Published in final edited form as:
Proteins. 2010 August 1; 78(10): 2329–2337. doi:10.1002/prot.22745.

Computational Alanine Scanning with Linear Scaling Semi-Empirical Quantum Mechanical Methods

David J. Diller[1,*], **Christine Humblet**[1], **Xiaohua Zhang**[2], and **Lance M. Westerhoff**[2]
[1] Pfizer Inc, 865 Ridge Road, Monmouth Junction, New Jersey 08543

[2] QuantumBio Inc. 200 Innovation Boulevard, Suite 261, State College, Pennsylvania 16802

Abstract

Alanine scanning is a powerful experimental tool for understanding the key interactions in protein-protein interfaces. Linear scaling semi-empirical quantum mechanical calculations are now sufficiently fast and robust to allow meaningful calculations on large systems such as proteins, RNA and DNA. In particular, they have proven useful in understanding protein-ligand interactions. Here we ask the question: can these linear scaling quantum mechanical methods developed for protein-ligand scoring be useful for computational alanine scanning? To answer this question, we assembled 15 protein-protein complexes with available crystal structures and sufficient alanine scanning data. In all, the data set contains ΔΔGs for 400 single point alanine mutations of these 15 complexes. We show that with only one adjusted parameter the quantum mechanics based methods out perform both buried accessible surface area and a potential of mean force and compare favorably to a variety of published empirical methods. Finally, we closely examined the outliers in the data set and discuss some of the challenges that arise from this examination.

Keywords

Protein-Protein interactions; linear scaling quantum mechanics; computational alanine scanning

Introduction

Protein-Protein interactions are critical in most biological systems. They comprise an important and exceptionally challenging family of drug targets[1–5]. A first step to targeting a protein-protein interaction is to understand the atomic interactions on an individual residue basis. Alanine scanning has proven to be a powerful experimental tool for dissecting and understanding protein-protein interactions at the atomic level[6–7]. The key consistent finding is that contributions to ΔG are not distributed evenly throughout residues in the interface[8–11]. Rather contributions to ΔG tend to be localized to a small number of key interactions, now referred to as hot spots. Further these hot spots tend to cluster as well. Identifying hot spots is particularly important because tenable small molecule binding sites that directly disrupt protein-protein interactions are typically found around hot spots[12–14].

Because of the importance of alanine scanning, a number of computational alanine scanning methods have been developed to predict the likely hot spots in protein-protein interactions. The majority of these approaches are empirically derived scoring functions or classifiers. The commonality between this set of methods is that they parameterize their scoring

[*] To whom correspondence should be addressed. djrdiller@gmail.com, phone (609)-216-4311.

functions or classifiers on a large data set of ΔΔGs. These methods include pattern recognition methods such as support vector machines[15], decision trees[16–17], and others[18–20] and incorporate empirical descriptors of the interaction such as sequence conservation, relative and absolute buried surface area, various atom or residue contact indices, shape specificity. Also included in the empirical methods are those that are parameterized starting from forcefield based terms[21–22] such as Leonard-Jones potentials, electrostatics, hydrogen bonding potentials, and solvation models.

A second approach to computational alanine scanning is the statistical or database-derived potentials. These approaches differ from the empirical methods in that they are not explicitly fit to binding data of any sort rather they are built on the frequency of occurrence of interactions of different types over a large data set of protein-protein complexes. These are most often potentials of mean force (PMF)[23–24].

A third approach to computational alanine scanning includes methods based on simulations[25–28]. These include a direct estimate of ΔΔG using techniques such as molecular mechanics/Poisson-Boltzmann/surface area or molecular mechanics/generalized Born/surface area methods typically from a trajectory derived from a simulation of the wild type complex[25–27]. In an alternate simulations based approach, Chong and co-workers[28] used high temperature dynamic simulations of the p53 dimer and correlated changes in unfolding kinetics with experimental ΔΔGs.

Linear scaling semi-empirical methods are now fast and robust enough to be used on large systems such as proteins, RNA, and DNA. In particular, they have been used extensively on protein-ligand complexes[29–40]. Merz and colleagues have developed a protein-ligand scoring function based on these quantum mechanical calculations[41]. Our question is simply: to what extent can these methods developed for protein-ligand complexes be extended to protein-protein interactions in the form of predicting ΔΔG from alanine scanning experiments.

To answer this question, we assembled a case of 15 protein-protein interactions from the AESDB[42–43] and the Binding Interface Database Wiki[44–46]. Complexes were chosen if they had sufficient alanine scanning data and a co-crystal structure of sufficient resolution. The full list of protein-protein complexes used is given in Table I. Only mutations to alanine were considered, and those in which the residue mutated was either glycine or proline were not considered as these mutations might lead to significant conformational changes. In all, 400 mutations with experimental ΔΔGs were used in this studied.

Materials and Methods

All proteins were prepared using the protein preparation work flow in Maestro[47] largely following the defaults. In particular, only hydrogens were minimized. All mutants were created using the OEChem SDK[48] by simply deleting the atoms beyond the Cβ and adding the one or two necessary hydrogens to the Cβ with ideal bond length. All calculations were preformed using divcon version 4.6.2[49–50] with defaults: in particular the PM3 basis set was used. All calculations reported here, unless otherwise noted, were performed with no explicit waters.

Divcon uses a number of terms in its protein-ligand scoring function: gas phase heat of formation, desolvation of the QM charges using the Poisson-Boltzmann equation, a vibrational entropy term, and the attractive portion of the Leonard-Jones potential using AMBER. For the protein-ligand calculations, to calculate the vibrational entropy term the ligand is minimized in the absence of the protein and all vibrational motion is assumed lost upon binding. A comparable term for a computational alanine scanning scoring function would be to minimize just the side chain of the residue to be mutated in its monomer and then calculate the vibrational entropy loss of the side chain by assuming again that all vibrational motion is lost. Because, at this point, this calculation is not practical for this many mutations, the vibrational entropy term was omitted. The assumption that all vibration is lost upon binding would, however, clearly not be valid in the case of protein-protein interactions. We would expect that a residue far from the interface would exhibit very little change in vibration upon binding whereas one completely buried in the interface would lose most vibrational motion. Thus without significant additional investigation the vibrational entropy term would likely add little to this study.

