

Computational Alanine Scanning of Protein-Protein Interfaces

Tanja Kortemme,^{1,2*} David E. Kim,² and David Baker^{2*}

(Published 10 February 2004)

INTRODUCTION

MATERIALS

EQUIPMENT

INSTRUCTIONS

Preparation of Input File
Submission of Input Data

TROUBLESHOOTING

RELATED TECHNIQUES

NOTES AND REMARKS

REFERENCES

¹Department of Biopharmaceutical Sciences and California Institute for Quantitative Biomedical Research, University of California, San Francisco, CA 94107, USA. ²Howard Hughes Medical Institute and Department of Biochemistry, University of Washington, Box 357350, Seattle, WA 98195, USA.

*Corresponding authors. Telephone, +1-206-543-1295; fax +1-206-685-1792; e-mail, dabaker@u.washington.edu (D.B.); e-mail, kortemme@u.washington.edu (T.K.)

Abstract

Protein-protein interactions are key components of all signal transduction processes, so methods to alter these interactions promise to become important tools in dissecting function of connectivities in these networks. We have developed a fast computational approach for the prediction of energetically important amino acid residues in protein-protein interfaces (available at <http://robetta.bakerlab.org/alaninescan>), which we, following Peter Kollman, have termed “computational alanine scanning.” The input consists of a three-dimensional structure of a protein-protein complex; output is a list of “hot spots,” or amino acid side chains that are predicted to significantly destabilize the interface when mutated to alanine, analogous to the results of experimental alanine-scanning mutagenesis. A total of 79% of hot spots and 68% of neutral residues were correctly predicted in a test of 233 mutations in 19 protein-protein complexes. A single interface can be analyzed in minutes. The computational methodology has been validated by the successful design of protein interfaces with new specificity and activity and has yielded new insights into the mechanisms of receptor specificity and promiscuity in biological systems.

Introduction

Protein-protein interactions are key components of all signal transduction processes, mediating the integration of linear pathways into the often complex interaction networks revealed by recent genome-scale studies (1). Tools to rationally alter and interfere with protein interactions offer great promise to help dissect the function of connectivities in these networks. The ability to alter protein interactions requires an understanding of the determinants of affinity and specificity in protein interfaces.

Experimental “alanine-scanning mutagenesis” is a powerful method for analyzing important interactions in protein-protein interfaces. Alanine scanning measures the effect of the deletion of an amino acid side chain beyond the C_β carbon atom on the affinity of a protein-protein complex. Individual substitutions of many amino acids with alanine yield a map of which interactions are critical in an interface and which ones are not (Fig. 1). Clackson and Wells (2) called these energetically important residues “hot spots” in their pioneering work on the binding of human growth hormone to its receptor, where only a small fraction of interface residues account for the majority of the binding energy. Subsequent studies suggested that the presence of binding energy hot spots comprising only a fraction of the complete interface area is a general property of most protein-protein complexes (3).

