



Molecular Dynamics—Solvated Interaction Energy Studies of Protein–Protein Interactions: The MP1–p14 Scaffolding Complex

Qizhi Cui¹, Traian Sulea¹, Joseph D. Schrag¹, Christine Munger², Ming-Ni Hung¹, Marwen Naïm¹, Mirosław Cygler^{1,2} and Enrico O. Purisima^{1*}

¹*Biotechnology Research Institute, National Research Council Canada, 6100 Royalmount Avenue, Montreal, Quebec, Canada H4P 2R2*

²*Department of Biochemistry, McGill University, Montreal, Quebec, Canada H3G 1Y6*

Received 11 January 2008;
received in revised form
2 April 2008;
accepted 11 April 2008
Available online
20 April 2008

Using the MP1–p14 scaffolding complex from the mitogen-activated protein kinase signaling pathway as model system, we explored a structure-based computational protocol to probe and characterize binding affinity hot spots at protein–protein interfaces. Hot spots are located by virtual alanine-scanning consensus predictions over three different energy functions and two different single-structure representations of the complex. Refined binding affinity predictions for select hot-spot mutations are carried out by applying first-principle methods such as the molecular mechanics generalized Born surface area (MM-GBSA) and solvated interaction energy (SIE) to the molecular dynamics (MD) trajectories for mutated and wild-type complexes. Here, predicted hot-spot residues were actually mutated to alanine, and crystal structures of the mutated complexes were determined. Two mutated MP1–p14 complexes were investigated, the p14(Y56A)-mutated complex and the MP1(L63A,L65A)-mutated complex. Alternative ways to generate MD ensembles for mutant complexes, not relying on crystal structures for mutated complexes, were also investigated. The SIE function, fitted on protein–ligand binding affinities, gave absolute binding affinity predictions in excellent agreement with experiment and outperformed standard MM-GBSA predictions when tested on the MD ensembles of Ras–Raf and Ras–RalGDS protein–protein complexes. For wild-type and mutant MP1–p14 complexes, SIE predictions of relative binding affinities were supported by a yeast two-hybrid assay that provided semiquantitative relative interaction strengths. Results on the MP1-mutated complex suggested that SIE predictions deteriorate if mutant MD ensembles are approximated by just mutating the wild-type MD trajectory. The SIE data on the p14-mutated complex indicated feasibility for generating mutant MD ensembles from mutated wild-type crystal structure, despite local structural differences observed upon mutation. For energetic considerations, this would circumvent costly needs to produce and crystallize mutated complexes. The sensitized protein–protein interface afforded by the p14(Y56A) mutation identified here has practical applications in screening-based discovery of first-generation small-molecule hits for further development into specific modulators of the mitogen-activated protein kinase signaling pathway.

Crown Copyright © 2008 Published by Elsevier Ltd. All rights reserved.

Edited by M. Levitt

Keywords: binding affinity calculations; MAP kinase signaling; virtual alanine scanning; yeast two-hybrid assay

*Corresponding author. E-mail address: rico@bri.nrc.ca.

Abbreviations used: MM-GBSA, molecular mechanics generalized Born surface area; SIE, solvated interaction energy; MD, molecular dynamics; MM-PBSA, molecular mechanics Poisson–Boltzmann surface area; MAPK, mitogen-activated protein kinase; DB, DNA binding; AD, activation domain; PDB, Protein Data Bank.

Introduction

Structural and energetic analysis of protein–protein interactions is a challenging endeavor both experimentally and theoretically. Experimental probing of protein–protein binding affinity entails the development of sophisticated yet accurate assays. Discovery of small-molecule modulators of protein–protein interactions adds another requirement, high-throughput portability. Computational alanine scanning provides an alternative to experimental mutagenesis toward identification of binding affinity hot spots at protein–protein interfaces.^{1–4} Time-efficient virtual scanning calculations typically model only the wild-type structure either as a single conformation² or as a conformational ensemble³ and are useful for prioritizing candidate hot spots. More rigorous estimations of the binding free energy change upon mutation are based on conformational ensemble representations of both wild-type and mutated structures.⁴

In recent years, molecular dynamics (MD) simulations have been used increasingly to study protein–protein interactions.^{3–13} MD simulations can provide not only plentiful structural–dynamical information on protein complex structures in solution but also a wealth of energetic information, including the free energy of binding between protein partners. The most rigorous MD-based approaches to estimate binding free energy are the free energy perturbation and thermodynamic integration methods.^{14–16} Because of long convergence time, they are computationally intensive and prohibitive on large systems such as protein–protein complexes. A more commonly used and tractable approach is the molecular mechanics Poisson–Boltzmann surface area (MM-PBSA) method.^{17,18} In this first-principle-based method, the gas-phase energy, calculated using conventional molecular mechanics force fields such as AMBER,^{19–21} is combined with a continuum model of solvation that includes a surface-area-based nonpolar contribution^{22–24} and a polar solvation free energy calculated with the Poisson–Boltzmann (PB) model.^{25,26} Solute entropy can be incorporated from statistical thermodynamics with normal mode analysis.²⁷ Recently, there has been an increased interest in the faster molecular mechanics generalized Born surface area (MM-GBSA) variant of MM-PBSA, which replaces the PB electrostatics with the generalized Born (GB) approximate model of electrostatics in water.^{17,28,29} Another advantage of MM-GBSA is that it utilizes a fully pairwise potential useful for decomposing the total binding free energy into atomic/group contributions, in a structurally nonperturbing formalism.⁷ While the MM-PB(GB)SA method is attractive for estimating binding affinities, the trend emerging from published applications is that it generally overestimates both absolute and relative binding free energies. Also, the time-averaged binding free energy values can be very sensitive to the region of MD trajectory used for calculations.

The solvated interaction energy (SIE) is another first-principle-based approach for estimating bind-

ing free energies in solution.³⁰ SIE is similar in spirit to the MM-PB(GB)SA formalism, in that it includes molecular mechanics force field intermolecular interaction energy terms supplemented by a continuum model of solvation. Typical choices include the AMBER force field^{19–21} and a boundary element method to solve the Poisson equation of electrostatic interactions in water.^{31,32} What distinguishes SIE from MM-PB(GB)SA is that the former has been explicitly calibrated on binding affinities in solution using a diverse dataset of protein–ligand complexes.³⁰ We have previously shown that even parametrizing a solvation model on small-molecule hydration free energy data may not be sufficient for ensuring parameter transferability to binding calculations in solution.³³ We have then proposed parametrization of first-principle-based methods on binding affinity datasets, which has materialized into the SIE function.

Here, we explored the range of applicability of the SIE function, fitted on protein–ligand binding affinities, by testing it on protein–protein complexes. Initial SIE tests were carried out on the Ras–Raf and Ras–RalGDS protein–protein complexes for which both MM-GBSA and experimental binding free energies are reported.^{7,8,13,34} We noted advantages of the SIE function over MM-PB(GB)SA in these systems. We then applied SIE calculations on another protein–protein complex of interest to us, in order to probe and characterize binding affinity hot spots at the protein–protein interface. First, binding affinity hot spots were located by virtual alanine scanning by exploring a consensus approach over various energy functions and different single-structure representations of the molecular system. Next, predictions of binding affinity changes upon hot-spot mutation were carried out with MM-GBSA and SIE using MD trajectories for mutant and wild-type complexes. We explored various protocols to generate the MD trajectories for mutant complexes, as well as their impact on the predicted binding affinities. To enable this analysis, we produced the mutant proteins and then determined the crystal structures of the mutated complexes. Binding experiments were also carried out to measure the validity and robustness of the computational approach.

The protein–protein interaction system featured in this primarily methodological study is represented by the MP1–p14 scaffolding complex that plays key roles in regulating mitogen-activated protein kinase (MAPK) signaling through the Ras, Raf, MEK, and ERK proteins.^{35,36} This pathway is activated by a large number of extracellular stimuli and plays a central role in the induction of processes such as proliferation, differentiation, cell survival, learning, migration, apoptosis, and oncogenic transformation, thus constituting one of the best-characterized signaling pathways involved in the development and propagation of human cancer.³⁷ Many efforts to control pathological consequences of aberrant signaling in the Ras network have targeted the key kinases involved in activation of the pathway (Raf, MEK, and ERK) with inhibitors with varying degrees

of success.^{38–40} An alternative strategy is to interfere only with selected signals by modulating the downstream effects of specific subcellularly localized signaling complexes. This makes the idea of targeting network interactions at the level of adapter or scaffolding components such as MP1–p14 very appealing, since these interactions are what define specificity within the system.

Despite sharing very low sequence identity, MP1 and p14 adopt the same structural fold and associate tightly into a heterodimeric complex. The MP1–p14 complex exhibits a large and shallow interface, characteristic of a *bona fide* protein–protein interaction. Two crystal structures of the wild-type MP1–p14 scaffolding complex have been previously reported,^{41,42} and two additional crystal structures for mutated MP1–p14 complexes have been determined in this study. Although these high-resolution crystal structures provide plentiful atomic-level structural information, our understanding of the dynamics and binding energetics in the MP1–p14 complex is currently very limited. An increasing comprehension of these elements can provide useful tools for modulating the stability of the MP1–p14 complex. This can lead to new opportunities toward the development of a new generation of therapeutic agents exhibiting MAPK pathway specificity, thus addressing a serious drawback of the current kinase inhibition-oriented strategy.

Results

MD of MP1, p14, and the MP1–p14 complex

The previously reported crystal structures of the MP1–p14 complex^{41,42} show that MP1 and p14 adopt the same fold (Fig. 1). Each protein has a five-stranded antiparallel β sheet flanked by three α helices. The interactions between the two proteins are primarily mediated by the edge-to-edge associa-

tion of their $\beta 3$ strands as well as by their $\alpha 2$ helices, effectively forming a 10-stranded β -sheet core in the MP1–p14 heterodimer. The $\alpha 2$ helices from the two proteins antiparallelly extend along the surface of the β -sheet core and are approximately perpendicular to the four $\alpha 1/\alpha 3$ helices on the opposite side of the β -sheet core.