The form of the scoring function is:

$$\text{QM} - \Delta\Delta\text{G}=\alpha^{*}(\Delta\text{Gas phase HOF}+\Delta\text{PB Desolvation})+\Delta\text{Attractive LJ} \tag{1}$$

where α is the lone parameter to be determined and the Δs in the right portion of the equation indicate the difference between each term in the mutated and wild type complexes. α was chosen so as to maximize the correlation between QM-ΔΔG and the experimental ΔΔG. We found α=0.26 to be optimal. Here the “Gas Phase HOF” term is the gas phase heat of formation of the complex minus that of the two domains in isolation and thus “ΔGas Phase HOF” is the difference in the “Gas Phase HOF” for the wild type complex and the mutant complex. The “PB desolvation” term is the desolvation of the CM2 atomic charges obtained from the PM3 calculation. Similarly, “ΔPB Desolvation” indicates the difference between the “PB desolvation” terms of the wild type and mutant complexes. Taken together these two terms represent the electrostatic portion of ΔG and are the only parts of the function that rely on quantum mechanics. The “Attractive LJ” term is the attractive component of the Leonard-Jones potential calculated using the AMBER forcefield.

A Potential of Mean Force

Given their success with protein-ligand scoring[51–52] and their simplicity, a potential of mean force was built for comparison purposes. To do so first, atoms were typed according to residue and atom type (i.e. atomic number and remoteness). Similar atoms, for example Oδ of Asp and Oε of Glu, were grouped together into a single atom type. In all, 19 different atom types were used. For two atoms of type i and j a distance r from one another, the energy of interaction is determined by:

$$E_{ij}(r)=-RT\log\left(\frac{C_{ij}(r)}{C_{ij}}\frac{C}{C(r)}\right) \tag{2}$$

where $C_{ij}(r)$ is the number of interactions between atoms of type i and j whose distance is between r and r+dr, C_{ij} is the total number of interactions between atoms of type i and j, C(r) is the total number of interactions between atoms of any type whose distance is between r and r+dr, and C is the total number of interactions between any two atoms. Here we use dr=0.25 Å.

A non-redundant data set was downloaded from the PDB with a resolution cutoff of 2.5 Å. Homologues were removed at 90% sequence identity. This gave a set of 8298 crystal structures. Interactions were counted in two different ways. These counting methods, described below, gave surprisingly different results.

The first PMF, (PMF1), was created by counting interactions between different but interacting monomers. Within a PDB file, two monomers were considered to be interacting if the buried surface area between the two was >600 Å^2, and the pair of monomers had the greatest buried surface area of any two monomers in the PDB file. The second PMF, (PMF2), was created by counting interactions between atoms in the same monomer. Here all interactions between pairs of atoms in the same monomer were counted unless the two atoms were in the same or sequential residues.

When we applied these two PMFs to the data set described in Table I, we found that PMF1 showed significant agreement with the alanine scanning data whereas PMF2 showed no agreement whatsoever. This result is consistent with the observations of Tsai and coworkers[46] regarding differences between packing in monomers and that in protein-protein interfaces. In particular, they find the architecture in monomers to be similar to that at an interface if the two domains fold cooperatively whereas when the domains fold independently prior to binding, they exhibit significant architectural differences.

We expect that the monomers of the complexes listed in Table I largely fold independently prior to binding whereas the set of complexes used to create PMF1 contains complexes that cooperatively fold and complexes with monomers that fold independently. This led to the conclusion that PMF1 is a combination of PMF2 and the ideal PMF for assessing protein-protein interactions (PMF-PPI), i.e.,

$$\mathrm{PMF1}=\alpha\mathrm{PMF}-\mathrm{PPI}+(1-\alpha)\mathrm{PMF2}, \tag{3}$$

or

$$\mathrm{PMF}-\mathrm{PPI}=[\mathrm{PMF1}-(1-\alpha)\mathrm{PMF2}]/\alpha. \tag{4}$$

PMF-PPI was then determined by optimizing its correlation within the data set over values of α between 0 and 1. A value of α=0.655 was found to be optimal. PMF-ΔΔG for a particular mutant is then the PMF-PPI score for the mutant complex minus that for the wild type complex.

Buried Accessible Surface Area

As a second comparison, buried accessible surface area was used. Here each residue was simply ranked by the change in its solvent accessible surface area upon formation of the complex. Surface areas were calculated using only non-hydrogen atoms excluding solvent molecules. Radii were set to Van der Waals radii plus 1.4 Å.

Results and Discussion

The overall results, expressed as correlation coefficients, on a case by case basis are shown in Fig. 1, for QM-ΔΔG, PMF-ΔΔG, and buried surface area. Of the three different methods for ranking the contributions of individual amino acids to binding, QM-ΔΔG is the best in 11 of the 15 cases and is nearly so in the other 4 cases.

Table II contains a variety of additional statistics on the entire data set to compare the three methods as well as to allow comparisons to other methods. For this table the overall correlation coefficients and two correct classification rates with different cutoffs are reported. The cutoff schemes are described below. In addition, these three statistics are reported for all 400 residues or for the 309 contact residues, i.e., those that bury at least 1 Å^2 of surface area in their complex. All the statistics indicate that QM-ΔΔG outperforms PMF-

ΔΔG and buried surface area. In particular, QM-ΔΔG has a correlation coefficient between 0.55 and 0.60 over the entire data set. We believe this to be a good degree of correlation given the fact that the data was gathered under many different conditions, with different assay methodologies and in different laboratories.

The first cutoff scheme was chosen in order to compare with the published computational alanine scanning results of Li and co-workers[19]. In this scheme, residues are classified as experimental hot spots if ΔΔG>1.5, as warm spots if 0.5<ΔΔG≤1.5, and as unimportant if ΔΔG≤0.5 with all units in kcal/mol. Further, only contact residues were included in the study. In our data set this left, 114 hot spots, 92 warm spots, and 103 unimportant residues. Finally, for the purposes of the classification assessment, Li and co-workers considered only the hot spots and unimportant residues. Cutoffs of 5.0, 66.0, and 37.0 were used for QM-ΔΔG, PMF-ΔΔG and buried surface area respectively to computationally classify an interface residue as either a hot spot or unimportant. These cutoffs were determined as those that gave the best correct classification rate. As can be seen in Table II, in the second to last row, the results with QM-ΔΔG were significantly better than with either PMF-ΔΔG or buried surface area. The overall correct classification rate for QM-ΔΔG is 75% compared to 68% for the PMF-ΔΔG and 64% for buried surface area. Further, the 75% success rate compares favorably to the four methods that Li and co-workers[19] tested. These methods include their own[19], PP_SITE[53], the alanine scanning method of Kortemme and Baker[22], and FOLDEF[21]. In this study the success rates for the four methods range from 66–72%. It should be noted that while there is significant overlap between the data set Li and co-workers used and the one used here, they are not identical.