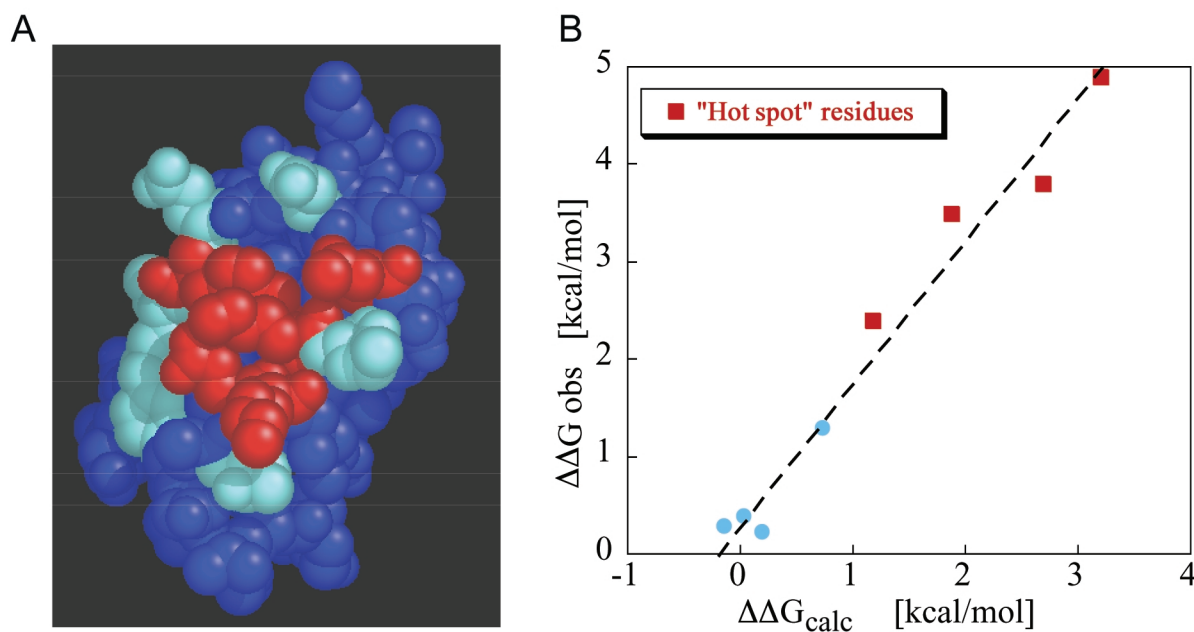


Fig. 1. Results of computational alanine scanning on the interface of the protein G B1 domain with an IgG Fc fragment (20). **(A)** View of the protein G interface contacting the Fc fragment, color coded according to the computed effect of alanine mutations on the binding free energy. Red, hot-spot residues with a computed effect on the binding free energy if mutated to alanine of more than 1 kcal/mol; light blue, neutral residues with a small predicted effect on the binding free energy. **(B)** Agreement of observed and predicted effects of alanine mutations on the binding free energy.

Although alanine-scanning mutagenesis can be scaled up by phage-display library techniques (4), it still represents a significant experimental effort that cannot easily be applied to a high-throughput analysis of protein-protein interactions. To answer this need, we have developed a “computational alanine-scanning” protocol [(Fig. 2) (5)] that, if a high-resolution structure of the protein-protein complex in question is available, allows the automatic scanning of a complete protein-protein interface within minutes on a single Linux PC processor. Computational alanine scanning uses a simple free energy function to calculate the effects of alanine mutations on the binding free energy of a protein-protein complex. The function consists of a linear combination of a Lennard-Jones potential to describe atomic packing interactions, an implicit solvation model (6), an orientation-dependent hydrogen-bonding potential derived from high-resolution protein structures (7), statistical terms approximating the backbone-dependent amino acid-type and rotamer probabilities (8), and an estimate of unfolded reference state energies [(Eq. 1) (5)]. In a test set of 19 protein-protein complexes with 233 mutations, 79% of the energetic “hot spots” (2) (defined as an experimentally observed change in binding free energy upon alanine substitution of more than 1 kcal/mol) were identified by the free energy function (5).’