We calculated solution structures for the MP1–p14 complex and the unbound MP1 and p14, from their respective MD simulations over 30-ns production trajectories at 300 K. Starting structures for these MD simulations were based on the crystal structure of the complex. The backbone root-mean-square deviation (RMSD) from initial structures along the 30-ns simulation period is shown in Fig. 2a. It is apparent that unbound p14 has larger RMSD fluctuations and a higher average RMSD value than free MP1. The latter protein has similar average RMSD value but larger RMSD fluctuations than the MP1–p14 complex. Hence, these MD simulations show that the conformations of MP1 and p14 are structurally stabilized upon their heterodimerization into the MP1–p14 complex.

To inspect the local flexibility of each protein, we calculated RMS fluctuations of backbone atoms, averaged for each residue in the unbound proteins and in the complex (Fig. 2b and c). Apart from the very flexible chain termini, the most significant fluctuations correspond to loop regions. It can be seen that the $\alpha 2/\beta 3$ loop is the most flexible region for both unbound proteins, particularly for p14. Upon binding, the conformation of the $\alpha 2/\beta 3$ loop of MP1 is fully stabilized (per-residue RMS fluctuations below 1 Å), whereas the $\alpha 2/\beta 3$ loop of p14 is only partially stabilized (some residues having RMS fluctuations as high as 2.4 Å, about half of the values in the free state).

The MD average structure of the MP1–p14 complex is overall very similar to its crystal structure (Fig. 1). Local conformational differences include slight shifts of the $\alpha 1$ and $\alpha 3$ helices from p14 and $\alpha 2$ and $\alpha 3$ helices from MP1, as well as conformational

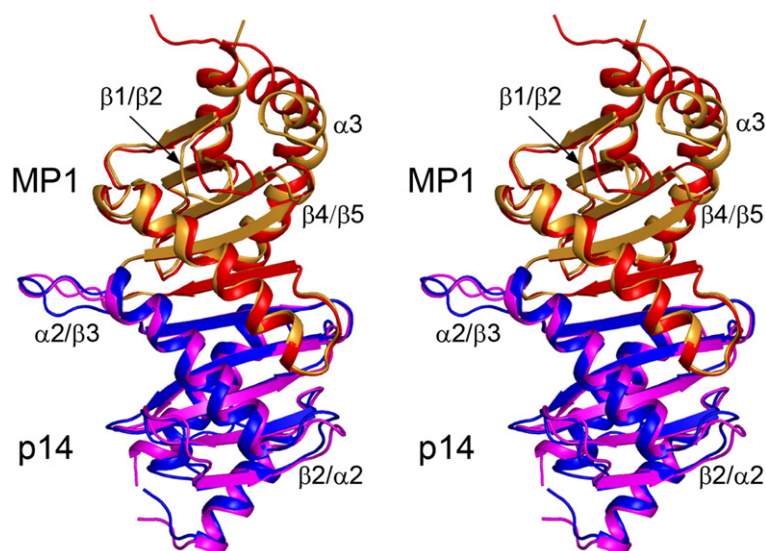


Fig. 1. Crystal and MD structures for the wild-type MP1–p14 complex. The stereoview shows the complex from its $\alpha 2$ helix side. The MD structure is an average over the last 5 ns of a 30-ns simulation. Select regions that show notable conformational differences between the experiment and simulation are labeled. Crystal structure: MP1, orange; p14, magenta. MD structure: MP1, red; p14, blue.

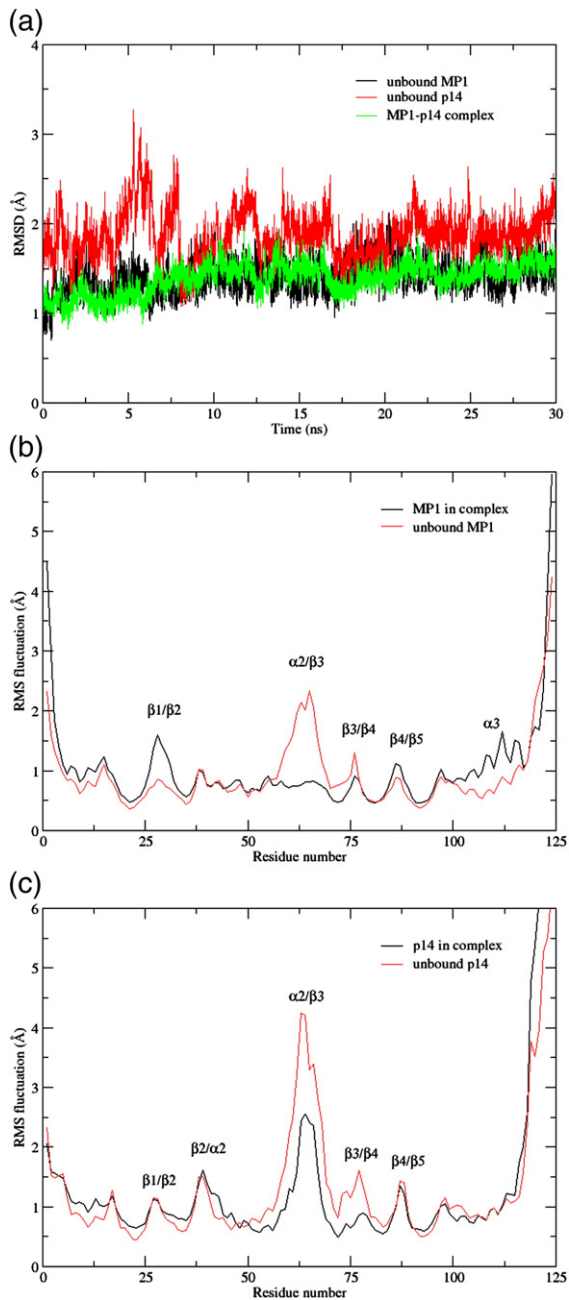


Fig. 2. Structural dynamics of unbound and complexed MP1 and p14. (a) RMSDs of backbone atoms of the unbound MP1, unbound p14, and MP1-p14 complex, with respect to their initial structures during the course of MD simulation. A total of 19 residues corresponding to protein termini (4 from each terminus of MP1, 4 from N-terminus, and 7 from C-terminus of p14) are excluded from the RMS fitting. (b and c) Per-residue RMS fluctuations of backbone atoms of MP1 and p14, respectively, in their unbound and bound states, over the 30-ns MD trajectory.

variations in the $\alpha 1/\beta 1$, $\beta 2/\alpha 2$, $\alpha 2/\beta 3$, $\beta 4/\beta 5$, and $\beta 5/\alpha 3$ loops from p14 and $\beta 1/\beta 2$ and $\beta 4/\beta 5$ loops from MP1. These loops, particularly $\beta 2/\alpha 2$ and $\alpha 2/\beta 3$ from p14, are more flexible than other regions in the complex (Fig. 2b and c) and, thus, are more

difficult to characterize reliably by crystallography. There are a number of conformational differences between the MD and crystal structures for the exposed side chains from these flexible p14 loops, for example, Tyr37 and Asp39 side chains from the $\beta 2/\alpha 2$ loop and Asp61, Glu62, Phe64, Glu66, and Lys70 from the $\alpha 2/\beta 3$ loop. The side chains of Asp61, Glu62, and Phe64 are also missing from the crystal structure of the complex.

Virtual alanine-scanning mutagenesis of the MP1-p14 complex

In order to identify binding affinity hot spots on the MP1-p14 complex, we performed “single-structure” computational alanine scanning of the complex, based on either the prepared crystal structure or the average structure over the last 10 ns of MD simulation, and using Robetta,^{1,2} MM-GBSA,^{17,18} and SIE³⁰ methods to estimate relative binding free energies. As seen in Fig. 3, the magnitudes of predicted $\Delta\Delta G_{\text{bind}}$ values upon mutation vary between methods, with MM-GBSA generally predicting the largest effects and SIE the smallest values. However, there is a good correlation between the relative binding affinity changes upon mutations predicted by the three methods. With respect to the use of the crystal structure or the MD average structure for scanning mutagenesis calculations, we obtained a similar pattern of binding affinity contributions, albeit using the MD average structure seems to lead to slightly larger values.

A consensus over the scanning results obtained on the two structures (crystal and MD average) with the three affinity scoring functions (Robetta, MM-GBSA, and SIE) highlights the side chain of Tyr56 from p14 as having the largest contribution to the binding free energy in the MP1-p14 complex (Fig. 3). Overall, scanning mutagenesis results indicate that from MP1, several side chains including those of Leu63, Leu65, Tyr74, and Tyr75 have moderate contributions to the binding free energy. Based on these scanning data, we set out to investigate further the MP1-p14 interface by characterizing two mutated MP1-p14 complexes: the p14(Y56A) single-mutant complex and the MP1(L63A,L65A) double-mutant complex.

Crystal and MD structures of MP1-p14 mutants

We first produced these mutant MP1-p14 complexes, determined their crystal structures, and compared those with the crystal structure of the wild-type complex (Fig. 4). The $\alpha 2/\beta 3$ loop of p14 undergoes a noteworthy main-chain conformational change between mutated and wild-type complexes. However, we have shown earlier by MD simulations that this loop is highly flexible, and it is located remotely from the mutation sites.

The crystal structure of the MP1(L63A,L65A) double mutant complexed with p14 shows no other significant main-chain conformational changes relative to the wild-type MP1-p14 complex (Fig. 4a).

