibilized various spatial regions, had very similar results on the main dataset when compared to the published method<sup>1</sup>, indicating that it is best to limit the flexibility to the immediate sequence neighborhood of the mutated residues. This validates the perturbative assumption underlying ZEMu, namely that the substitution mutations have the largest effect in the near neighborhood of the mutation site, and less effect farther from that position. Anecdotal observations suggest MD potentials introduce structural artifacts, so leaving as much as possible of the crystallographic structure unmodified may be key to ZEMu’s success. We introduce an MMB modeling protocol and show that it leads to a versatile method to predict  $\Delta\Delta G$  on modeled and fitted protein-protein complexes.

## Methods

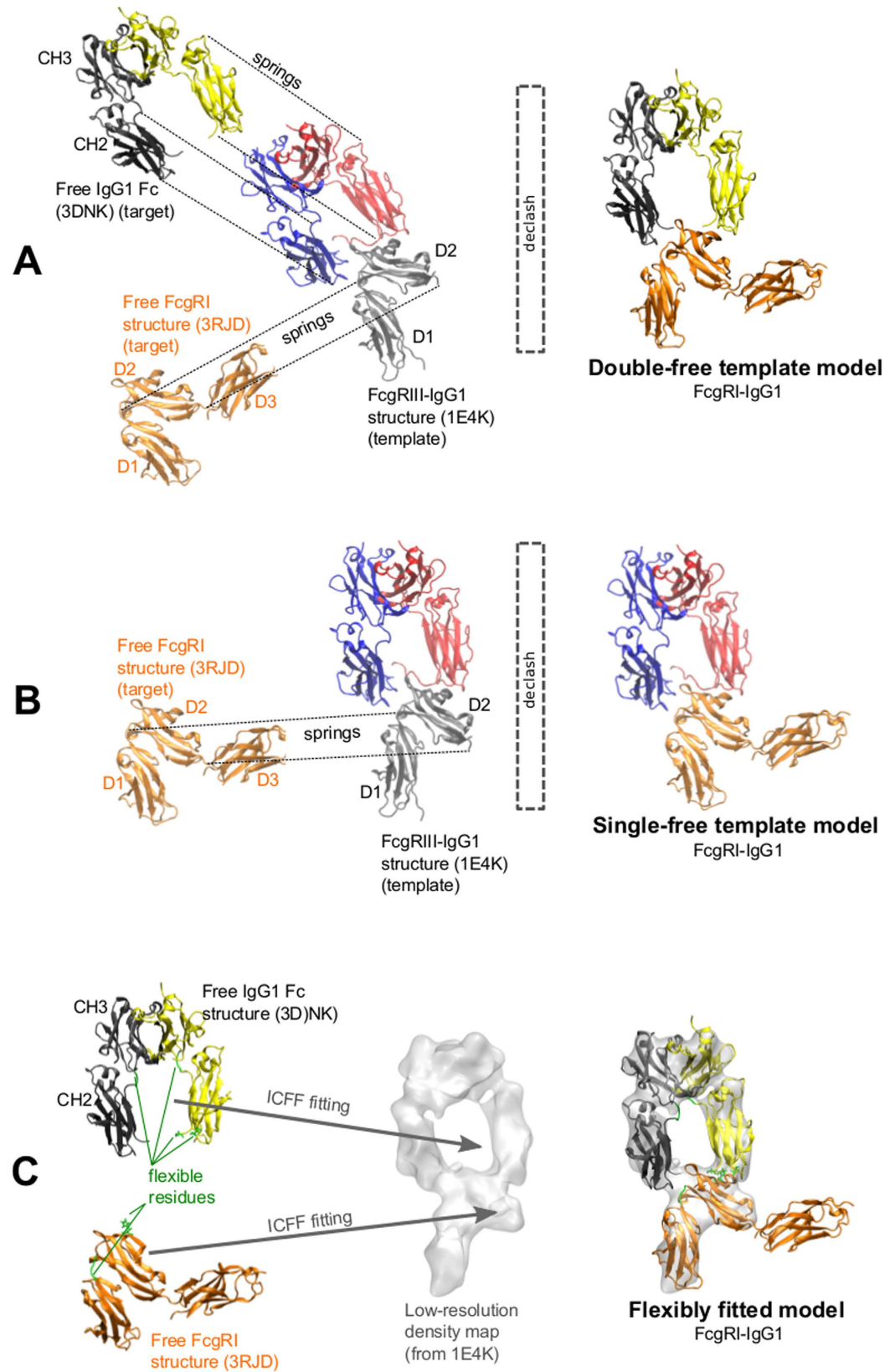
We built models of protein-protein complexes using MMB template modeling to align the free protein structures to relevant protein crystal complexes and also by fitting to a low-resolution density maps using ICF<sup>21</sup> (described below). We then used ZEMu<sup>1</sup> to predict  $\Delta\Delta G$  upon mutation and compared to the results of using FoldX-only<sup>13</sup>. Note that in all cases FoldX-only analysis was conducted with the MMB modeled or fitted structures. ZEMu was mostly used as described in<sup>1-3</sup>. Though we also tested the effect of flexibilizing additional residues in the spatial neighborhood (up to 14 Å) of the mutation sites.