The second cutoff scheme we use for Table II is similar to the first except that we classify a residue as a hot spot if ΔΔG>2.0, as a warm spot if 1.0<ΔΔG≤2.0, and as unimportant ifΔΔG≤1.0 kcal/mol. Again the correct classification rate given uses only the hot spots and the unimportant residues. For the contact residues this left 82 hotspots and 148 unimportant residues. Again, QM-ΔΔG performed significantly better than either PMF-ΔΔG or buried surface area: on the data set of interface residues QM-ΔΔG had a 80% correct classification rate compared to 67% and 64% for the PMF-ΔΔG and buried surface area methods.

Individual cases

1dvf - D1.3 and E5.2

Metals in the interface - This case is noteworthy in that this is the only case where there is a critical metal involved in the interface. QM-ΔΔG is able to perform well (correlation coefficient of 0.78) without additional parameterization. In contrast, both buried surface area and the PMF (correlation coefficient < 0.4) performed significantly worse. In principle, the PMF could be further parameterized given sufficient data with metal containing complexes. It is, however, not clear that this would be straightforward as the presence of the metal greatly perturbs the neighboring residues and may very well impact the subsequent interactions these residues make. This effect may be difficult to describe with simple pairwise potentials without significant additional parameterization.

1gc1 - CD4 and gp120

In this case, none of the methods perform particularly well. QM-ΔΔG works the best with a modest correlation coefficient of 0.30. A visual inspection reveals that many of the residues make no direct contact with gp120. In fact of all the mutated residues, the greatest ΔΔG occurs with Thr81 (1.5 kcal/mol) which is at least 20Å from gp120. As can be seen in Fig. 2, excluding those amino acids that have no buried surface area in the complex leads to a significantly better correlation coefficient of 0.64.

1jtg - TEM-1 β-lactamase (TEM) and β-lactamase inhibitor protein II (BLIP)

Breaking interactions upon complex formation - In this case, all three methods perform seemingly poorly with QM-ΔΔG having the best correlation coefficient of 0.21. Two of these mutations (Tyr105 of TEM and Tyr50 of BLIP) are significant outliers. Without these two mutations the correlation coefficient between QM-ΔΔG and the experimental ΔΔG increases to a more respectable 0.52. The reason for these residues being such significant outliers is quite telling of the limitations of using a static structure for these calculations.

In the complex, TEM Tyr105 makes a number of seemingly good interactions with BLIP including stacking against BLIP Phe142, hydrophobic contact with the aliphatic portion of BLIP Lys74 and a hydrogen bond with BLIP Gly141. Overall, TEM Tyr105 buries over 141 $Å^2$ of surface area in the complex yet its mutation to alanine has essentially no effect on the stability of the complex. A comparison of a TEM apo structure (1zg4[54]) to that in the complex reveals a potential explanation as to why these interactions of Tyr105 observed in the complex do not lead to a significant contribution to the binding affinity. As shown in Fig. 3a, the backbone structure of TEM changes very little between the bound and apo structures. In fact, side chain conformations change very little with the exception of TEM Tyr105. In the apo structure, it stacks against TEM Pro107 and forms a hydrogen bond with the backbone carbonyl of TEM Met127. Thus while TEM Tyr105 makes significant interactions in the BLIP/TEM complex, it must break significant interactions in order to form these interactions. The net effect being that it contributes nothing to the binding affinity.

Mutation of BLIP Tyr50 to alanine increased the stability of the complex by 2.2 kcal/mol. This is by far the largest stability gain with any of the mutations considered in this study. The comparison of the TEM apo structure to the TEM-BLIP complex structure again suggests a potential reason for this. As is apparent from Fig. 3b, the reason TEM Tyr105 cannot adopt the same conformation in the complex as it does in the apo structure is because of a steric clash with BLIP Tyr50. Based on the complex structure it seems likely that not only would TEM Tyr105 be able to maintain its apo conformation in a Y50A BLIP mutant it might also be able to interact with the Cβ of the resulting alanine thereby leading to the gain in stability with the Y50A BLIP mutant.

Other examples of interactions in a complex that simply replace those in one of the monomers leading to significant over estimation of a residue's importance in binding affinity include Im9 Glu41 of 1emv[55]. In this case, Im9 Glu41 makes a salt bridge with E9DNase Arg54. In the apo structure of E9DNase, E9DNase Arg54 is observed to be making a salt bridge with E9DNase Asp51 along with a pair of hydrogen bonds. Thus while the Im9 Glu41-E9DNase Arg54 interaction appears to be energetically significant, it simply replaces energetically comparable interactions in the E9DNase apo structure, and as a result, does not lead to a significant contribution to the binding affinity. It is difficult to assess how frequently this phenomenon occurs as in most cases suitable apo structures are not available.

1brs – Barnase/Barstar

The role of water and stabilization of secondary structure - In this case, the correlation coefficients are all reasonable with the QM-ΔΔG having a correlation coefficient of 0.64. Of the 14 mutations studied in this case, 8 have a measuredΔΔG>3.0 kcal/mol whereas for the full dataset only 53 of 409 mutations have a measured ΔΔG>3.0 kcal/mol. Thus there are an unusual number of significant hotspots in the barnase/barstar interface. The calculated change in binding of most of the mutations is significantly smaller than expected relative to the magnitude of the experimental ΔΔGs. This potentially arises because of the number of water mediated hydrogen bonds in the interface. There are at least 10 observed water molecules that interact with one of these hot spots and also with a residue from the other member of the complex. Fig. 4 shows the interaction of two of these water molecules which help mitigate the interaction between the two domains.