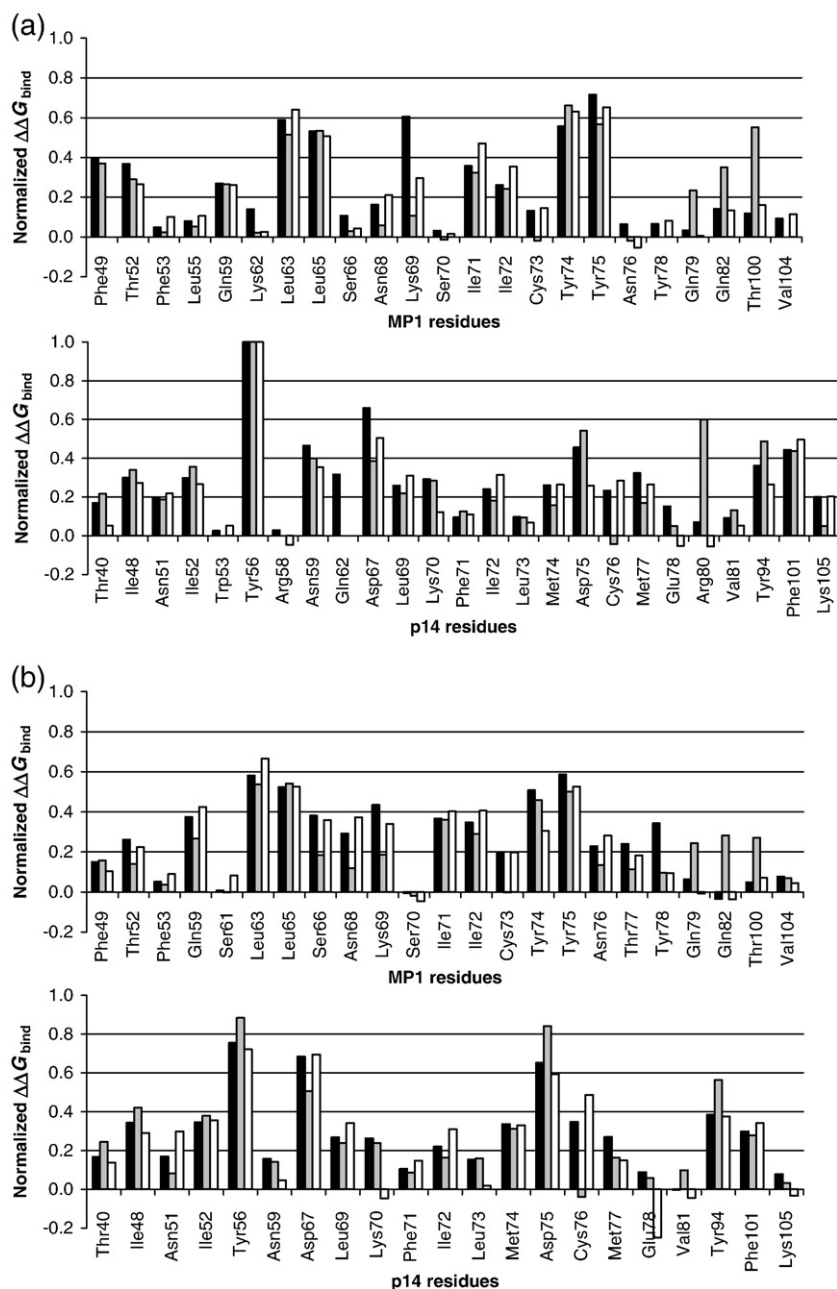


Fig. 3. Single-structure-based computational alanine scanning of the MP1-p14 complex. Calculations are based on (a) the minimized average MD structure over the last 10 ns of the 30-ns MD trajectory and (b) the prepared crystal structure (PDB code: 1VET). Three methods are used to estimate relative binding free energy: SIE³⁰ (black bars), Robetta^{1,2} (gray bars), and MM-GBSA^{17,18} (open bars). Data are normalized to the largest binding free energy change ($\Delta\Delta G_{\text{bind}}$) calculated with a given method over the entire complex. These largest values are (in kilocalories per mole) 1.50 for SIE, 4.39 for Robetta, and 11.4 for MM-GBSA.

Around the mutation sites, the main chain of the $\alpha 2/\beta 3$ loop that carries the mutated residues remains in the wild-type conformation. The Y94 side chain from the interacting $\beta 2/\alpha 2$ loop of p14 changes its conformation from the wild-type complex to fill a cavity created by the two Leu-to-Ala mutations (Fig. 4b). We note, however, that the Y94 side chain of p14 has already been recorded in more than one conformation in previously determined crystal structures of the wild-type MP1-p14 complex.^{41,42}

The crystal structure of the p14(Y56A) single mutant complexed with MP1 exhibits additional conformational changes relative to the wild-type and MP1-mutated complexes. These conformational changes occur mainly in the $\alpha 2$ helix of MP1, whose axis, slightly curved in the wild-type complex, becomes somewhat straighter (Fig. 4c). As a result, the

end of the $\alpha 2$ helix and the following $\alpha 2/\beta 3$ loop of MP1 shift away from the β -sheet core, while the main chain of the $\alpha 2/\beta 3$ loop of MP1 is displaced by as much as 7 Å away from the adjacent $\beta 2/\alpha 2$ loop of p14. The latter exhibits the same Y94 side-chain rotamer as in the MP1-mutated complex. The space created at the protein-protein interface upon the Y56A mutation, located in the $\alpha 2$ helix of p14, is partially filled by the phenyl ring of F53 from the interacting $\alpha 2$ helix of MP1. This compensation occurs *via* a conformational change of the MP1 F53 side chain, pointing toward p14 in the mutant complex, which is almost opposite to its orientation in the wild-type complex (Fig. 4d). This may trigger the slight change in the axis of the $\alpha 2$ helix from MP1.

We carried out 30-ns MD simulations starting from these crystal structures of the MP1-p14 mutant

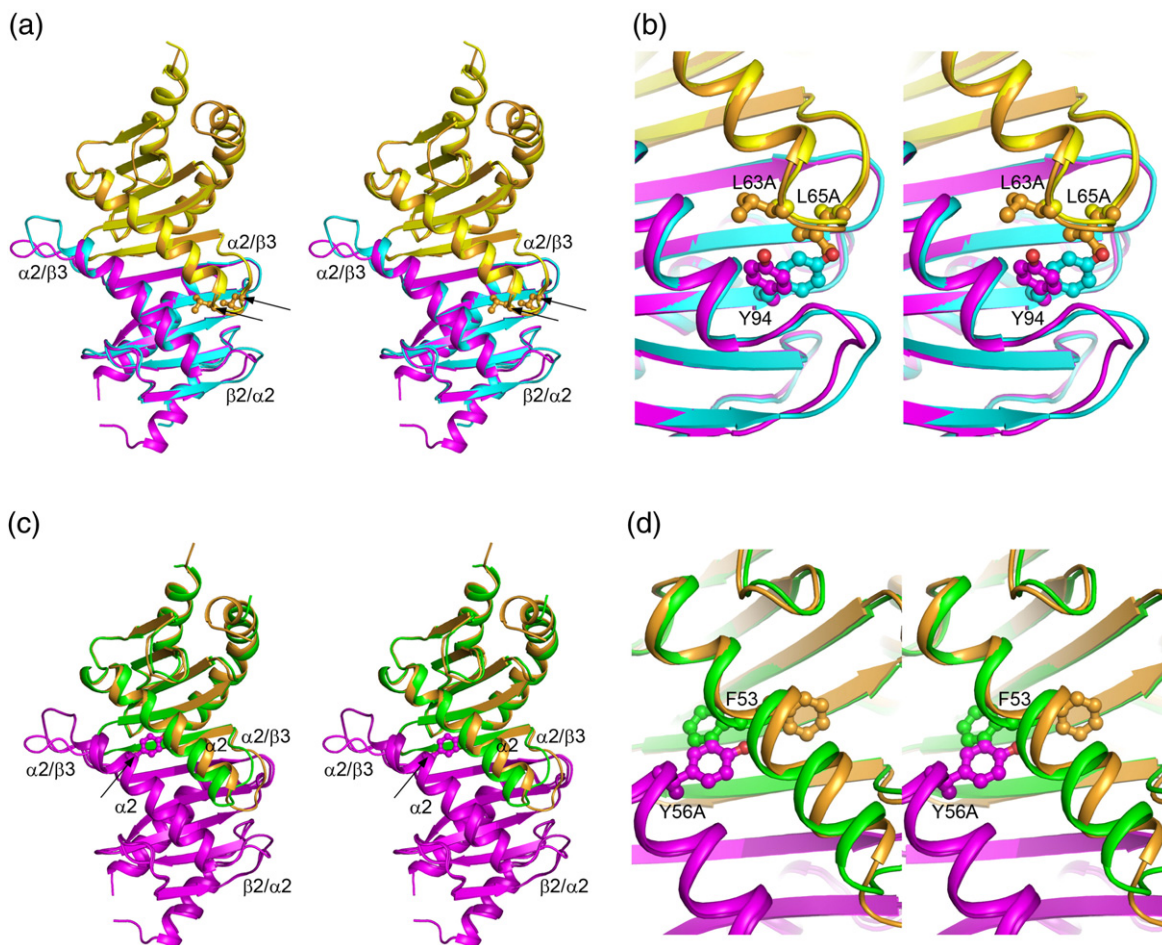


Fig. 4. Crystal structures of mutated MP1-p14 complexes. (a) Global and (b) mutation site focused stereoviews of the MP1(L63A,L65A)-mutated complex superimposed onto the crystal structure of the wild-type complex. Mutant crystal structure: MP1, yellow; p14, cyan. Wild-type crystal structure: MP1, orange; p14, magenta. (c) Global and (d) mutation site focused stereoviews of the p14(Y56A)-mutated complex superimposed onto the crystal structure of the wild-type complex. Mutant crystal structure: MP1, green; p14, purple. Wild-type crystal structure: MP1, orange; p14, magenta. Arrows in the global stereoviews indicate mutation site locations. Side chains at mutation positions and additional side chains that experience notable conformational transitions upon mutation are shown in ball-and-stick representation.

complexes. The calculated MD structures of the mutant complexes remain similar to the crystal structures, with RMSD values within 1–1.5 Å range over the course of the simulation (Fig. 5a). After excluding the flexible protein termini from RMS fitting, one can see that the mutated complexes have overall smaller local RMS fluctuations (per residue) than the wild-type complex, except for slightly increased RMS fluctuations in the $\alpha 2/\beta 3$ region of MP1 and in the $\beta 2/\alpha 2$ loop of p14 (Fig. 5b and c).

For methodological comparison, we also carried out 30-ns MD simulations of the same complexes but starting from *in silico* mutated structures of the wild-type MP1-p14 complex, rather than from the crystal structures of mutant complexes. Average structures over the last 5 ns of MD simulations from different starting structures are compared in Fig. 6 for the MP1- and p14-mutated complexes. For the MP1-mutated complex, the two MD simulations predicted quite similar average structures (except for the termini), including the two most flexible loops, $\beta 2/\alpha 2$ and $\alpha 2/\beta 3$ of p14, predicted to lean

toward MP1 as seen in the crystal structure of this mutated complex (Fig. 4a). Near the mutation sites, the phenol ring of Y94 from p14 orients toward MP1 and is very close to that in the crystal structure of the MP1-mutated complex (Figs. 4b and 6b). However, for the p14-mutated complex, it was rather difficult to reconstitute the experimentally observed structural changes upon mutation by starting the MD simulation from the wild-type conformation (Fig. 4c). In particular, near the p14 mutation site (A56), the F53 side chain of MP1 does not change conformation to fill the void created by the mutation. As a result, the shift in the $\alpha 2$ helix and the displacement of the $\alpha 2/\beta 3$ loop of MP1 are not accomplished, thus rather reflecting the starting structure of the wild-type complex (Figs. 4d and 6d).