**Dataset.** The validation dataset comprises 846 mutants (each with 1-6 simultaneous substitutions) of 11 different protein-protein complexes models for which  $\Delta\Delta G_{\text{exp}}$  is available (Table 1)<sup>36,38-60</sup>. The dataset consists mostly of double-free template models (two targets). In some cases the templates were the same proteins as the targets (self-templates, Table S1); such systems are useful for benchmarking purposes. In other cases we used homologous templates (Fig. 3A; Table S2). We also created a single-free template model, for which only one of the two targets comes from a free structure, while the other is retained from the template (Fig. 3B; Table S2). Finally, we also generated a model of the biomedically important Fc $\gamma$ RII/IgG1<sup>61</sup> by fitting to a low-resolution synthetic density map of Fc $\gamma$ RIII/IgG1<sup>33</sup> using ICF<sup>21</sup>.

**Template modeling in MMB.** Several good template-based modeling methods exist<sup>62-64</sup>. Our procedure starts with a sequence alignment<sup>65</sup> between unbound (target) and bound (template) target and template residues, followed by structural alignment. We then resolve any steric clashes, after which the model is ready to be used for  $\Delta\Delta G$  prediction with ZEMu or potentially other purposes. The procedure (described in detail below) is straightforward and convenient to do in MMB.

**Structural alignment based on gapped sequence alignment.** We have previously described morphing<sup>21,66</sup> and homology modeling of RNA<sup>20</sup> and ribonucleoprotein complexes<sup>19</sup>, using springs which connect residues in a rigid molecule of known structure, with corresponding residues in a flexible molecule of unknown structure. The mentioned correspondence is determined by a simple SeqAn<sup>65</sup> gapped sequence alignment now available in MMB, with a mismatch and gap opening penalty of  $-1$ . In contrast with our homology modeling work (in which a fully-flexible unstructured target chain is aligned with a rigid template)<sup>19</sup>, or our morphing work (in which a partially-flexible structured target is aligned with a rigid template)<sup>21,66</sup>, here those springs align one or more unbound targets (or their binding domains) onto corresponding domains in the template, (Fig. 3) with both the targets and the template remaining fully rigid during the process. The template is then deleted, leaving the modeled targets in their place (Fig. 3A,B).

**Declashing.** The thus-modeled complex typically has a small number of clashing residues, as the binding interface of the target was previously exposed to solvent allowing greater freedom to the side chains and to some extent the backbone. These clashes must be removed for accurate  $\Delta\Delta G$  calculation (Fig. 3A,B). The problem of side chains that clash due to modeling is not unlike that of side chains that clash due to *in silico* mutagenesis, the problem that is addressed by ZEMu. Therefore we use a ZEMu-like method<sup>1</sup> to declash. We flexibilize only