Barnase Asn58 is particularly telling from two aspects. First, its experimental ΔΔG is 3.1 kcal/mol whereas it buries essentially no surface area in the complex and is calculated to have almost no impact on binding. As can be seen in Fig. 4, it forms water mediated interactions with barnase Asp35 (ΔΔG=4.5 kcal/mol) which may lead to a contribution to the binding affinity. Second, barnase Asn58 forms a pair of hydrogen bonds with the backbone of barnase Leu63 across a β-turn. The loss of these hydrogen bonds might destabilize the β-turn thereby affecting the interactions of barnase Arg59 (ΔΔG=5.2 kcal/mol) with barstar Glu76 and Trp78. Thus the large ΔΔG with the mutation of Barnase Asn58 could in part be due to its impact on the stability of the β-turn.

A similar stabilization of a secondary structural element is seen with the case 3d9a[56] in which the HYHEL Asp401Ala mutant leads to a 3.75 kcal/mol loss in affinity despite Asp401 being no closer than 6.5 Å to HEL. Given its distance from the interface, it is calculated to have no effect on the binding. In this case, Asp401 hydrogen bonds via its side chain across a β-turn to HYHEL Asn397. Though its effect on binding has not been studied via mutagenesis, HYHEL Trp398 appears to be a likely candidate for a binding hotspot in this interface. As HYHEL Trp398 is in the turn between HYHEL Asn397 and Asp401, the loss of the hydrogen bonding between the two may lead to a destabilization of this loop and disruption of the interaction between HYHEL Trp398 and HEL.

Many examples of water mediated interactions can be found throughout this dataset. Examples include the 1emv[55] case with Im9 Asp51 and the case 1dfj[57] with RNase Inhibitor Asp431. In both of these cases, there is a water molecule that is entirely bound mediating significant interactions between these residues and their respective partners. The calculated importance of mutation of either of these two residues to alanine is significantly underestimated likely because the water molecules are omitted from the calculation. When we redo the calculations with the assumption that the water molecule is displaced as a result of the mutation of the corresponding residue then the calculations are much more consistent with the experimental ΔΔG. It should be noted, however, that isolated bound waters are less common than networks of bound waters such as that in the barnase-barstar interface. Thus simply removing individual waters when they are involved with the mutated residue will not fully address the role of water in determining ΔΔG.

Interleukin 2 (IL2) and its α receptor (IL2rα) - An example identification of a small molecule binding site

A potential use for alanine-scanning is the identification of a small molecule binding site in a protein-protein interaction. Here we present an example of a prototypical use for computational alanine scanning and ligand binding site identification within a protein-protein interface. In this case, we use IL2 and IL2rα because there is an available co-crystal structure of the two (1z92[58]) and a co-crystal structure of IL2 bound to a small molecule disruptor of the complex(1pw6[59]). Here we calculated the QM-ΔΔG for any residue that has more than 10 Å^2 of buried surface area in the 1z92 complex. This included 15 residues from IL2 and 20 residues from IL2rα. The top ranked residues are shown in Fig. 5. Though these residues represent just over a third of the interface in terms of buried surface area, they localize in one hot spot thereby suggesting this region as a small molecule binding site. Indeed, as is evident in Fig. 5 the small molecule binding site on IL2 is completely engulfed in the calculated hot spots.

Conclusion

The overall results with the QM-ΔΔG computational alanine scanning are encouraging. It is particularly so because with only a single adjustable parameter a scoring function developed for protein-ligand interactions performed for protein-protein interactions as well as empirical methods with tens of adjustable parameters. Furthermore, it is encouraging that electrostatics play a significant role in producing reasonable agreement with experiment. This is likely only possible with the more rigorous approach which explicitly includes effects such as polarization and charge transfer. The importance of electrostatics would likely increase with more accurate basis sets.

The individual cases discussed above highlight several challenges with computational alanine scanning particularly as it relates to applying quantum mechanics. The first difficulty is in considering how side chains reorient upon complex formation. In many cases, it appears to be a reasonable assumption that the conformations adopted by side chains in the complex are not significantly higher in energy than those adopted in the apo structure. There are, however, going to be examples where this assumption is not valid. The only way in which this effect can be accurately captured is through side chain sampling in both the complex and apo structures. This challenge is well demonstrated with the 1jtg example. It is worth noting that identifying those residues that seemingly make strong interactions across the interface but result in no gain in binding affinity due to the loss of similar interactions in one of the apo forms could be very important. These sites are areas where small molecules may be able to improve upon the protein-protein interactions. In this regard, computational alanine scanning is complementary to experimental alanine scanning

A second challenge is to properly handle the rearrangement of discrete water molecules at the interface. This is likely a straightforward exercise when there is a single tightly bound water affected by a mutation. In a situation such as barnase/barstar where the interface is highly solvated and the effects of a single mutation might propagate further along the interface, properly handling the waters may prove to be quite challenging. Indeed, Jiang and co-workers have shown improvement with their computational alanine scanning by explicitly handling discrete waters through a solvated rotamer approach.[60]

The biggest challenge in general with applying quantum mechanics to computational alanine scanning and ultimately protein design is speed. Even with the linear scaling algorithms, quantum mechanical calculations require on the order of hours of CPU time for a single point energy calculation compared to milliseconds for traditional force field based or empirical methods. This is further exasperated when addressing side chain and water sampling issues discussed in the previous paragraph. Given the resources needed, quantum mechanics is best suited as a final filter in the selection of mutants or design of new proteins. The results we present here suggest that current quantum mechanical tools could be a valuable asset as a final filter and offer more evidence that improvements in rigor will lead to improved calculations.

Acknowledgments

We thank Dr. Roger I Martin for source-level improvements to the DivCon package for the treatment of protein/protein interactions. XZ and LW acknowledge the National Institutes of Health Small Business Innovation Research Grant Number R44GM080115 for financial support.