Binding free energy calculations

Next, we used the trajectories obtained from the MD simulations started with the crystal structures of wild-type and mutant MP1-p14 complexes to carry

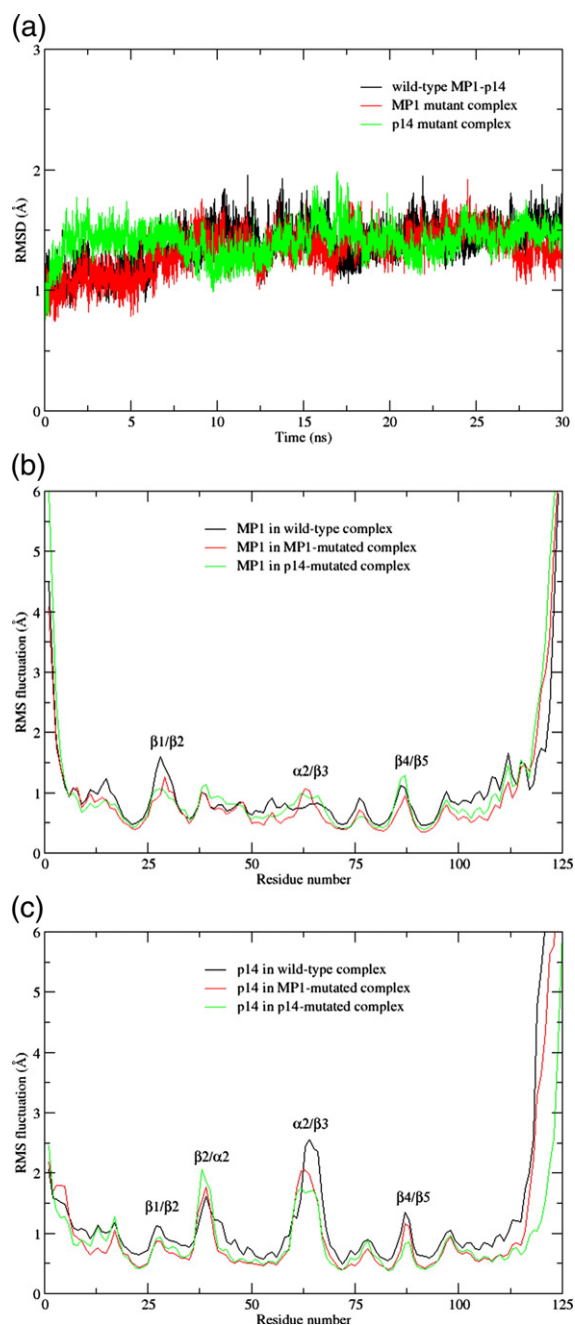


Fig. 5. Structural dynamics of the MP1-p14 mutant complexes. (a) RMSDs of backbone atoms of wild-type and mutated MP1-p14 complexes, with respect to their initial structures during the course of simulation. A total of 19 residues corresponding to protein termini (4 from each terminus of MP1, 4 from N-terminus, and 7 from C-terminus of p14) are excluded from the RMS fitting. (b and c) Per-residue RMS fluctuations of backbone atoms of MP1 and p14, respectively, from wild-type and mutated MP1-p14 complexes over the 30-ns MD trajectory.

out protein-protein binding free energy calculations using the conventional MM-GBSA method.^{17,18} These binding free energy calculations have been shown to be sensitive to the portion of MD trajectory from which the snapshots are selected.¹³ To address the reliability of binding free energy calculations

with respect to this convergence issue in our systems, we calculated total energies as gas-phase energies plus continuum solvation free energies over the entire 30-ns MD trajectories, for each unbound protein and the complex, using the MM-GBSA method. Inspection of these total energy profiles indicated that all systems become energetically stable in the last 10 ns of the 30-ns MD simulation, with energy drift values in the range of 0.1×10^{-3} to 2.1×10^{-3} kcal/mol/ps. Therefore, all binding free energy calculations presented in this study were based on the last 10 ns of MD simulation for the corresponding molecular systems (500 snapshots at 20-ps intervals).

Applying the single (complex) trajectory protocol, the MM-GBSA method predicted an unreasonably strong binding affinity of -69.2 ± 3.8 kcal/mol for the wild-type MP1-p14 complex (Table 1). Not including entropic contributions leaves an even more negative binding free energy estimate of -127.83 kcal/mol. For the two mutated complexes, the predicted binding free energies are even stronger than for the wild type (Table 1). We then used separate trajectories for unbound proteins and complexes to perform the MM-GBSA calculations on these systems. Compared with the single-trajectory results, the separate-trajectory MM-GBSA protocol predicted a more reasonable binding free energy value of -29.0 ± 5.3 kcal/mol for the wild-type MP1-p14 complex. However, the binding free energy differences among the three complexes seem unreasonably large, a 15.7-kcal/mol effect upon MP1 (L63A,L65A) double mutation and a -17.9 -kcal/mol effect upon the p14(Y56A) mutation (Table 2). The calculated stabilizing effect upon p14(Y56A) mutation is particularly puzzling and not consistent with the virtual alanine mutagenesis results showing an expected destabilizing effect (Fig. 3). Replacing the GB continuum solvation model with the more rigorous PB solvent model led to similar results (data not shown). Taken together, these results indicate that although convergence appears to have been reached during MD simulations, the MM-GB(PB)SA method does not predict reliable binding free energies for these protein-protein interaction systems.

The SIE method was calibrated on binding affinities between proteins and small-molecule ligands.³⁰ In order to test its ability to predict binding affinities for protein-protein interaction systems, we applied it on the Ras-Raf and Ras-RalGDS protein-protein complexes for which both experimental and MM-GBSA computational data are available.^{7,13,34} As shown in Table 3, the SIE method predicted protein-protein binding free energies that are closer to experimental values than predictions with the standard MM-GBSA method. In these test cases, SIE slightly overestimates absolute binding free energies, by 21% and 33% for the Ras-RalGDS and Ras-Raf complexes, respectively. In the same examples, standard MM-GBSA overestimated absolute binding free energies by 56% to 132% using the three-trajectory protocol and underestimated the binding affinity by 67% in the single-trajectory mode.^{7,13}

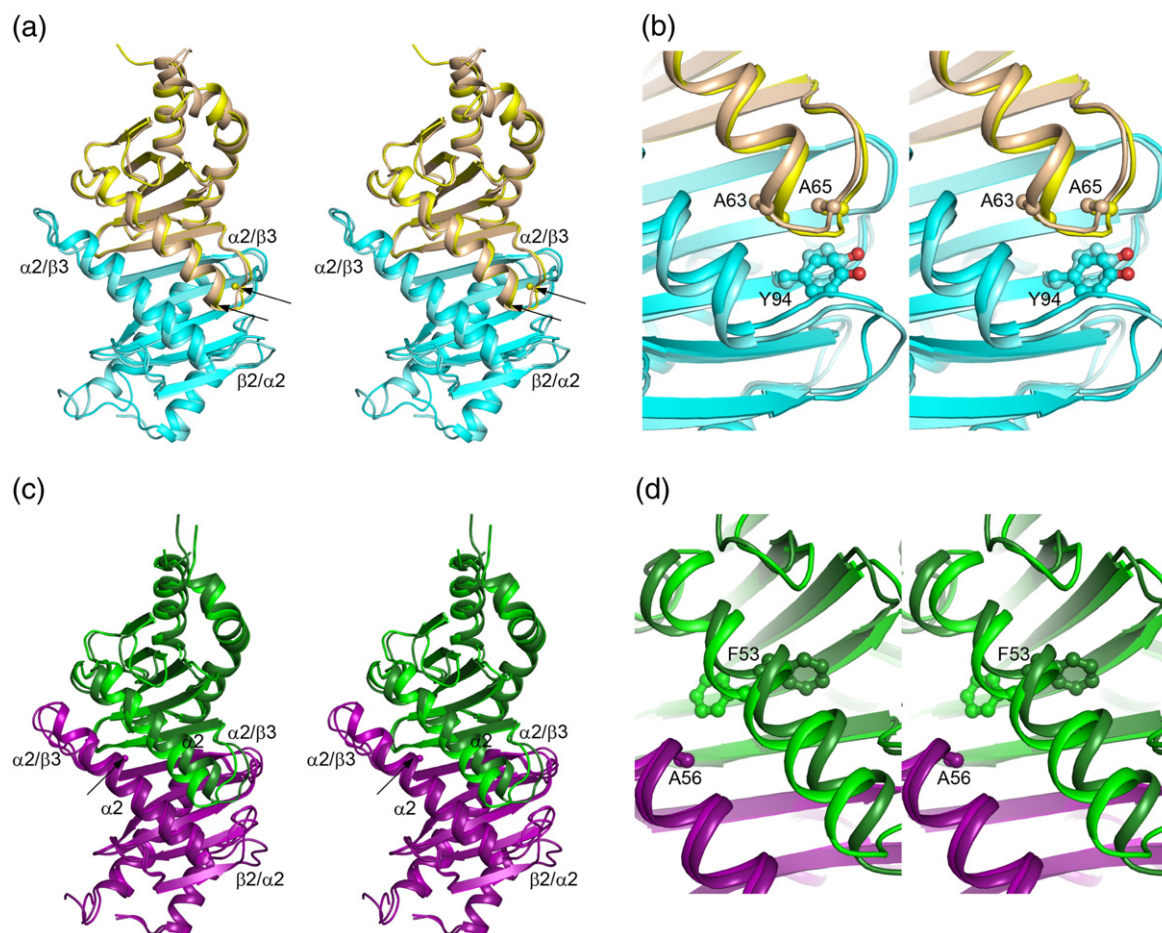


Fig. 6. Comparison of MD structures of mutant MP1-p14 complexes obtained from the crystal structures of the mutant complexes and from *in silico* mutated crystal structure of the wild-type complex. The structures shown are averages over the last 5 ns of the corresponding 30-ns MD simulations. (a) Global and (b) mutation site focused stereoviews of superimposed MP1(L63A,L65A)-mutated complexes. Average MD structure from the mutant crystal structure: MP1, yellow; p14, cyan. Average MD structure from the wild-type crystal structure: MP1, wheat; p14, aquamarine. (c) Global and (d) mutation site focused stereoviews of superimposed p14(Y56A)-mutated complexes. Average MD structure from the mutant crystal structure: MP1, green; p14, purple. Average MD structure from the wild-type crystal structure: MP1, forest green; p14, deep purple. Arrows in the global stereoviews indicate mutation site locations. Side chains at mutation positions and additional side chains shown experimentally to experience notable conformational transitions upon mutation are indicated in ball-and-stick representation.