**Figure 3. Creating complexes with MMB template modeling and ICFE.** As an example, we create an Fc $\gamma$ RI-IgG1 model based on the experimentally observed Fc $\gamma$ RIII-IgG1 complex. (A) A double-free template model is created as follows. We rigidify all chains. The two chains comprising the IgG1 Fc are constrained to each other, for the free (from PDB ID: 3DNK) structure. For the template (1E4K), all chains are constrained to ground. Springs connect the binding domain (D2) of the free Fc $\gamma$ RI (3RJD) to the D2 domain of the template Fc $\gamma$ RIII (1E4K). The springs connect residues which correspond based on a gapped sequence alignment. Similarly, springs connect the two CH2 domains of the free to the CH2 domains of the template IgG1. Once the free

are superimposed on the template proteins, both of the template proteins are deleted. The complex is then declassified (see text), leaving the modeled complex ready for  $\Delta\Delta G$  evaluation. The models in our main dataset were prepared in this way, except for the two complexes highlighted in Table 1. **(B)** A *single-free template model* is prepared as above, except that we have only one target (Fc $\gamma$ RI) and the IgG1 is retained from the template rather than being deleted. **(C)** A fitted model is created using ICFF as described in<sup>21</sup>. We require an electron density map, which may be at low resolution (here we created a simulated map at 10 Å resolution from 1E4K using MDFF). We rigidly and approximately fit both protein structures into the available density map. Then, we allow flexibility in hinge residues (to enable domain motions) as well as certain residues in the protein-protein interface of interest (which would otherwise clash). We activate an MD force field in the spatial neighborhood of the flexible residues. We then turn on forces which push atoms along the density gradient, until the flexible fitting has converged<sup>21</sup>. Note significant domain motions occurred during the fitting process. Fitting is on the basis of all domains which are present in the low-resolution density map (i.e. D3 is not fitted).

the clashing residue or, if needed, a 5-residue zone centered on the main clashing residue, and equilibrate under PARM99<sup>67</sup> interactions with near neighbors. We then minimize the structure under the FoldX potential<sup>13</sup>.

**Flexible fitting to low-resolution density maps.** In addition to the template models described above, we also flexibly fitted free Fc $\gamma$ RI and IgG1 structures into a synthetic 10 Å-resolution density map (based on PDB ID 1E4K) of the Fc $\gamma$ RIII-IgG1 complex. We previously described Internal Coordinate Flexible Fitting (ICFF). In ICFF, each atom in the molecule (or relevant fragment thereof) perceives a force which is proportional to the electronic density gradient at the nuclear position<sup>21</sup>. The molecule in question can have hinge and interface flexibility, and MD forces can be applied about such points of flexibility. Hinges between the D1 and D2 domains on Fc $\gamma$ RI, and between the CH2 and CH3 domains on Fc, were made flexible, to allow domain motions. Interface residues 1236 and 1298 (on Fc) and 148 (on Fc $\gamma$ RI) were granted side-chain flexibility, to relieve clashes which would otherwise occur as the two subunits come into contact to form the complex (Fig. 3C).

**Evaluating  $\Delta\Delta G$  for the modeled complexes.** ZEMu involves first specifying a flexibility zone comprising five residues consecutive in sequence, centered on the mutation site. The flexibility zone is treated in torsion space, leaving the remainder of the protein rigid and fixed. Also, a physics zone of 12 Å around each flexible residue is defined, inside of which electrostatic and van der Waals forces are active. The system is then minimized by dynamics<sup>1</sup>.

The interaction energy of the equilibrated complex is evaluated with the Knowledge Based (KB) potential FoldX<sup>13</sup>. We conduct the calculation for the MMB-equilibrated wild type and mutant complexes to obtain  $\Delta G_{\text{wild type}}$  and  $\Delta G_{\text{mutant}}$ , respectively. An estimate of  $\Delta\Delta G_{\text{exp}}$  is obtained as follows<sup>68</sup>:

$$\Delta\Delta G_{\text{ZEMu}} = \Delta G_{\text{mutant}} - \Delta G_{\text{wild type}} \approx \Delta\Delta G_{\text{exp}}$$

**ZEMu+ additional flexibility radius.** We also tested a variation of ZEMu which grants flexibility to residues in the spatial neighborhood of the mutation sites. This was defined as all residues within a certain radius of the mutation site, potentially including residues of the adjacent protein. We evaluated radii ranging from 0 (equivalent to ordinary ZEMu) to 14 Å. After equilibration we evaluated  $\Delta\Delta G$  with FoldX<sup>13</sup> as before.