References

1. Fry DC. Protein-protein interactions as targets for small molecule drug discovery. Biopolymers. 2006; 84(6):535–552. [PubMed: 17009316]

2. Fry DC, Vassilev LT. Targeting protein-protein interactions for cancer therapy. J Mol Med. 2005; 83(12):955–963. [PubMed: 16283145]

3. White AW, Westwell AD, Brahemi G. Protein-protein interactions as targets for small-molecule therapeutics in cancer. Expert Rev Mol Med. 2008; 10:e8. [PubMed: 18353193]

4. Arkin MR, Whitty A. The road less traveled: modulating signal transduction enzymes by inhibiting their protein-protein interactions. Curr Opin Chem Biol. 2009; 13(3):284–290. [PubMed: 19553156]

5. Arkin M. Protein-protein interactions and cancer: small molecules going in for the kill. Curr Opin Chem Biol. 2005; 9(3):317–324. [PubMed: 15939335]

6. Clackson T, Wells JA. A hot spot of binding energy in a hormone-receptor interface. Science. 1995; 267(5196):383–386. [PubMed: 7529940]

7. Wells JA. Systematic mutational analyses of protein-protein interfaces. Methods Enzymol. 1991; 202:390–411. [PubMed: 1723781]

8. Bogan AA, Thorn KS. Anatomy of hot spots in protein interfaces. J Mol Biol. 1998; 280(1):1–9. [PubMed: 9653027]

9. DeLano WL. Unraveling hot spots in binding interfaces: progress and challenges. Curr Opin Struct Biol. 2002; 12(1):14–20. [PubMed: 11839484]

10. Keskin O, Ma B, Nussinov R. Hot regions in protein--protein interactions: the organization and contribution of structurally conserved hot spot residues. J Mol Biol. 2005; 345(5):1281–1294. [PubMed: 15644221]

11. Moreira IS, Fernandes PA, Ramos MJ. Hot spots--a review of the protein-protein interface determinant amino-acid residues. Proteins. 2007; 68(4):803–812. [PubMed: 17546660]

12. Wells JA, McClendon CL. Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. Nature. 2007; 450(7172):1001–1009. [PubMed: 18075579]

13. Yin H, Hamilton AD. Strategies for targeting protein-protein interactions with synthetic agents. Angew Chem Int Ed Engl. 2005; 44(27):4130–4163. [PubMed: 15954154]

14. Arkin MR, Wells JA. Small-molecule inhibitors of protein-protein interactions: progressing towards the dream. Nat Rev Drug Discov. 2004; 3(4):301–317. [PubMed: 15060526]

15. Cho KI, Kim D, Lee D. A feature-based approach to modeling protein-protein interaction hot spots. Nucleic Acids Res. 2009; 37(8):2672–2687. [PubMed: 19273533]

16. Darnell SJ, LeGault L, Mitchell JC. KFC Server: interactive forecasting of protein interaction hot spots. Nucleic Acids Res. 2008; 36(Web Server issue):W265–269. [PubMed: 18539611]

17. Darnell SJ, Page D, Mitchell JC. An automated decision-tree approach to predicting protein interaction hot spots. Proteins. 2007; 68(4):813–823. [PubMed: 17554779]

18. del Sol A, O'Meara P. Small-world network approach to identify key residues in protein-protein interaction. Proteins. 2005; 58(3):672–682. [PubMed: 15617065]

19. Li L, Zhao B, Cui Z, Gan J, Sakharkar MK, Kangueane P. Identification of hot spot residues at protein-protein interface. Bioinformation. 2006; 1(4):121–126. [PubMed: 17597870]

20. Tuncbag N, Gursoy A, Keskin O. Identification of computational hot spots in protein interfaces: combining solvent accessibility and inter-residue potentials improves the accuracy. Bioinformatics. 2009; 25(12):1513–1520. [PubMed: 19357097]

21. Guerois R, Nielsen JE, Serrano L. Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. J Mol Biol. 2002; 320(2):369–387. [PubMed: 12079393]

22. Kortemme T, Baker D. A simple physical model for binding energy hot spots in protein-protein complexes. Proc Natl Acad Sci U S A. 2002; 99(22):14116–14121. [PubMed: 12381794]

23. Gilis D, Rooman M. Predicting protein stability changes upon mutation using database-derived potentials: solvent accessibility determines the importance of local versus nonlocal interactions along the sequence. J Mol Biol. 1997; 272(2):276–290. [PubMed: 9299354]

24. Jiang L, Gao Y, Mao F, Liu Z, Lai L. Potential of mean force for protein-protein interaction studies. Proteins. 2002; 46(2):190–196. [PubMed: 11807947]

25. Huo S, Massova I, Kollman PA. Computational alanine scanning of the 1:1 human growth hormone-receptor complex. J Comput Chem. 2002; 23(1):15–27. [PubMed: 11913381]