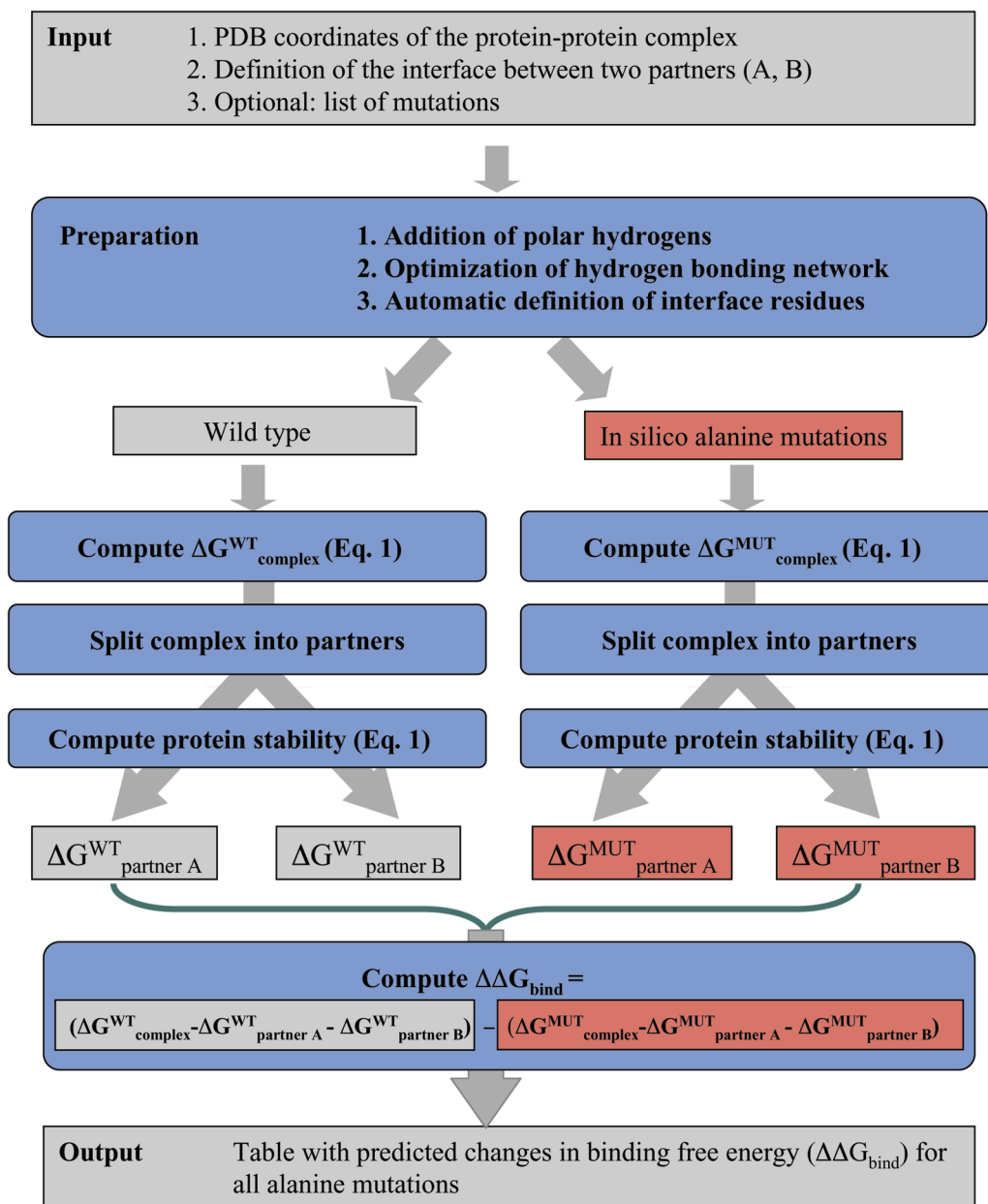


Fig. 2. Flow chart illustrating the computational alanine scanning procedure.

$$\Delta G = W_{attr} E_{LJattr} + W_{rep} E_{LJrep} + W_{HB(sc-bb)} E_{HB(sc-bb)} + W_{HB(sc-sc)} E_{HB(sc-sc)} + W_{sol} G_{sol} + W_{\phi/\psi} E_{\phi/\psi}(aa) + \sum_{aa=1}^{20} n_{aa} E_{aa}^{ref}$$

E_{LJattr}	Attractive part of a Lennard-Jones potential
E_{LJrep}	A linear distance-dependent repulsive term
$E_{HB(sc-bb)}$	Orientation-dependent side chain-backbone hydrogen bond potential
$E_{HB(sc-sc)}$	Orientation-dependent side chain-side chain hydrogen bond potential
G_{sol}	Implicit solvation model
W	Relative weights of the different energy terms [for details on the parameterization see (5)].
$E_{\phi/\psi}(aa)$	Amino-acid type (aa) dependent backbone torsion angle propensity
E_{aa}^{ref}	Amino-acid type dependent reference energy, which approximates the interactions made in the unfolded state ensemble (8) (n_{aa} is the number of amino acids of a certain type); the last two terms were included to model changes in protein stability upon mutation but do not contribute to free energy changes of protein protein interactions.

Eq. 1. Simple free energy function for calculating effects of alanine mutation on binding free energy of a protein-protein complex.