We then applied SIE-based binding affinity calculations to the wild-type and mutated MP1-p14 complexes from MD simulations. Three different protocols were employed to generate MD conformational ensembles for the mutated complexes. In the order of decreasing computational and experimental demands, these protocols are as follows: (i) actual mutation to alanine, crystal structure determination for the mutant complex, followed by 30-ns MD simulation of the mutant complex; (ii) *in silico* mutation to alanine in the crystal structure of the wild-type complex, followed by 30-ns MD simulation of the mutant complex; and (iii) *in silico* mutation to alanine in snapshots from the MD trajectory of the wild-type complex.

Relative to the MM-GBSA results, the SIE method predicted a more reasonable absolute binding free energy value of -23.08 ± 0.03 kcal/mol for the wild-type MP1-p14 complex (Table 4). For the MP1-p14

(Y56A) singly mutated complex, SIE yielded destabilization effects of 1.30, 1.11, and 1.43 kcal/mol relative to the wild-type complex, by using protocols (i), (ii), and (iii) that differ in the way that MD snapshots are generated for the mutant complex, respectively. These values are also similar with the Y56A side-chain contribution from single-structure virtual alanine scanning (Fig. 3) on the wild-type crystal structure (1.14 kcal/mol) or wild-type MD average structure (1.50 kcal/mol). It can be seen that p14(Y56A) virtual mutation on the single wild-type MD average structure and on 500 snapshots from the wild-type MD trajectory yielded almost the same average relative binding affinity values (1.50 versus 1.43 kcal/mol) by SIE. These results show that SIE can provide protein-protein binding affinity predictions robust to various structural treatment protocols ranging from single structures to MD ensembles, even when local structural changes may

Table 1. Binding free energy for wild-type and mutant MP1-p14 complexes calculated with the MM-GBSA method using MD snapshots extracted from the single complex trajectory

Energy (kcal/mol) ^a	Wild type	MP1(L63A,L65A) mutant	p14(Y56A) mutant
$\langle \Delta E_{\text{ele}} \rangle$	-256.82 ± 1.19	-275.35 ± 1.22	-339.65 ± 1.23
$\langle \Delta E_{\text{vdw}} \rangle$	-162.08 ± 0.28	-165.57 ± 0.29	-142.74 ± 0.27
$\langle \Delta E_{\text{int}} \rangle$	0	0	0
$\langle \Delta E_{\text{gas}} \rangle$	-418.90 ± 1.24	-440.92 ± 1.26	-482.38 ± 1.26
$\langle \Delta G_{\text{nonpolar}} \rangle$	-21.06 ± 0.03	-21.29 ± 0.03	-19.77 ± 0.02
$\langle \Delta G_{\text{GB}} \rangle$	312.14 ± 1.15	326.15 ± 1.17	373.81 ± 1.14
$\langle \Delta G_{\text{sol}} \rangle$	291.07 ± 1.13	304.86 ± 1.16	354.05 ± 1.13
$\langle \Delta(E_{\text{gas}} + G_{\text{sol}}) \rangle$	-127.83 ± 0.28	-136.06 ± 0.28	-128.33 ± 0.30
$\langle \Delta TS \rangle$	-58.64 ± 3.81	-57.00 ± 3.82	-49.85 ± 2.13
$\langle \Delta(E_{\text{gas}} + G_{\text{sol}}) - \Delta TS \rangle$	-69.20 ± 3.82	-79.06 ± 3.83	-78.48 ± 2.15

^a $\langle \Delta \text{Energy} \rangle = \langle \text{Energy}^{\text{MP1-p14}} - \text{Energy}^{\text{MP1}} - \text{Energy}^{\text{p14}} \rangle$, where (...) denotes the average over 500 snapshots at 20-ps intervals from the last 10 ns MD single complex trajectory. The uncertainties shown in Tables 1-4 are mean standard errors. The entropy average was calculated over 20 snapshots at 500-ps intervals from the same trajectory.

exist between the underlying structural models, as described earlier.

For the MP1(L63A,L65A)-p14 double-mutated complex, using protocol (iii) consisting of mutated snapshots from the wild-type MD trajectory, SIE predicted a weakening of binding affinity by 1.51 kcal/mol, similar with the combined contributions of Leu63 and Leu65 side chains based on single-structure virtual alanine scanning (Fig. 3) either from the wild-type crystal structure (1.66 kcal/mol) or from the wild-type MD average structure (1.68 kcal/mol). However, SIE predicted a slightly more favorable binding free energy for the mutated complex relative to the wild-type complex if the MD trajectory of the mutated complex was used, simulated either from *in silico* mutated wild-type complex [protocol (ii)] or from the crystal structure of the mutant [protocol (i)],

Table 2. Binding free energy for wild-type and mutant MP1-p14 complexes calculated with the MM-GBSA method using MD snapshots extracted from separate trajectories of MP1, p14, and MP-p14 complex

Energy (kcal/mol) ^a	Wild type	MP1(L63A,L65A) mutant	p14(Y56A) mutant
$\Delta \langle E_{\text{ele}} \rangle$	-66.24 ± 6.16	-110.03 ± 6.97	-387.26 ± 5.64
$\Delta \langle E_{\text{vdw}} \rangle$	-114.07 ± 1.55	-108.42 ± 1.57	-126.68 ± 1.51
$\Delta \langle E_{\text{int}} \rangle$	-9.32 ± 2.68	10.06 ± 2.69	12.42 ± 2.61
$\Delta \langle E_{\text{gas}} \rangle$	-189.63 ± 6.66	-208.39 ± 7.19	-501.50 ± 6.10
$\Delta \langle G_{\text{nonpolar}} \rangle$	-17.33 ± 0.11	-15.03 ± 0.09	-19.09 ± 0.10
$\Delta \langle G_{\text{GB}} \rangle$	125.18 ± 5.75	162.23 ± 6.39	423.50 ± 5.34
$\Delta \langle G_{\text{sol}} \rangle$	107.83 ± 5.68	147.19 ± 6.37	404.41 ± 5.29
$\Delta \langle (E_{\text{gas}} + G_{\text{sol}}) \rangle$	-81.78 ± 2.75	-61.19 ± 2.49	-97.09 ± 2.63
$\Delta \langle TS \rangle$	-52.74 ± 4.51	-47.83 ± 2.96	-50.20 ± 3.64
$\Delta \langle (E_{\text{gas}} + G_{\text{sol}}) - \Delta TS \rangle$	-29.04 ± 5.28	-13.36 ± 3.87	-46.89 ± 4.49

^a $\Delta \langle \text{Energy} \rangle = \langle \text{Energy}^{\text{MP1-p14}} \rangle - \langle \text{Energy}^{\text{MP1}} \rangle - \langle \text{Energy}^{\text{p14}} \rangle$, where (...) denotes the average over 500 snapshots at 20-ps intervals from the last 10 ns MD trajectory of MP1, p14, or MP-p14 complex. The entropy average was calculated over 20 snapshots at 500-ps intervals from the same trajectory.

Table 3. Binding free energies for Ras-Raf and Ras-RalGDS complexes calculated with the SIE and the MM-GBSA methods

Energy (kcal/mol)	Ras-Raf	Ras-RalGDS
$\langle E_C \rangle$	-554.72 ± 1.09	9.82 ± 2.35
$\langle E_{\text{vdw}} \rangle$	-60.89 ± 0.23	-59.78 ± 0.44
$\langle \gamma \cdot \Delta \text{MSA} \rangle$	-11.05 ± 0.03	-10.21 ± 0.07
$\langle \Delta G_{\text{bind}}^R \rangle$	531.86 ± 1.02	-9.56 ± 1.94
$\langle \Delta G_{\text{bind}} \rangle$ (SIE)	-12.82 ± 0.03 ^a	-10.19 ± 0.06 ^b
$\langle \Delta G_{\text{bind}} \rangle$ (MM-GBSA)	$\begin{cases} -15.0 \pm 6.3^c \\ -19.3 \pm 3.4^e \\ -3.1 \pm 1.6^f \end{cases}$	-19.5 ± 5.9 ^d
ΔG_{bind} (Experiment)	-9.6 ± 0.2 ^g	-8.4 ± 0.2 ^g

^a Average over 500 snapshots at 20-ps intervals from single complex trajectory for Ras-Raf.

^b Average over 150 snapshots at 20-ps intervals from single complex trajectory for Ras-RalGDS.

^c Average over 150 snapshots at 20-ps intervals from separate trajectories for Ras, Raf, and Ras-Raf. Data taken from Ref. 7.

^d Average over 150 snapshots at 20-s intervals from separate trajectories for Ras, RalGDS, and Ras-RalGDS. Data taken from Ref. 7.

^e Average over 500 snapshots at 20-ps intervals from separate trajectories for Ras, Raf, and Ras-Raf. Data taken from Ref. 13.

^f Average over 500 snapshots at 20-ps intervals from single complex trajectory for Ras-Raf. Data taken from Ref. 13.

^g Values determined by isothermal titration calorimetry taken from Ref. 34.

by -0.21 and -1.06 kcal/mol, respectively. These changes in predicted binding affinities could not be associated with any global or significant local conformational changes between the underlying structural models used, as described earlier.