26. Moreira IS, Fernandes PA, Ramos MJ. Computational alanine scanning mutagenesis--an improved methodological approach. J Comput Chem. 2007; 28(3):644–654. [PubMed: 17195156]
27. Zoete V, Meuwly M. Importance of individual side chains for the stability of a protein fold: computational alanine scanning of the insulin monomer. J Comput Chem. 2006; 27(15):1843–1857. [PubMed: 16981237]
28. Chong LT, Swope WC, Pitera JW, Pande VS. Kinetic computational alanine scanning: application to p53 oligomerization. J Mol Biol. 2006; 357(3):1039–1049. [PubMed: 16457841]
29. Raha K, Peters MB, Wang B, Yu N, Wollacott AM, Westerhoff LM, Merz KM Jr. The role of quantum mechanics in structure-based drug design. Drug Discov Today. 2007; 12(17–18):725–731. [PubMed: 17826685]
30. Peters MB, Raha K, Merz KM Jr. Quantum mechanics in structure-based drug design. Curr Opin Drug Discov Devel. 2006; 9(3):370–379.
31. Raha K, Merz KM Jr. A quantum mechanics-based scoring function: study of zinc ion-mediated ligand binding. J Am Chem Soc. 2004; 126(4):1020–1021. [PubMed: 14746460]
32. Zhou T, Huang D, Caflisch A. Is quantum mechanics necessary for predicting binding free energy? J Med Chem. 2008; 51(14):4280–4288. [PubMed: 18578469]
33. Yu N, Hayik SA, Wang B, Liao N, Reynolds CH, Merz KM. Assigning the protonation states of the key aspartates in beta-Secretase using QM/MM X-ray structure refinement. J Chem Theory Comput. 2006; 2(4):1057–1069. [PubMed: 19079786]
34. Khandogin J, York DM. Quantum descriptors for biological macromolecules from linear-scaling electronic structure methods. Proteins. 2004; 56(4):724–737. [PubMed: 15281126]
35. Xiong Y, Lu HT, Li Y, Yang GF, Zhan CG. Characterization of a catalytic ligand bridging metal ions in phosphodiesterases 4 and 5 by molecular dynamics simulations and hybrid quantum mechanical/molecular mechanical calculations. Biophys J. 2006; 91(5):1858–1867. [PubMed: 16912214]
36. Cross JB, Duca JS, Kaminski JJ, Madison VS. The active site of a zinc-dependent metalloproteinase influences the computed pK(a) of ligands coordinated to the catalytic zinc ion. J Am Chem Soc. 2002; 124(37):11004–11007. [PubMed: 12224947]
37. Westerhoff LM, Merz KM Jr. Quantum mechanical description of the interactions between DNA and water. J Mol Graph Model. 2006; 24(6):440–455. [PubMed: 16199192]
38. He X, Mei Y, Xiang Y, Zhang DW, Zhang JZ. Quantum computational analysis for drug resistance of HIV-1 reverse transcriptase to nevirapine through point mutations. Proteins. 2005; 61(2):423–432. [PubMed: 16114038]
39. Mei Y, He X, Xiang Y, Zhang DW, Zhang JZ. Quantum study of mutational effect in binding of efavirenz to HIV-1 RT. Proteins. 2005; 59(3):489–495. [PubMed: 15789428]
40. Zhang DW, Xiang Y, Gao AM, Zhang JZ. Quantum mechanical map for protein-ligand binding with application to beta-trypsin/benzamidine complex. J Chem Phys. 2004; 120(3):1145–1148. [PubMed: 15268233]
41. Raha K, Merz KM Jr. Large-scale validation of a quantum mechanics based scoring function: predicting the binding affinity and the binding mode of a diverse set of protein-ligand complexes. J Med Chem. 2005; 48(14):4558–4575. [PubMed: 15999994]
42. http://nic.ucsf.edu/asedb/index.php.
43. Thorn KS, Bogan AA. ASEdb: a database of alanine mutations and their effects on the free energy of binding in protein interactions. Bioinformatics. 2001; 17(3):284–285. [PubMed: 11294795]
44. http://tsailab.org/wikiBID/index.php/Main_Page.
45. Tsai CJ, Lin SL, Wolfson HJ, Nussinov R. Protein-protein interfaces: architectures and interactions in protein-protein interfaces and in protein cores. Their similarities and differences. Crit Rev Biochem Mol Biol. 1996; 31(2):127–152. [PubMed: 8740525]
46. Tsai CJ, Xu D, Nussinov R. Structural motifs at protein-protein interfaces: protein cores versus two-state and three-state model complexes. Protein Sci. 1997; 6(9):1793–1805. [PubMed: 9300480]
47. Maestro. 9.0. New York: Schrodinger, LLC; 2009.
48. OEChem Toolkit. 1.61. Santa Fe, NM: OpenEye Scientific Software, Inc; 2009.

49. Dixon SL, Merz KM. Fast, accurate semiempirical molecular orbital calculations for macromolecules. The Journal of Chemical Physics. 1997; 107(3):879–893.

50. divcon. 4.6.2. State College, PA: QuantumBio, Inc; 2009.

51. Gohlke H, Hendlich M, Klebe G. Knowledge-based scoring function to predict protein-ligand interactions. J Mol Biol. 2000; 295(2):337–356. [PubMed: 10623530]

52. Muegge I, Martin YC. A general and fast scoring function for protein-ligand interactions: a simplified potential approach. J Med Chem. 1999; 42(5):791–804. [PubMed: 10072678]

53. Gao Y, Wang R, Lai L. Structure-based method for analyzing protein-protein interfaces. J Mol Model. 2004; 10(1):44–54. [PubMed: 14634848]

54. Stec B, Holtz KM, Wojciechowski CL, Kantrowitz ER. Structure of the wild-type TEM-1 beta-lactamase at 1.55 A and the mutant enzyme Ser70Ala at 2.1 A suggest the mode of noncovalent catalysis for the mutant enzyme. Acta Crystallogr D Biol Crystallogr. 2005; 61(Pt 8):1072–1079. [PubMed: 16041072]

55. Kuhlmann UC, Pommer AJ, Moore GR, James R, Kleanthous C. Specificity in protein-protein interactions: the structural basis for dual recognition in endonuclease colicin-immunity protein complexes. J Mol Biol. 2000; 301(5):1163–1178. [PubMed: 10966813]

56. Acchione M, Lipschultz CA, Desantis ME, Shanmuganathan A, Li M, Wlodawer A, Tarasov S, Smith-Gill SJ. Light chain somatic mutations change thermodynamics of binding and water coordination in the HyHEL-10 family of antibodies. Mol Immunol. 2009

57. Kobe B, Deisenhofer J. A structural basis of the interactions between leucine-rich repeats and protein ligands. Nature. 1995; 374(6518):183–186. [PubMed: 7877692]

58. Rickert M, Wang X, Boulanger MJ, Goriatcheva N, Garcia KC. The structure of interleukin-2 complexed with its alpha receptor. Science. 2005; 308(5727):1477–1480. [PubMed: 15933202]

59. Thanos CD, Randal M, Wells JA. Potent small-molecule binding to a dynamic hot spot on IL-2. J Am Chem Soc. 2003; 125(50):15280–15281. [PubMed: 14664558]

60. Jiang L, Kuhlman B, Kortemme T, Baker D. A "solvated rotamer" approach to modeling water-mediated hydrogen bonds at protein-protein interfaces. Proteins. 2005; 58(4):893–904. [PubMed: 15651050]

61. Papageorgiou AC, Shapiro R, Acharya KR. Molecular recognition of human angiogenin by placental ribonuclease inhibitor--an X-ray crystallographic study at 2.0 A resolution. Embo J. 1997; 16(17):5162–5177. [PubMed: 9311977]

62. Buckle AM, Schreiber G, Fersht AR. Protein-protein recognition: crystal structural analysis of a barnase-barstar complex at 2.0-A resolution. Biochemistry. 1994; 33(30):8878–8889. [PubMed: 8043575]

63. Scheidig AJ, Hynes TR, Pelletier LA, Wells JA, Kossiakoff AA. Crystal structures of bovine chymotrypsin and trypsin complexed to the inhibitor domain of Alzheimer's amyloid beta-protein precursor (APPI) and basic pancreatic trypsin inhibitor (BPTI): engineering of inhibitors with altered specificities. Protein Sci. 1997; 6(9):1806–1824. [PubMed: 9300481]