If computational alanine scanning is applied to a complex structure without existing experimental data, the algorithm automatically identifies all interface residues in a protein-protein interface. An interface residue is defined as (i) a residue that has at least one atom within a sphere with a 4 Å radius of an atom belonging to the other partner in the protein complex, or (ii) a residue that becomes significantly buried upon complex formation, as measured by an increase in the number of C_{β} atoms within a sphere with a radius of 8 Å around the C_{β} atom of the residue of interest. The program then replaces each of the interface residues individually with alanine residues, and computes the effect of this mutation on the binding free energy of the complex. An example of the results of computational alanine scanning is given in Table 1.

Computational analysis of binding-energy hot spots in protein-protein interfaces is useful for many applications. In the simplest case, it can suggest mutations likely to dramatically affect the binding affinity of a given protein-protein complex. Conversely, it can help the annotation of single nucleotide polymorphisms that might have functional consequences by affecting protein-mediated interactions. In fact, the speed of the computational protocol makes large-scale applications to all protein-protein interfaces with known structures feasible. Identification of hot spots provides a starting point for the design of protein-protein interactions. For example, we used hot-spot analysis of interfaces between protein domains to create a novel chimeric protein by fusing two related protein domains and creating a new interface between them (9).

Finally, the comparative analysis of hot spots in homologous interfaces can reveal determinants of specificity and cross-reactivity. We applied computational alanine scanning to multiple sets of interactions formed by the NKG2D immunoreceptor with major histocompatibility (MHC) class I-like ligands (10) and to the gp130 shared signaling receptor with divergent cytokines (11). In both cases, the receptor recognizes multiple ligands through the same interface with virtually the same conformation of all contacting residues. Hot-spot analysis, together with focused experiments confirming the predictions, helped explain this promiscuity. In the case of the NKG2D immunoreceptor, the ligands use different strategies to recognize a conserved hot spot on the receptor (10). In contrast, different hot spots on gp130 are utilized by its divergent binding partners that each satisfy some, but not all, of the binding capacities in the receptor interface (11). It appears that the energetically important residues in this interface display some specificity that might be exploited for the design of specific interaction modulators by targeting different hot-spot regions of the interface (2).

Materials

Structure of the protein complex in question, in Protein Data Bank (pdb) format (see http://www.rcsb.org/pdb/info.html#File_Formats_and_Standards)

Experimental data on the effect of alanine mutations on the binding free energy of the protein protein complex in question

Note: This is optional.

Equipment

Internet access

Note: No special computer equipment is needed; all calculations can be performed with the computational alanine scanning Web server (<http://rosetta.bakerlab.org/alaninescan>).

RasMol or similar structure display program (<http://www.openrasmol.org/>)

Instructions

The following instructions explain the file formats and the submission interface that users will encounter when submitting a protein interaction for computational alanine scanning. Users can enter information about known mutations if that information is available or can allow the program to automatically identify all of the interface residues to be virtually mutated and the effects determined by computation.

An example of the results of a computational alanine scan is given in Table 1, with the column descriptors given in the Table legend. Column 6 contains the predicted changes in binding free energy ($\Delta\Delta G_{\text{bind}}$) upon alanine mutation. Positive values mean that replacement by alanine is predicted to destabilize the complex; negative values predict a stabilizing effect. Column 8 shows the predicted stabilizing or destabilizing effect of the alanine mutation on the mutated protein complex partner in isolation.

virtual alanine scanning, 1fcc.alascan							
pdb#	chain	int_id	res#	aa	$\Delta\Delta G(\text{bind})$	$\Delta\Delta G(\text{bind, obs})$	$\Delta G(\text{partner})$
25	C	0	231	17	0.10	0.24	1.06
27	C	1	233	4	3.21	4.90	0.16
28	C	1	234	9	0.99	1.30	-0.53
31	C	1	237	9	1.91	3.50	0.68
35	C	1	241	12	1.18	2.40	-0.18
40	C	1	246	3	-0.15	0.30	-0.61
42	C	1	248	4	0.03	0.40	-0.24
43	C	1	249	19	2.71	3.80	3.74