## References

- Dourado, D. F. & Flores, S. C. A multiscale approach to predicting affinity changes in protein-protein interfaces. *Proteins*, doi: 10.1002/prot.24634 (2014).
- Fiesel, F. C. *et al.* Structural and Functional Impact of Parkinson Disease-Associated Mutations in the E3 Ubiquitin Ligase Parkin. *Hum Mutat* **36**, 774–786, doi: 10.1002/humu.22808 (2015).
- Caulfield, T. R. *et al.* Phosphorylation by PINK1 releases the UBL domain and initializes the conformational opening of the E3 ubiquitin ligase Parkin. *PLoS Comput Biol* **10**, e1003935, doi: 10.1371/journal.pcbi.1003935 (2014).
- Kazlauskas, R. J. & Bornscheuer, U. T. Finding better protein engineering strategies. *Nat Chem Biol* **5**, 526–529, doi: 10.1038/nchembio0809-526 (2009).
- Levin, A. M. & Weiss, G. A. Optimizing the affinity and specificity of proteins with molecular display. *Mol Biosyst* **2**, 49–57, doi: 10.1039/b511782h (2006).
- Getts, D. R., Getts, M. T., McCarthy, D. P., Chastain, E. M. L. & Miller, S. D. Have we overestimated the benefit of human(ized) antibodies? *Mabs-Austin* **2**, 682–694 doi:DOI 10.4161/mabs.2.6.13601 (2010).
- Marelius, J., Kolmodin, K., Feierberg, I. & Aqvist, J. Q: a molecular dynamics program for free energy calculations and empirical valence bond simulations in biomolecular systems. *J Mol Graph Model* **16**, 213–225, 261 (1998).
- Massova, I. & Kollman, P. A. Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding. *Perspect Drug Discov* **18**, 113–135, doi: 10.1023/A:1008763014207 (2000).
- Jorgensen, W. L. Free-Energy Calculations - a Breakthrough for Modeling Organic-Chemistry in Solution. *Accounts Chem Res* **22**, 184–189, doi: 10.1021/Ar00161a004 (1989).
- Kollman, P. Free-Energy Calculations - Applications to Chemical and Biochemical Phenomena. *Chem Rev* **93**, 2395–2417, doi: 10.1021/Cr00023a004 (1993).
- Dehouck, Y., Kwasigroch, J. M., Rooman, M. & Gilis, D. BeAtMuSiC: prediction of changes in protein-protein binding affinity on mutations. *Nucleic Acids Res* **41**, W333–339, doi: 10.1093/nar/gkt450 (2013).
- Douglas, E. V., Pires, D. B. A., Tom, L. & Blundell, mCSM: predicting the effects of mutations in proteins using graph-based signatures. *Bioinformatics* **30**, 335–342, doi: 10.1093/bioinformatics/btt691 (2014).
- Guerois, R., Nielsen, J. E. & Serrano, L. Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *J Mol Biol* **320**, 369–387, doi: 10.1016/S0022-2836(02)00442-4 (2002).
- Kortemme, T. & Baker, D. A simple physical model for binding energy hot spots in protein-protein complexes. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 14116–14121, doi: 10.1073/Pnas.202485799 (2002).
- Kamisetty, H., Ramanathan, A., Bailey-Kellogg, C. & Langmead, C. J. Accounting for conformational entropy in predicting binding free energies of protein-protein interactions. *Proteins-Structure Function and Bioinformatics* **79**, 444–462, doi: 10.1002/Prot.22894 (2011).