64. Li Y, Li H, Smith-Gill SJ, Mariuzza RA. Three-dimensional structures of the free and antigen-bound Fab from monoclonal antilysozyme antibody HyHEL-63(,). Biochemistry. 2000; 39(21): 6296–6309. [PubMed: 10828942]

65. Braden BC, Fields BA, Ysern X, Dall'Acqua W, Goldbaum FA, Poljak RJ, Mariuzza RA. Crystal structure of an Fv-Fv idiotope-anti-idiotope complex at 1.9 A resolution. J Mol Biol. 1996; 264(1): 137–151. [PubMed: 8950273]

66. Zhang E, St Charles R, Tulinsky A. Structure of extracellular tissue factor complexed with factor VIIa inhibited with a BPTI mutant. J Mol Biol. 1999; 285(5):2089–2104. [PubMed: 9925787]

67. Kwong PD, Wyatt R, Robinson J, Sweet RW, Sodroski J, Hendrickson WA. Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. Nature. 1998; 393(6686):648–659. [PubMed: 9641677]

68. Sundstrom M, Lundqvist T, Rodin J, Giebel LB, Milligan D, Norstedt G. Crystal structure of an antagonist mutant of human growth hormone, G120R, in complex with its receptor at 2.9 A resolution. J Biol Chem. 1996; 271(50):32197–32203. [PubMed: 8943276]

69. Sogabe S, Stuart F, Henke C, Bridges A, Williams G, Birch A, Winkler FK, Robinson JA. Neutralizing epitopes on the extracellular interferon gamma receptor (IFNgammaR) alpha-chain

characterized by homolog scanning mutagenesis and X-ray crystal structure of the A6 fab-IFNgammaR1-108 complex. J Mol Biol. 1997; 273(4):882–897. [PubMed: 9367779]

70. Lim D, Park HU, De Castro L, Kang SG, Lee HS, Jensen S, Lee KJ, Strynadka NC. Crystal structure and kinetic analysis of beta-lactamase inhibitor protein-II in complex with TEM-1 beta-lactamase. Nat Struct Biol. 2001; 8(10):848–852. [PubMed: 11573088]

71. Hart PJ, Deep S, Taylor AB, Shu Z, Hinck CS, Hinck AP. Crystal structure of the human TbetaR2 ectodomain--TGF-beta3 complex. Nat Struct Biol. 2002; 9(3):203–208. [PubMed: 11850637]

72. Bhat TN, Bentley GA, Boulot G, Greene MI, Tello D, Dall'Acqua W, Souchon H, Schwarz FP, Mariuzza RA, Poljak RJ. Bound water molecules and conformational stabilization help mediate an antigen-antibody association. Proc Natl Acad Sci U S A. 1994; 91(3):1089–1093. [PubMed: 8302837]

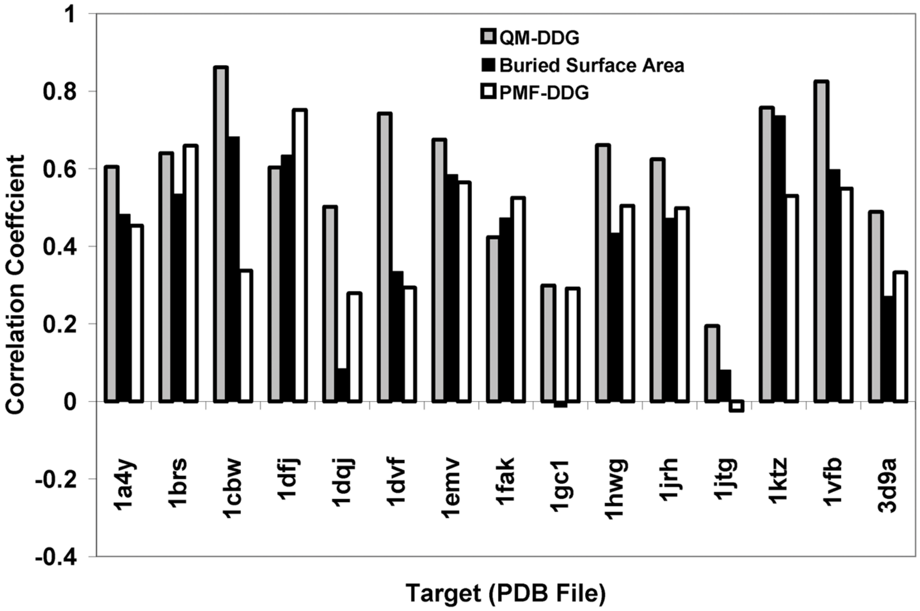

Figure 1.

The correlation coefficients for the three methods: QM-ΔΔG (gray), buried surface area (black) and PMF-ΔΔG (white)

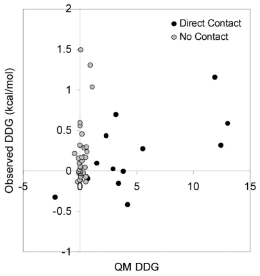

Figure 2.

The results for the 1gc1 case (CD4/gp120) with the QM-ΔΔG scoring function. All mutations are mutations of CD4. The gray are those for which the wild type residue has no contact with gp120 as measured by having no buried surface area upon complex formation. The black are those that make direct contact with gp120.

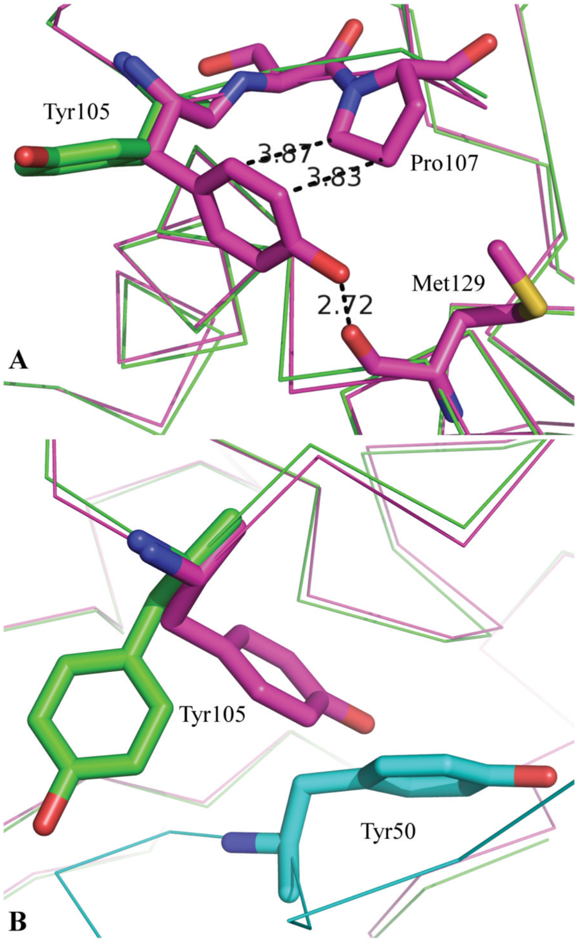

Figure 3.