Table 1. Example of results of computational alanine scanning on the structure 1FCC.pdb. the columns contain the following information. **(Column 1)** pdb#, number of mutated residue in the pdb file; **(Column 2)** chain, pdb chain identifier; **(Column 3)** int_id: measure of whether a residue side chain atom is within 4 Å of an atom on the other partner (1) or not contacting directly, but buried upon binding (0); **(Column 4)** res#, continuous residue numbering of all partners, starting with residue number 1; **(Column 5)** aa, amino acid type according to one-residue nomenclature in alphabetical order (1-A, 2-C, 3-D, 4-E ...); **(Column 6)** $\Delta\Delta G(\text{bind})$, predicted change in binding free energy upon alanine mutation, computed according to scheme in Fig. 2; **(Column 7)** $\Delta\Delta G(\text{bind, obs})$, observed changes in binding free energy upon alanine mutation (user input in mutation list, otherwise set to zero); and **(Column 8)** $\Delta G(\text{partner})$, predicted change in protein stability of the mutated complex partner upon alanine mutation, computed according to Eq. 1.

Hot-spot residues can be defined operationally as those for which alanine mutations have destabilizing effects on $\Delta\Delta G_{\text{bind}}$ of more than 1 kcal/mol. For a comparison with experimental data, a correctly identified hot spot means a residue with a predicted and observed $\Delta\Delta G_{\text{bind}}$ value of greater than or equal to 1 kcal/mol, and a correctly identified neutral residue has both predicted and observed $\Delta\Delta G_{\text{bind}}$ values of less than 1 kcal/mol (alanine substitutions with experimentally measured stabilizing effects are rare and not larger than -0.8 kcal/mol; these are included in the neutral category). By these criteria, in a test set of 233 mutations in 19 protein-protein complexes, 79% of all hot spots in the interface were correctly predicted, as were 68% of all neutral residues in the interface (5).

The quantitative agreement between predicted and observed $\Delta\Delta G_{\text{bind}}$ values is less precise. The average unsigned error was 1.06 kcal/mol for the 233 alanine mutations in protein-protein interfaces, and 0.81 kcal/mol for the effects of alanine mutation on the stability of globular proteins (743 mutations). In general, predictions for aromatic and hydrophobic amino acid residues are more reliable than for charged amino acids, for which inaccuracies of the energy function modeling electrostatic and solvation effects are particularly severe. For further information in the quantitative interpretation of the results, see the Notes and Remarks section, below.

Some caveats are associated with the interpretation of the experimental results; see (12) for an excellent discussion of this topic.

Preparation of Input Data

1. Obtain the coordinates of the structure or structural model of the protein-protein complex in question in Protein Data Bank (pdb) format.

Note: Interactions involving nonprotein components of the protein-protein complex (bound metal ions, lipids, glycosaccharides, nucleotides and nucleic acids, other cofactors) cannot currently be modeled. Also, amino acids not comprising the 20 naturally occurring amino acids, such as selenomethione, are currently ignored.

2. Identify a binary protein-protein interface, for example, by visual inspection using a display program for pdb files such as RasMol.

Note: Each of the two binding interfaces, or “partners,” can consist of several peptide chains.

3. Create a file with a list of mutations to be virtually scanned (see explanations on the Web site for details on the file format).

Note: This step is optional. If it is not available, the program will automatically identify all interface residues to be scanned.

Submission of Input Data

1. Go to <http://robetta.bakerlab.org/alaninescan> (Fig. 3).
2. Upload the structure or structural model (in pdb format) into the browser.
3. Define the protein interface: List all relevant polypeptide chains of the protein complex structure in the corresponding input fields on the web site (Fig. 3) and indicate for each chain to which partner in the protein complex it belongs.
4. Upload a list of alanine mutations with experimentally determined values into the browser.

Note: This step is optional. If experimental values are not available, leave this field blank.

5. To perform an automatic scan of all interface residues, leave the mutation list blank.
6. Click “Submit.”
7. Identify residues that when mutated destabilize the complex by their positive values greater than 1kcal/mol in column 6 of the output table (Table 1).