16. Smith, C. A. & Kortemme, T. Backrub-like backbone simulation recapitulates natural protein conformational variability and improves mutant side-chain prediction. *Journal of Molecular Biology* **380**, 742–756, doi: 10.1016/j.jmb.2008.05.023 (2008).
17. Berliner, N., Teyra, J., Colak, R., Garcia Lopez, S. & Kim, P. M. Combining structural modeling with ensemble machine learning to accurately predict protein fold stability and binding affinity effects upon mutation. *PLoS One* **9**, e107353, doi: 10.1371/journal.pone.0107353 (2014).
18. Brender, J. R. & Zhang, Y. Predicting the Effect of Mutations on Protein-Protein Binding Interactions through Structure-Based Interface Profiles. *PLoS Comput Biol* **11**, e1004494, doi:10.1371/journal.pcbi.1004494 (2015).
19. Flores, S., Zemora, G. & Waldsich, C. Insights into diseases of human telomerase from dynamical modeling. *Pacific Symposium on Biocomputing* **18**, 200–211 (2013).
20. Flores, S., Sherman, M., Bruns, C., Eastman, P. & Altman, R. Fast flexible modeling of macromolecular structure using internal coordinates. *IEEE Transactions in Computational Biology and Bioinformatics* **8**, 1247–1257 (2010).
21. Flores, S. C. Fast fitting to low resolution density maps: elucidating large-scale motions of the ribosome. *Nucleic Acids Res* **42**, e9, doi: 10.1093/nar/gkt906 (2014).
22. Hu, Z. *et al.* Structural and biochemical basis for induced self-propagation of NLRC4. *Science* **350**, 399–404, doi: 10.1126/science.aac5489 (2015).
23. Nogales, E. & Scheres, S. H. W. Cryo-EM: A Unique Tool for the Visualization of Macromolecular Complexity. *Molecular Cell* **58**, 677–689, doi: 10.1016/j.molcel.2015.02.019 (2015).
24. Svergun, D. I. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing (vol 76, pg 2879, 1999). *Biophysical Journal* **77**, 2896–2896 (1999).
25. Chapman, H. N. *et al.* Femtosecond diffractive imaging with a soft-X-ray free-electron laser. *Nature Physics* **2**, 839–843, doi: 10.1038/nphys461 (2006).
26. Sinha, R., Kundrotas, P. J. & Vakser, I. A. Docking by structural similarity at protein-protein interfaces. *Proteins-Structure Function and Bioinformatics* **78**, 3235–3241, doi: 10.1002/prot.22812 (2010).
27. Kundrotas, P. J., Zhu, Z., Janin, J. & Vakser, I. A. Templates are available to model nearly all complexes of structurally characterized proteins. *P Natl Acad Sci USA* **109**, 9438–9441, doi: 10.1073/pnas.1200678109 (2012).
28. Chothia, C. & Lesk, A. M. The Relation between the Divergence of Sequence and Structure in Proteins. *Embo Journal* **5**, 823–826 (1986).
29. Kundrotas, P. J. & Vakser, I. A. Global and local structural similarity in protein-protein complexes: implications for template-based docking. *Proteins* **81**, 2137–2142, doi: 10.1002/prot.24392 (2013).
30. Murakami, H. *et al.* Phase I study of ganitumab (AMG 479), a fully human monoclonal antibody against the insulin-like growth factor receptor type I (IGF1R), in Japanese patients with advanced solid tumors. *Cancer Chemother Pharmacol* **70**, 407–414, doi: 10.1007/s00280-012-1924-9 (2012).
31. Walsh, S. T. & Kossiakoff, A. A. Crystal structure and site 1 binding energetics of human placental lactogen. *J Mol Biol* **358**, 773–784, doi: 10.1016/j.jmb.2006.02.038 (2006).
32. Ramsland, P. A. *et al.* Structural basis for Fc gammaRIIIa recognition of human IgG and formation of inflammatory signaling complexes. *J Immunol* **187**, 3208–3217, doi: 10.4049/jimmunol.1101467 (2011).
33. Sondermann, P., Huber, R., Oosthuizen, V. & Jacob, U. The 3.2-Å crystal structure of the human IgG1 Fc fragment-Fc gammaRIII complex. *Nature* **406**, 267–273, doi: 10.1038/35018508 (2000).
34. Lu, J. *et al.* Structure of Fc gammaRI in complex with Fc reveals the importance of glycan recognition for high-affinity IgG binding. *P Natl Acad Sci USA* **112**, 833–838, doi: 10.1073/pnas.1418812112 (2015).
35. Rost, B. Twilight zone of protein sequence alignments. *Protein Eng* **12**, 85–94 (1999).