A comparison of the Apo and bound structures of TEM. **3a:** The green structure is that of TEM as bound to the BLIP from 1jtg[70]. The purple structure is that of apo TEM from the structure 1zg4[54]. **3b:** The green and purple structures are those of 3a. The cyan structure is that of BLIP from the complex with TEM. From 3b it is evident that Tyr105 of TEM-1 cannot adopt the same conformation in the bound (green) as observed in the apo (purple) due largely to the presence of Tyr50 of BLIP.

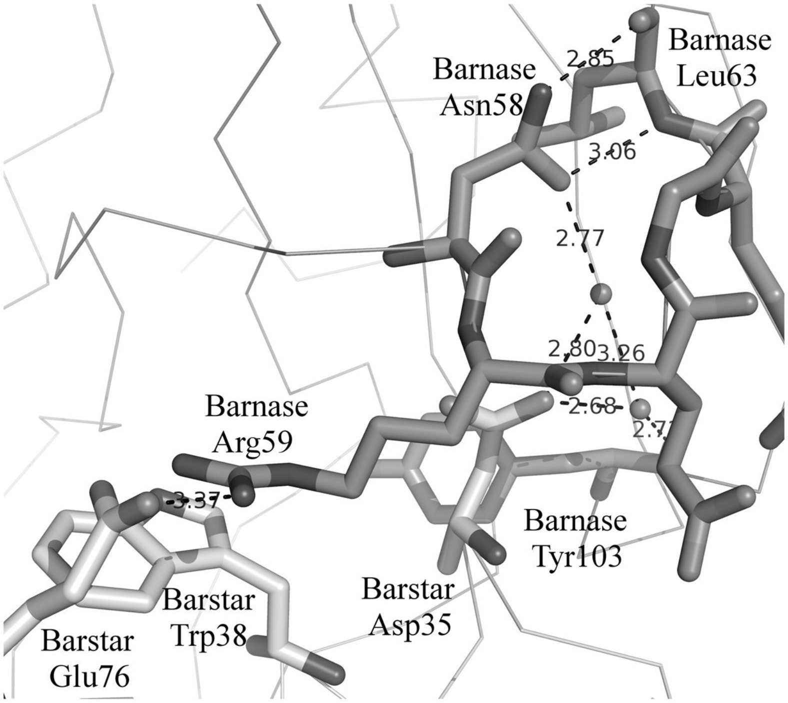

Figure 4.

Barnase/Barstar. Barnase is dark. Barstar is light. Here we highlight two aspects of the role of barnase Asn 58. First, barnase Asn 58 interacts with barstar only through water mediated interactions with barstar Asp35. Second, the hydrogen bonds between barnase Asn 58 and barnase Leu63 may stabilize the β-turn and thereby affect the ability of barnase Arg59 to interact with barstar.

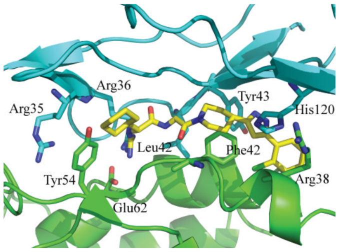

Figure 5.

Computational hot spots for the IL2/IL2rα complex. IL2 is shown in green. IL2rα is shown in cyan. The complex is taken from the structure 1z92[58]. The displayed residues are those that have the greatest QM-ΔΔG. Shown in yellow is a disruptor of this interaction. It is extracted from its co-crystal structure, 1pw6[59], with IL2 and aligned via aligning the IL2 domains from the two structures. The key point is that the calculated hot spots essentially engulf the ligand binding site.

Table I

The Data Set Summary

PDB Code	Number of Mutations	ΔΔG std dev kcal/mo	Resolution Å	Protein1	Protein2
1a4y[61]	28	0.96	2.0	Ribonuclease Inhibitor	Angiognenin
1brs[62]	14	2.46	2.0	Barnase	Barstar
1cbw[63]	8	0.61	2.6	Chymotrypsin	BPTI
1dfj[57]	14	1.40	2.5	Ribonuclease Inhibitor	Ribonuclease A
1dqj[64]	21	1.86	2.0	Lysozyme	HYHEL-63
1dvf[65]	24	1.22	1.9	D1.3	E5.2
1emv[55]	38	1.41	1.7	Im9	Colcin E9
1fak[66]	19	0.85	2.1	Factor VIIa	Tissue Factor
1gc1[67]	49	0.39	2.5	CD4	gp120
1hwg[68]	67	1.10	2.5	Growth Hormone	Growth Hormone Receptor
1jrh[69]	31	1.38	2.8	Antibody A6	Interferon γ-receptor
1jtg[70]	23	1.8	1.7	TEM-1 β-lactamase	β-lactamase inhibitor protein II
1ktz[71]	19	0.70	2.2	TGF-β3	TGF-β3r
1vfb[72]	29	0.98	1.8	Lysozyme	D1.3
3d9a[56]	16	2.13	1.2	Lysozyme	HYHEL-10

Table II

The classification statistics for QM-ΔΔG, PMF-ΔΔG and Buried Surface Area.

Data Set	Statistic	QM-ΔΔG	PMF-ΔΔG	Buried Surface Area
All Residues	Correlation Coefficient	0.60	0.10	0.43
	Correct Classification Rate: 1.5/0.5 Cutoffs	81%	74%	71%
	Correct Classification Rate: 2.0/1.0 Cutoffs	83%	76%	76%
Contact Residues	Correlation Coefficient	0.55	0.01	0.33
	Correct Classification Rate: 1.5/0.5 Cutoffs	75%	68%	64%
	Correct Classification Rate: 2.0/1.0 Cutoffs	80%	67%	64%