Note: The results will be e-mailed to you and will also be available on the Web site.

Fig. 3. Screen shot of the computational interface alanine scanning server.

Troubleshooting

In case of problems with the computational alanine scanning Web server, please e-mail the authors at kortemme@u.washington.edu or dabaker@u.washington.edu. “Frequently Asked Questions” and “Known Problems” lists will be maintained and updated on the server.

Related Techniques

A number of published reports describe the calculation of the effect of mutations on the binding free energy of specific protein-protein complexes (13–16). Guerois and Serrano used an approach similar to ours and also compared their computational results to experimental data on a large number of mutations on monomeric proteins and protein-protein complexes (16). A more sophisticated approach to the calculation of changes in binding free energy upon alanine mutation are molecular mechanics-Poisson Boltzmann surface area (MM-PBSA) methods (14, 15), which are computationally much more expensive but can give a more accurate estimation of long-range electrostatic effects in proteins. There is a good agreement of the results between the MM-PBSA energy function and our simple energy model for the interface between mouse double minute 2 (mdm2) and a peptide derived from p53 (5, 14), with a correlation coefficient of 0.96 for 10 alanine mutations.

Notes and Remarks

Several assumptions are made in our implementation of computational alanine scanning that affect the applicability of the method and the interpretation of the results.

First, the terms in the energy function are pairwise additive. This simplification is necessary for a fast computational evaluation of residue-residue energies used in protein design methods. Because of this approximation, coupling effects between different mutations cannot be taken into account, and multiple mutations are always assumed to be additive. Thus, experimental data on nonadditivity effects observed in double-mutant cycles (17) cannot be reproduced by these methods. For the same reason, indirect effects on the binding energy exerted by residues not making direct interactions in the interface are generally not captured [but see the discussion of environment-dependent effects of the energy function below].

Second, the isolated partners in the protein-protein complex have the same bound and unbound conformations during the calculations; it is assumed that there are no conformational changes upon binding. Consequently, intramolecular interactions are equivalent in the isolated complex partners and in the complex, and their contribution cancels out in the calculation of the binding energy (Fig. 2). The only exceptions are electrostatic interactions that are considered to be dependent on their environment. For example, an intramolecular hydrogen bond can become more buried in the complex than in the monomer, and is therefore predicted to be stronger, according to the environment-dependent energy function. Thus, if one of the side chains participating in the hydrogen bond is mutated to alanine, there will be a net destabilizing effect on the binding free energy even if the hydrogen bonding interaction is not intermolecular [see (5) for a more detailed discussion of the energy function]. For the same reason, residues not directly participating in contacts across the interface can affect the binding free energy by changing the environment of hydrogen-bonding interactions.

Third, all structures are viewed as static; possible changes in side-chain and backbone mobility of the interacting proteins are not taken into account. This approximation is related to the second point above, because binding can lead to both structural as well as dynamical changes. Ignoring these effects can lead to errors, especially where disordered regions become ordered upon binding [some protein parts may also become more flexible upon binding (18)]. The entropy gains and losses associated with these effects can significantly affect the binding free energy.

Fourth, cofactors, metal ions, hydrogen-bonding water molecules bridging side-chains in the protein interface, or other nonpeptide ligands or binding partners (such as nucleic acids) are not taken into account. This approximation will lead to an underestimation of the energetic effects of mutating residues contacting these nonpeptide ligands. An example is given in Fig. 1B in (5) for amino acid residues in the barnase-barstar interface forming water-mediated interactions that are not modeled accurately using the implicit solvation model included in our simple free energy function. Explicitly modeling water molecules observed in the crystal structure improves the accuracy the predictions in this case (19).

Fifth, symmetry in a protein-protein complex is not taken into account. During computational alanine scanning, only one residue at a time is considered. Therefore, in a symmetrical interface with one mutated position, the corresponding residue on the other partner in a dimer is modeled as wild type. However, because of the assumption of additivity, the single mutations of corresponding residues can be added to a first approximation (although this will ignore the environment-dependent effects described in the third point above).