Yeast two-hybrid assay of MP1-p14 interactions

We were not successful in obtaining quantitative measurements of dissociation constants for MP1-p14 by surface plasmon resonance experiments due to poor solubilities of the individual proteins (Jason Baardsnes and Maureen O'Connor-McCourt, personal communication). In contrast, the complexes themselves, which were obtained through coexpression, were better behaved and allowed crystallization for X-ray studies. Alternatively, we employed a qualitative yeast two-hybrid assay in order to experimentally verify the interactions between MP1, p14, and their investigated mutants. We observed growth of diploid yeast cells expressing the GAL4 DNA binding (DB) and activation domain (AD) pair fused to the following protein combinations: MP1 and p14, MP1 and p14(Y56A), MP1(L63A,L65A) and p14, and MP1(L63A,L65A) and p14(Y56A). The growth was insensitive to BD/AD swapping. Qualitatively, this indicates that the mutations of these three residues have marginal effects on the MP1-p14 interaction, not detectable in the growth assay.

To semiquantify the relative strength of these protein-protein interactions, we measured β -galactosidase activities of the yeast two-hybrid clones of MP1 and p14, MP1 and p14(Y56A), MP1(L63A,L65A) and p14, and MP1(L63A,L65A) and p14(Y56A), with DB-fused MP1 variant and AD-fused

Table 4. Binding free energies for wild-type and mutant MP1-p14 complexes calculated with the SIE method using MD snapshots extracted from single complex trajectories generated using different protocols

Energy (kcal/mol) ^a	Wild type	MP1(L63A,L65A) mutant			p14(Y56A) mutant		
		Protocol (i) ^b	Protocol (ii) ^c	Protocol (iii) ^d	Protocol (i) ^b	Protocol (ii) ^c	Protocol (iii) ^d
$\langle E_C \rangle$	-113.97±0.52	-122.64±0.56	-131.66±0.47	-114.14±0.53	-151.65±0.54	-99.55±0.63	-110.35±0.53
$\langle E_{vdw} \rangle$	-162.33±0.28	-166.31±0.29	-160.80±0.26	-149.00±0.27	-142.71±0.26	-157.35±0.27	-151.93±0.27
$\langle \gamma \Delta \text{MSA} \rangle$	-27.78±0.04	-28.11±0.03	-28.26±0.03	-26.23±0.04	-26.06±0.03	-26.73±0.03	-26.23±0.04
$\langle \Delta G_{\text{bind}}^R \rangle$	111.30±0.48	114.14±0.48	125.96±0.43	111.10±0.49	140.10±0.44	101.45±0.55	109.45±0.50
$\langle \Delta G_{\text{bind}} \rangle$	-23.08±0.03	-24.14±0.03	-23.29±0.03	-21.57±0.03	-21.78±0.03	-21.97±0.03	-21.65±0.03
$\Delta \langle \Delta G_{\text{bind}} \rangle$	0	-1.06±0.04	-0.21±0.04	1.51±0.04	1.30±0.04	1.11±0.04	1.43±0.04

^a $\langle \dots \rangle$ denotes the average over 500 snapshots at 20-ps intervals from the last 10 ns MD of the complex single-trajectories.

^b Actual mutation to alanine, crystal structure determination for the mutant MP1-p14 complex, followed by 30-ns MD simulation of the mutant complex.

^c *In silico* mutation to alanine in the crystal structure of the wild-type MP1-p14 complex, followed by 30-ns MD simulation of the mutant complex.

^d *In silico* mutation to alanine in snapshots from the MD trajectory of the wild-type MP1-p14 complex.

p14 variant (Fig. 7). Pairing of MP1 and p14(Y56A) and pairing of MP1(L63A,L65A) and p14(Y56A) lead to lower β -galactosidase activity than pairing of wild-type MP1 and p14. Consistently, pairing of the mutated MP1 with wild-type p14 leads to a similar level of activity as for the wild-type MP1 and p14 pair. Thus, the Y56A mutation of p14 has weakened the interaction between MP1 and p14 to some extent that can be measured by the β -galactosidase assay, whereas the L63A,L65A double mutation of MP1 has no significant impact on the interaction. This agrees well with the trend of relative binding free energies calculated with the SIE method based on MD simulations of wild-type and mutated complexes (Table 4), which predicted similar (slightly stronger) binding in the MP1-mutated complex and weaker binding (by 1 order of magnitude in K_d) in the p14-mutated complex, relative to the wild-type complex.

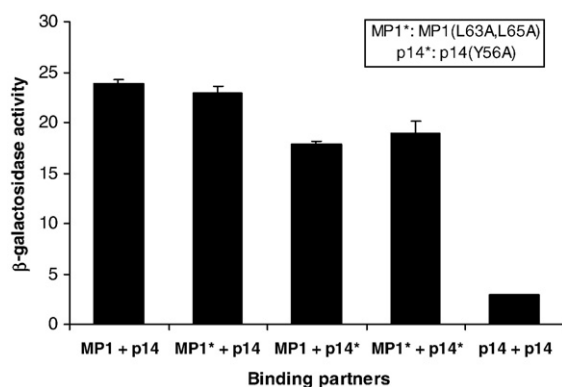


Fig. 7. β -Galactosidase activities of the yeast two-hybrid clones expressing MP1-DB and p14-AD, MP1-DB and p14(Y56A)-AD, MP1(L63A,L65A)-DB and p14-AD, and MP1(L63A,L65A)-DB and p14(Y56A)-AD, where DB and AD are the GAL4 DNA binding and activation domains, respectively. Data represent averages from triplicate experiments. The β -galactosidase activity was expressed as follows: $\text{OD}_{420}/\text{OD}_{600}/\text{volume of the reaction (ml)}/\text{incubation time of the assay (min)} \times 1000$. The activity of the two-hybrid clone expressing DB-p14 and AD-p14 was measured at the same time as a negative control.

Discussion

Using the MP1-p14 scaffolding complex from the MAPK signaling pathway as model system, we explored a structure-based computational protocol to probe and characterize binding affinity hot spots at protein-protein interfaces. In the first step, binding affinity hot spots are located by virtual alanine scanning. For this step, we obtained consensus predictions from three different binding energy functions (Robetta,^{1,2} MM-GBSA,^{17,18} and SIE³⁰) and two single-structure representations of the wild-type complex (the crystal structure and the MD average structure). In the second step, refined predictions of binding affinity changes upon hot-spot mutation are attempted by applying two first-principle methods (MM-GBSA and SIE) to the MD trajectories for mutated and wild-type complexes. For this step, predicted hot-spot residues were actually mutated to alanine, and crystal structures of the mutated complexes were determined. Alternate ways to generate MD trajectories for mutant complexes that do not rely on obtaining the crystal structures for mutated complexes were investigated at this point. Binding experiments were also carried out to measure the validity and robustness of the computational approach.

Single-structure-based virtual scanning

Ranking of side-chain binding affinity contribution by single-structure-based alanine-scanning calculations appears relatively robust with respect to the binding energy function employed (Fig. 3). This is not too surprising, given that Robetta,^{1,2} MM-GBSA,^{17,18} and SIE³⁰ are all first-principle energy functions that employ molecular mechanics force field terms and continuum treatments of solvation. In terms of absolute binding affinity predictions, one can see an advantage of Robetta, and particularly of SIE, over MM-GBSA. The more reasonable magnitudes obtained with the former potentials are probably related to their original calibrations on binding affinity data. The same applies to the more accurate

binding free energy estimates as averages over MD trajectories (Tables 1 and 4).

Computational alanine scanning is also not overly sensitive to using the wild-type crystal structure or average MD structure, albeit differences exist (Fig. 3). Even though the MD simulation of wild-type MP1-p14 complex reached a steady state leading to an average structure structurally similar to the starting crystal structure, some variations of calculated energy terms from slightly different single-conformation structures are to be expected with first-principle methods. The overall robustness to the underlying structure and to the energy function, while not overlooking the noted sensitivities to these parameters, points toward a consensus approach to fast computational hot-spot profiling calculations.

Estimating protein-protein binding affinity with SIE

Although calculating side-chain contributions to protein-protein binding affinity from a single-structure representation of the complex is fast, it is also arguably less accurate than one based on conformational ensemble representations, for example, obtained by MD or Monte Carlo simulations. The SIE function was fitted up to a 1.3-kcal/mol mean unsigned error to experimental binding affinity on a dataset consisting of 99 complexes of 11 proteins with small-molecule ligands with available crystal structures.³⁰ In these protein-ligand calculations, the SIE method was typically applied to a single-structure representation of the complex from which the free state is approximated by a rigid separation of the complex. In terms of MD-ensemble-based binding calculations, this translates to the single-trajectory approach [Eq. (3)]. Applying the SIE function on the MD trajectories of Ras-Raf and Ras-RalGDS protein-protein complexes yielded excellent absolute binding affinity predictions in these systems, -12.8 and -10.2 kcal/mol, respectively, compared to corresponding experimental values of -9.6 and -8.4 kcal/mol. Although these SIE predictions show a slight overestimation of absolute binding affinity, they are superior to the standard MM-GBSA predictions obtained using either single-trajectory or three-trajectory protocols.^{7,13} This mirrors the alanine-scanning results and reflects the applicability and benefits to protein-protein binding calculations of a first-principle energy function fitted on binding affinities in aqueous solution.

The conclusions from the tests on Ras-Raf and Ras-RalGDS complexes are also seen in the results obtained on the MP1-p14 complex. The absolute binding affinity in the wild-type complex is -23.1 kcal/mol, hence stronger than in the other studied complexes. Again, the predicted value appears somewhat overestimated, also considering the reported K_d value of 12.8 nM ($\Delta G_{\text{bind}} = -10.6$ kcal/mol).⁴¹ Yet, a critical examination of the underlying surface plasmon resonance data (Fig. 1b of Ref. 41) indicates a level of uncertainty for this fitted dissociation

constant value, which, in fact, may be smaller. MM-GBSA data for MP1-p14 obtained from a single trajectory were clearly out of range (-69.2 kcal/mol), whereas those from separate trajectories (-29.0 kcal/mol) were closer to the SIE results. However, all MM-GBSA calculations on the MP1-p14 system fail to predict realistic values for the change in binding affinity from wild-type to the two mutant complexes investigated here (Tables 1 and 2). In contrast, SIE not only provided realistic magnitudes for the corresponding relative binding affinities but also was in qualitative agreement with our experimental binding data from a yeast two-hybrid assay (Table 4; Fig. 7).