36. Lippow, S. M., Witttrup, K. D. & Tidor, B. Computational design of antibody-affinity improvement beyond *in vivo* maturation. *Nat Biotechnol* **25**, 1171–1176, doi: 10.1038/nbt1336 (2007).
37. Moreira, I. S., Martins, J. M., Coimbra, J. T., Ramos, M. J. & Fernandes, P. A. A new scoring function for protein-protein docking that identifies native structures with unprecedented accuracy. *Phys Chem Chem Phys* **17**, 2378–2387, doi: 10.1039/c4cp04688a (2015).
38. Lim, D. *et al.* Crystal structure and kinetic analysis of beta-lactamase inhibitor protein-II in complex with TEM-1 beta-lactamase. *Nat Struct Biol* **8**, 848–852, doi: 10.1038/nsb1001-848 (2001).
39. Petrosino, J., Rudgers, G., Gilbert, H. & Palzkill, T. Contributions of aspartate 49 and phenylalanine 142 residues of a tight binding inhibitory protein of beta-lactamases. *J Biol Chem* **274**, 2394–2400, doi: 10.1074/jbc.274.4.2394 (1999).
40. Albeck, S., Unger, R. & Schreiber, G. Evaluation of direct and cooperative contributions towards the strength of buried hydrogen bonds and salt bridges. *J Mol Biol* **298**, 503–520, doi: 10.1006/jmbi.2000.3656 (2000).
41. Selzer, T., Albeck, S. & Schreiber, G. Rational design of faster associating and tighter binding protein complexes. *Nat Struct Biol* **7**, 537–541 (2000).
42. Reichmann, D. *et al.* The modular architecture of protein-protein binding interfaces. *P Natl Acad Sci USA* **102**, 57–62, doi: 10.1073/pnas.0407280102 (2005).
43. Braden, B. C. *et al.* 3-Dimensional Structures of the Free and the Antigen-Complexed Fab from Monoclonal Antilysozyme Antibody-D44.1. *J Mol Biol* **243**, 767–781, doi: 10.1016/0022-2836(94)90046-9 (1994).
44. Bhat, T. N. *et al.* Bound water molecules and conformational stabilization help mediate an antigen-antibody association. *P Natl Acad Sci USA* **91**, 1089–1093 (1994).
45. Dall'Acqua, W., Goldman, E. R., Eisenstein, E. & Mariuzza, R. A. A mutational analysis of the binding of two different proteins to the same antibody. *Biochemistry-Us* **35**, 9667–9676, doi: 10.1021/Bi960819i (1996).
46. Dall'Acqua, W. *et al.* A mutational analysis of binding interactions in an antigen-antibody protein-protein complex. *Biochemistry-Us* **37**, 7981–7991, doi: 10.1021/bi980148j (1998).
47. De Crescenzo, G. *et al.* Three key residues underlie the differential affinity of the TGFbeta isoforms for the TGFbeta type II receptor. *J Mol Biol* **355**, 47–62, doi: 10.1016/j.jmb.2005.10.022 (2006).
48. Hart, P. J. *et al.* Crystal structure of the human TbetaR2 ectodomain-TGF-beta3 complex. *Nat Struct Biol* **9**, 203–208, doi: 10.1038/nsb766 (2002).
49. Deisenhofer, J. 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-Å resolution. *Biochemistry-Us* **20**, 2361–2370 (1981).
50. Bottomley, S. P. *et al.* The Stability and Unfolding of an Igg Binding-Protein Based Upon the B-Domain of Protein-a from *Staphylococcus-Aureus* Probed by Tryptophan Substitution and Fluorescence Spectroscopy. *Protein Eng* **7**, 1463–1470, doi: 10.1093/Protein/7.12.1463 (1994).
51. Cedergren, L., Andersson, R., Jansson, B., Uhlen, M. & Nilsson, B. Mutational Analysis of the Interaction between Staphylococcal Protein-a and Human Igg(1). *Protein Eng* **6**, 441–448, doi: 10.1093/Protein/6.4.441 (1993).
52. Jendeberg, L. *et al.* Kinetic-Analysis of the Interaction between Protein-a Domain Variants and Human Fc Using Plasmon Resonance Detection. *J Mol Recognit* **8**, 270–278, doi: 10.1002/jmr.300080405 (1995).
53. Li, Y. L., Li, H. M., Smith-Gill, S. J. & Mariuzza, R. A. Three-dimensional structures of the free and antigen-bound Fab from monoclonal antilysozyme antibody HyHEL-63. *Biochemistry-Us* **39**, 6296–6309, doi: 10.1021/Bi000054l (2000).
54. Li, Y. L., Urrutia, M., Smith-Gill, S. J. & Mariuzza, R. A. Dissection of binding interactions in the complex between the anti-lysozyme antibody HyHEL-63 and its antigen. *Biochemistry-Us* **42**, 11–22, doi: 10.1021/Bi020589+ (2003).