References

1. A. H. Tong , B. Drees, G. Nardelli, G. D. Bader, B. Brannetti, L. Castagnoli, M. Evangelista, S. Ferracuti, B. Nelson, S. Paoluzi, M. Quondam, A. Zucconi, C. W. Hogue, S. Fields, C. Boone, G. Cesareni, A combined experimental and computational strategy to define protein interaction networks for peptide recognition modules. *Science* **295**, 321–324 (2002).
2. T. Clackson, J. A. Wells, A hot spot of binding energy in a hormone-receptor interface. *Science* **267**, 383–386 (1995).
3. A. A. Bogan, K. S. Thorn, Anatomy of hot spots in protein interfaces. *J. Mol. Biol.* **280**, 1–9 (1998).
4. G. A. Weiss, C. K. Watanabe, A. Zhong, A. Goddard, S. S. Sidhu, Rapid mapping of protein functional epitopes by combinatorial alanine scanning. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 8950–8954 (2000).
5. T. Kortemme, D. Baker, A simple physical model for binding energy hot spots in protein-protein complexes. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 14116–14121 (2002).
6. T. Lazaridis, M. Karplus, Effective energy function for proteins in solution. *Proteins* **35**, 133–152 (1999).
7. T. Kortemme, A. V. Morozov, D. Baker, An orientation-dependent hydrogen bonding potential improves prediction of specificity and structure for proteins and protein-protein complexes. *J. Mol. Biol.* **326**, 1239–1259 (2003).
8. B. Kuhlman, D. Baker, Native protein sequences are close to optimal for their structures. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 10383–10388 (2000).
9. B. S. Chevalier , T. Kortemme, M. S. Chadsey, D. Baker, R. J. Monnat, B. L. Stoddard, Design, activity, and structure of a highly specific artificial endonuclease. *Mol. Cell* **10**, 895–905 (2002).
10. B. J. McFarland, T. Kortemme, S. F. Yu, D. Baker, R. K. Strong, Symmetry recognizing asymmetry: Analysis of the interactions between the C-type lectin-like immunoreceptor NKG2D and MHC class I-like ligands. *Structure (Camb)* **11**, 411–422 (2003).
11. M. J. Boulanger, A. J. Bankovich, T. Kortemme, D. Baker, K. C. Garcia, Convergent mechanisms for recognition of divergent cytokines by the shared signaling receptor gp130. *Mol. Cell* **12**, 577–589 (2003).
12. W. L. DeLano, Unraveling hot spots in binding interfaces: Progress and challenges. *Curr. Opin. Struct. Biol.* **12**, 14–20 (2002).
13. K. A. Sharp, Calculation of HyHel10-lysozyme binding free energy changes: Effect of ten point mutations. *Proteins* **33**, 39–48 (1998).
14. I. Massova, P. A. Kollman, Computational alanine scanning to probe protein-protein interactions: A novel approach to evaluate binding free energies. *J. Am. Chem. Soc.* **121**, 8133–8143 (1999).
15. S. Huo, I. Massova, P. A. Kollman, Computational alanine scanning of the 1:1 human growth hormone-receptor complex. *J. Comput. Chem.* **23**, 15–27 (2002).
16. R. Guerois, J. E. Nielsen, L. Serrano, Predicting changes in the stability of proteins and protein complexes: A study of more than 1000 mutations. *J. Mol. Biol.* **320**, 369–387 (2002).
17. G. Schreiber, A. R. Fersht, Energetics of protein-protein interactions: Analysis of the barnase-barstar interface by single mutations and double mutant cycles. *J. Mol. Biol.* **248**, 478–486 (1995).
18. J. D. Forman-Kay, The 'dynamics' in the thermodynamics of binding. *Nat. Struct. Biol.* **6**, 1086–1087 (1999).
19. L. Jiang, T. Kortemme, B. Kuhlman, D. Baker, unpublished data.
20. D. J. Sloan, H. W. Hellinga, Dissection of the protein G B1 domain binding site for human IgG Fc fragment. *Protein Sci.* **8**, 1643–1648 (1999).

Citation: T. Kortemme, D. E. Kim, D. Baker, Computational alanine scanning of protein-protein interfaces. *Sci. STKE* **2004**, pl2 (2004).