The underlying assumption of the SIE calculation is the existence of entropy-enthalpy compensation in binding, that is, anticorrelation between entropy and enthalpy that effectively allows the free energy to be expressed as a scaled (solvated) enthalpic term. This is, of course, an approximation since there is no physical requirement for entropy-enthalpy compensation to always occur and examples to the contrary can in fact lead to enhanced binding affinity.⁴³ Hence, the lack of perfect entropy-enthalpy anticorrelation in real systems does contribute to some of the scatter in the predicted absolute binding free energies obtained from the SIE calculations. Notwithstanding this limitation, this treatment of an “averaged out” entropic effect appears to be a useful approximation and preferable to more explicit but inherently inaccurate treatments of entropy that may just introduce noise to the calculation. For example, the use of a normal mode analysis for estimating entropy can be problematic in the presence of multiple energy minima in the system. It is interesting to note that the value of our empirically derived scaling factor, α [Eq. (4)], is similar in magnitude to the degree of entropy-enthalpy compensation found in Monte Carlo calculations of absolute binding free energies of cyclodextrin-ligand complexes.⁴⁴ The cancellation of a large fraction of the enthalpy by the configurational entropic contribution to binding has also been observed in MD simulations of T4 lysozyme L99A complexes carried out in full atomic detail and with explicit water molecules.⁴⁵

A shared limitation of MM-GBSA and the SIE calculations is the use of a term proportional to surface area for the nonelectrostatic solvation contribution. The shortcomings of a surface-area-based functional form have been pointed out in the literature.⁴⁶ The use of a more sophisticated treatment of nonelectrostatic solvation could further improve the agreement of SIE with experiment.

Impact of mutation protocol on binding calculations

The SIE results on the MP1-mutated complex indicate that contributions to protein-protein interactions are overestimated by employing only the MD trajectory of the wild-type complex in the absence of MD ensembles generated for mutated complexes. This behavior is also seen for the p14-

mutated complex, albeit with a reduced effect (Table 4). This simplified treatment, in which mutations are operated in the snapshots from the MD trajectory generated for the wild-type complex, in fact leads to similar results to those from single-structure-based alanine scanning. This overestimation effect is explained by a lack of relaxation after *in silico* mutation.

Simulation of a mutant complex is typically carried out starting from the structure of the wild-type complex or, preferably, if available, from the crystal structure of the mutated complex. Obtaining a faithful energetic and/or structural representation of the mutant by simulations starting from the wild-type crystal structure is not always attainable. For the protein-protein complex investigated here, we determined the crystal structures of two mutant complexes and found that one mutant, p14(Y56A), undergoes a local conformational change upon mutation relative to the wild-type complex. This conformational change could not be reproduced after 30 ns of MD simulation starting from the *in silico* mutated structure of the wild-type complex. Because this mutation site of p14 is located at the center of the MP1-p14 complex interface, the room temperature MD simulation does not provide sufficient sampling of conformational minima separated by large energy barriers, in the vicinity of the mutation site. The MD simulations initiated from the crystal structures of mutated complexes preserved their overall structures.

Although costly production and crystallographic efforts on mutant complexes may thus appear necessary from a structural standpoint, they may not necessarily be so from an energetic standpoint. Despite the noted structural differences in the MD conformational ensembles of the p14-mutated complex generated from the crystal structure of the mutant or from the *in silico* mutated crystal structure of the wild type, these ensembles lead to very similar relative binding affinities calculated by SIE between the mutant and wild-type complexes (Table 4). A plausible explanation is that the MD solution structure of a mutant simulated from the wild-type structure represents a feasible conformational state that is only marginally destabilized relative to the conformational state observed experimentally (e.g., crystallographically), while a large energetic barrier may separate these states.

Sensitization of the MP1-p14 interface

Computational identification of a single-point mutation of p14, Y56A, which only partially weakens the binding affinity to MP1, has the important practical application of providing a sensitized protein-protein interface in the MAPK signaling pathway. This and other medically relevant kinase signaling pathways have traditionally been targeted with small-molecule kinase inhibitors that, unfortunately, suffer from lack of specificity and, thus, have side effects. Chemical interference with the scaffold components rather than kinases is increasingly being perceived as conferring elevated pathway specificity to small-molecule agents. The MP1-p14

scaffolding complex has the characteristics of a *bona fide* protein-protein interface, that is, flat and large interface and high binding affinity, making detection of small-molecule modulators, particularly disruptors, as well as stabilizers, a difficult task by conventional screening assays. Hence, a sensitized MP1-p14 interface has the potential to facilitate the identification of first-generation small-molecule hits, either disruptors or stabilizers, all conceivably weak binders, which will then have to undergo further binding affinity optimization by focused medicinal chemistry efforts.

In addition to the available crystal structures, the MD solution structure of the MP1-p14 complex generated here brings new structural and dynamical elements that will benefit virtual screening calculations, particularly aiming for binding stabilizers. Our MD simulations have revealed that the MP1 in both free and complexed states is more rigid than the p14, with most p14 loops showing increased flexibility relative to the corresponding loops of the MP1 in the complex. Since the flexible $\beta 2/\alpha 2$ and $\alpha 2/\beta 3$ loops on p14 form transient pockets at the surface of the MP1-p14 complex, the MD solution structure has a critical impact on virtual ligand docking and scoring in these regions of the complex. This becomes even more important since the highly flexible $\alpha 2/\beta 3$ loop of p14 may represent a possible interacting site with another functional partner for this scaffolding complex, as already noted.⁴¹

Methods

MD simulations

Starting structures for MD simulations of wild-type and mutant MP1-p14 complexes were based on their crystal structures [Protein Data Bank (PDB) code 1VET, and this study], which were prepared by adding missing side chains, terminal residues, and hydrogen atoms using Sybyl 6.9 (Tripos, Inc., St. Louis, MO) and Xleap (in AMBER)⁴⁷ programs. Wild-type and mutated MP1-p14 complexes prepared for MD simulations had identical sequences, except for the mutated sites. Starting structures for MD simulations of wild-type and mutated free MP1 and p14 were adopted from their complexes. The AMBER 9 suite of programs⁴⁷ together with AMBER FF03 force field^{19,20} were used to perform all MD simulations. Each system was solvated in a truncated octahedron TIP3P water⁴⁸ box. The distance between the wall of the box and the closest atom of the solute was 12.0 Å, and the closest distance between the solute and solvent atoms was 0.8 Å. Counterions (Na^+ , Cl^-) were added to maintain electroneutrality of the system. Each system was minimized first, applying harmonic restraints with force constants of 10 kcal/mol/Å² to all solute atoms, followed by heating from 100 to 300 K over 25 ps in the canonical ensemble (NVT) and by equilibrating to adjust the solvent density under 1 atm pressure over 25 ps in the isothermal-isobaric ensemble (NPT) simulation. The harmonic restraints were then gradually reduced to zero with four rounds of 25-ps NPT simulations. After additional 25-ps simulation, a 30-ns production run was obtained with snapshots collected every 1 ps. For all simulations, 2 fs time step and 9 Å

nonbonded cutoff were used. The particle mesh Ewald method⁴⁹ was used to treat long-range electrostatics, and bond lengths involving bonds to hydrogen atoms were constrained by SHAKE.⁵⁰

MM-GBSA binding calculations

The protein–protein binding free energy can be expressed as:

$$\Delta G_{\text{bind}} = \langle G^{\text{complex}} \rangle - \langle G^{\text{protein 1}} \rangle - \langle G^{\text{protein 2}} \rangle \quad (1)$$

where $\langle \dots \rangle$ denote averages over snapshots from independent MD trajectories. The free energy, G , for each species can be calculated by the following scheme using the MM-GBSA method:^{17,18}

$$\begin{aligned} G &= E_{\text{gas}} + G_{\text{sol}} - TS \\ E_{\text{gas}} &= E_{\text{int}} + E_{\text{ele}} + E_{\text{vdw}} \\ E_{\text{int}} &= E_{\text{bond}} + E_{\text{angle}} + E_{\text{torsion}} \\ G_{\text{sol}} &= G_{\text{GB}} + G_{\text{nonpolar}} \\ G_{\text{nonpolar}} &= \gamma \text{ SAS} \end{aligned} \quad (2)$$

Here, E_{gas} is the gas-phase energy; E_{int} is the internal energy; E_{bond} , E_{angle} , and E_{torsion} are the bond, angle, and torsion energies, respectively; and E_{ele} and E_{vdw} are the Coulomb and van der Waals energies, respectively. E_{gas} was calculated using the AMBER molecular mechanics force field.^{19,20} G_{sol} is the solvation free energy and can be decomposed into polar and nonpolar contributions. G_{GB} is the polar solvation contribution calculated by solving the GB equation.²⁹ Dielectric constants for solute and solvent were set to 1 and 80, respectively. G_{nonpolar} is the nonpolar solvation contribution and was estimated by the solvent accessible surface area, SAS,²³ determined using a water probe radius of 1.4 Å. The surface tension constant γ was set to 0.0072 kcal/mol/Å². T and S are the temperature and the total solute entropy, respectively. Vibrational entropy contributions were estimated by classical statistical thermodynamics,²⁷ using normal mode analysis. To obtain the normal modes, we carried out minimization of each snapshot in gas phase using the conjugate gradient method with a distance-dependent dielectric constant of $4r$ (where r is the interatomic distance), until the RMS of the elements of the gradient vector was less than 10^{−4} kcal/mol/Å.

If snapshots of unbound proteins are taken from the trajectory of the protein complex, the binding free energy can be calculated as:

$$\Delta G_{\text{bind}} = \langle G^{\text{complex}}(i) \rangle - \langle G^{\text{protein 1}}(i) \rangle - \langle G^{\text{protein 2}}(i) \rangle \quad (3)$$

where i corresponds to a snapshot of the protein complex. This procedure effectively cancels out the E_{int} terms in Eq. (2) between the bound and free states.

The mm_pbsal.pl script in AMBER 9 suite⁴⁷ was used for all MM-GBSA binding free energy calculations, at an ionic strength of 150 mM based on the experimental conditions. For comparison, we also performed calculations at ionic strengths of 0 and 50 mM. The differences in free energy or in binding free energy calculated at various ionic strengths were less than 1 kcal/mol.