55. Pelletier, H. & Kraut, J. Crystal-Structure of a Complex between Electron-Transfer Partners, Cytochrome-C Peroxidase and Cytochrome-C. *Science* **258**, 1748–1755, doi: 10.1126/Science.1334573 (1992).
56. Pielak, G. J. & Wang, X. M. Interactions between yeast iso-1-cytochrome c and its peroxidase. *Biochemistry-Us* **40**, 422–428, doi: 10.1021/Bi002124u (2001).
57. Shields, R. L. *et al.* High resolution mapping of the binding site on human IgG1 for Fc gamma RI, Fc gamma RII, Fc gamma RIII, and FcRn and design of IgG1 variants with improved binding to the Fc gamma R. *J Biol Chem* **276**, 6591–6604, doi: 10.1074/jbc.M009483200 (2001).
58. Tamm, A., Kister, A., Nolte, K. U., Gessner, J. E. & Schmidt, R. E. The IgG binding site of human Fc gamma RIIIb receptor involves CC' and FG loops of the membrane-proximal domain. *J Biol Chem* **271**, 3659–3666 (1996).
59. Hibbs, M. L., Tolvanen, M. & Carpen, O. Membrane-proximal Ig-like domain of Fc gamma RIII (CD16) contains residues critical for ligand binding. *J Immunol* **152**, 4466–4474 (1994).
60. Richards, J. O. *et al.* Optimization of antibody binding to Fc gamma RIIa enhances macrophage phagocytosis of tumor cells. *Mol Cancer Ther* **7**, 2517–2527, doi: 10.1158/1535-7163.MCT-08-0201 (2008).
61. Lu, J., Ellsworth, J. L., Hamacher, N., Oak, S. W. & Sun, P. D. Crystal structure of Fc gamma receptor I and its implication in high affinity gamma-immunoglobulin binding. *J Biol Chem* **286**, 40608–40613, doi: 10.1074/jbc.M111.257550 (2011).
62. Zhang, Y. Template-based modeling and free modeling by I-TASSER in CASP7. *Proteins-Structure Function and Bioinformatics* **69**, 108–117, doi: 10.1002/prot.21702 (2007).
63. Zhang, Y. & Skolnick, J. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Research* **33**, 2302–2309, doi: 10.1093/nar/gki524 (2005).
64. Tuncbag, N., Keskin, O., Nussinov, R. & Gursoy, A. Fast and accurate modeling of protein-protein interactions by combining template-interface-based docking with flexible refinement. *Proteins* **80**, 1239–1249, doi: 10.1002/prot.24022 (2012).
65. Kehr, B., Weese, D. & Reinert, K. STELLAR: fast and exact local alignments. *BMC Bioinformatics* **12**, doi: 10.1186/1471-2105-12-S9-S15 (2011).
66. Tek, A., Korostelev, A. & Flores, S. C. MMB-GUI: a fast morphing method demonstrates a possible ribosomal tRNA translocation trajectory. *Nucleic Acids Res.* **44**(1), 95–105, doi: 10.1093/nar/gkv1457 (2015).
67. Cornell, W. D. *et al.* A second generation force field for the simulation of proteins, nucleic acids, and organic molecules (vol 117, pg 5179, 1995). *J Am Chem Soc* **118**, 2309–2309, doi: 10.1021/Ja955032e (1996).
68. Brandsdal, B. O., Aqvist, J. & Smalas, A. O. Computational analysis of binding of P1 variants to trypsin. *Protein Sci* **10**, 1584–1595, doi: 10.1110/ps.940101 (2001).

## Acknowledgements

The authors acknowledge funding from eSENCE (essenceofscience.se), The Swedish Foundation for International Cooperation in Research and Higher Education (STINT), and Uppsala University, and a generous allocation of supercomputer time from the Swedish National Infrastructure for Computing (SNIC).

## Author Contributions

S.F. conceived the idea and designed the methodology. D.D. compiled the dataset and validated the methodology. Both authors wrote the paper.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Dourado, D. F.A.R. and Flores, S. C. Modeling and fitting protein-protein complexes to predict change of binding energy. *Sci. Rep.* **6**, 25406; doi: 10.1038/srep25406 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>