SIE binding calculations

The SIE function for estimating protein–ligand binding free energy is written as:³⁰

$$\Delta G_{\text{bind}}(\rho, D_{\text{in}}, \alpha, \gamma, C) = \alpha \cdot [E_C(D_{\text{in}}) + \Delta G_{\text{bind}}^R(\rho, D_{\text{in}}) + E_{\text{vdw}} + \gamma \cdot \Delta \text{MSA}(\rho)] + C \quad (4)$$

where E_C and E_{vdw} are the intermolecular Coulomb and van der Waals interaction energies in the bound state, respectively, calculated using AMBER molecular mechanics force field (FF03) with an optimized dielectric constant. The ΔG_{bind}^R is the change in the reaction field energy between the bound and free states, calculated by solving the Poisson equation with the boundary element method program, BRI BEM,^{31,32} and using a molecular surface generated with a variable-radius solvent probe.⁵¹ The ΔMSA term is the change in molecular surface area upon binding. The AMBER van der Waals radii linear scaling coefficient (ρ), the solute interior dielectric constant (D_{in}), the molecular surface area coefficient (γ), the global proportionality coefficient relating to the loss of configurational entropy upon binding (α), and a constant (C) are parameters calibrated by fitting to the absolute binding free energies for a set of 99 protein–ligand complexes.³⁰ The optimized values of these parameters are $\alpha=0.1048$, $D_{\text{in}}=2.25$, $\rho=1.1$, $\gamma=0.0129$ kcal/mol/Å², and $C=-2.89$ kcal/mol SIE calculations were carried out with the program sietraj (<http://www.bri.nrc.ca/ccb/pub/software>).

Virtual alanine-scanning mutagenesis

The Robetta,^{1,2} MM-GBSA,^{17,18} and SIE³⁰ methods were independently used to perform “single-structure” computational alanine-scanning mutagenesis of MP1 and p14, based either on the prepared crystal structure⁴¹ (PDB code: 1VET) or on the last 10-ns average MD structure of the wild-type MP1-p14 complex.

Protocols for binding free energy calculations

Binding affinity hot-spot residues identified from single-structure scanning calculations were subjected to refined calculations of binding affinity changes upon mutations, using MD conformational ensembles. Three different protocols were employed to generate such ensembles for the mutated complexes. In the order of decreasing computational and experimental demands, these protocols are as follows: (i) actual mutation to alanine, crystal structure determination for the mutant MP1-p14 complex, followed by 30-ns MD simulation of the mutant complex; (ii) *in silico* mutation to alanine in the crystal structure of the wild-type MP1-p14 complex, followed by 30-ns MD simulation of the mutant complex; and (iii) *in silico* mutation to alanine in snapshots from the MD trajectory of the wild-type MP1-p14 complex. All three protocols were employed to obtain SIE estimates [Eq. (4)] of relative binding affinities upon mutation. Only protocol (i) using crystallographic starting points for mutant simulations was employed for MM-GBSA-based calculations of relative binding affinities, using either the complex single-trajectory treatment [Eq. (3)] or the three separate trajectories treatment [Eq. (1)]. The latter required MD simulations of unbound wild-type and mutant proteins, carried out as described earlier. For each protocol, average energies were collected from 500 snapshots at 20-ps intervals over the last 10 ns of each respective 30-ns MD trajectories. Due to the high computational demand in the minimization steps required for normal mode analysis, average vibrational entropies were based on 20 snapshots at 500-ps intervals from the last 10 ns of each respective simulation.

Crystallization, data collection, and structure refinement of mutant MP1-p14 complexes

MP1 and p14 were mutated using standard molecular biology approaches. Mutant MP1-p14 complexes were

expressed in *Escherichia coli* and purified as previously described for the wild-type complex.⁴² The purified mutant MP1-p14 complexes were concentrated to ~4.5 mg/ml in 50 mM Tris, pH 8.5, and 50 mM NaCl. The complexes were crystallized at room temperature (~20 °C) by hanging drop vapor diffusion from conditions similar to those used for the wild-type complex.⁴² Reservoir solution typically contained 10–11% PEG (polyethylene glycol) 3350 in 100 mM Hepes, pH 7.5. Each drop contained equal volumes of protein and reservoir solutions. The crystals are orthorhombic and belong to space group $P2_12_12_1$ with unit cell dimensions $a=43.0$ Å, $b=122.5$ Å, $c=105.4$ Å. There are two molecules in the asymmetric unit.

Crystals were transferred to a cryoprotectant solution containing 28% PEG 3350 and 12% glycerol in 100 mM Hepes, pH 7.5, and were flash frozen in a stream of nitrogen gas at –180 °C. Diffraction data were collected on a Q315 detector (ADSC, San Diego, CA) at beamline X29 at National Synchrotron Light Source, Brookhaven National Lab, and data were processed using d*TREK.⁵² Models of the mutant complex were generated from PDB entry 1SKO.^{42,53,54} The models were refined using Refmac5⁵⁵ and were rebuilt using Coot.⁵⁴ The quality of the models was evaluated using PROCHECK.⁵⁶ Data collection and refinement statistics are shown in Table 5.

Table 5. X-ray crystallographic data collection and refinement statistics

	MP1*-p14	MP1-p14*
<i>Data collection</i>		
Space group	$P2_12_12_1$	$P2_12_12_1$
Unit cell		
<i>a</i> (Å)	44.29	44.95
<i>b</i> (Å)	62.44	63.14
<i>c</i> (Å)	74.98	75.78
$\alpha = \beta = \gamma$ (°)	90	90
Resolution range	32.5–2.0 (2.07–2.0)	26.3–1.9 (1.97–1.9)
No. of reflections		
Observed	80,763 (8040)	87,978 (8831)
Unique	14,618 (1421)	17,610 (1725)
Mean redundancy	5.5 (5.6)	5.0 (5.1)
Completeness (%)	100	100
Mean $I/\sigma I$	12.7 (3.9)	13.4 (4.5)
R_{sym}^a (%)	5.6 (34.2)	5.6 (28.8)
<i>Refinement</i>		
R_{work}^b (%)	21.0 (23.7)	20.9 (24.3)
R_{free} (%)	26.4 (28.0)	28.3 (30.3)
RMSDs		
Bonds (Å)	0.01	0.009
Angles (°)	1.301	1.158
Mean <i>B</i> -factors (Å ²)		
MP1 (no. of atoms)	30.2 (917)	26.4 (930)
p14 (no. of atoms)	37.2 (880)	25.1 (878)
Solvent (no. of atoms)	42.6 (116)	37.6 (163)
<i>Model quality</i>		
Residues in model	3–121 (1–116)	4–122 (1–116)
Ramachandran analysis (%) [MP1(p14)]		
Most favored	94.3 (94.2)	92.4 (95.2)
Allowed	5.8 (4.8)	7.6 (4.8)
Disallowed	0 (1.0)	0 (0)

MP1*: MP1(L63A,L65A).

p14*: p14(Y56A).

^a $R_{\text{sym}} = (\sum |I_{\text{obs}} - I_{\text{avg}}|) / \sum I_{\text{obs}}$

^b $R_{\text{work}} = (\sum |F_{\text{obs}} - F_{\text{calc}}|) / \sum F_{\text{obs}}$

Yeast two-hybrid assay

Yeast two-hybrid assays were performed using two expression vectors, pDEST32 and pDEST22, from the ProQuest system (Invitrogen). Site-directed mutagenesis of the codons of Leu63 and Leu65 of MP1, and Tyr56 of p14, was done by using an oligonucleotide-directed mutagenesis kit (Amersham). Plasmid pJW173-46 encoding MP1 and p14 was the source of the DNA template for mutagenesis. The DNA fragments of wild-type MP1 and p14, mutants MP1(L63A,L65A) and p14(Y56A), were excised from pJW173 and the mutated pJW173, respectively, and inserted into the BamHI–EcoRI site of the pENTRY vector (Invitrogen) to generate the entry clones for the Gateway recombination technology (Invitrogen). The DNA fragments of wild-type p14 and mutant p14(Y56A) were amplified by PCR using pJW173 and mutated pJW173, respectively, as templates with a 5' primer containing a BamHI site and a 3' primer containing a NotI site and inserted into the BamHI–NotI site of the pENTRY. The inserts in the entry clones were recombined into pDEST32 and pDEST22 vectors, which encode GAL4 DB and AD, respectively. The resulting yeast two-hybrid expression vectors were transformed into *Saccharomyces cerevisiae* strain PJ69; mating types α and α and the diploid yeasts were obtained by mating. The diploid yeasts expressing MP1 or p14 as GAL4 (DB) fusion, and MP1 or p14 as GAL4(AD) fusion, were assayed for (i) growth, on dropout medium lacking leucine, tryptophan, and histidine, and containing 25 mM 3-amino-1,2,4-triazole, and (ii) β -galactosidase activity, to characterize the strength of interaction.

PDB coordinates

Coordinates and structure factors for the MP1(L63A, L65A)–p14 complex and MP1–p14(Y56A) complex have been deposited in the PDB (<http://www.rcsb.org/>) with accession codes 2ZL1 and 3CPT, respectively.

Acknowledgements

Q.C. gratefully acknowledges the Visiting Fellowship in Canadian Government Laboratory. Some of the MD simulations were run on the Linux cluster servers at The Scripps Research Institute, La Jolla, CA. We thank Prof. David Case (The Scripps Research Institute) and Prof. Holger Gohlke (J.W. Goethe-Universität, Frankfurt) for providing the original MD snapshots of Ras–Raf and Ras–RalGDS protein complexes and for helpful discussions. M.-N.H. would like to acknowledge Dr. Gregor Jansen (McGill University, Montreal) for the gift of pDEST32 and pDEST22. We thank Dr. Christophe Deprez (Biotechnology Research Institute, Montreal) for providing computing support. This work was partially supported by the Genome and Health Initiative Program. This is NRCC publication number 49567.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.jmb.2008.04.035](https://doi.org/10.1016/j.jmb.2008.04.035)

References

- Kortemme, T., Kim, D. E. & Baker, D. (2004). Computational alanine scanning of protein-protein interfaces. *Sci. STKE*, **2004**, 12.
- Kortemme, T. & Baker, D. (2002). A simple physical model for binding energy hot spots in protein-protein complexes. *Proc. Natl Acad. Sci. USA*, **99**, 14116–14121.
- Massova, I. & Kollman, P. A. (1999). Computational alanine scanning to probe protein-protein interactions: a novel approach to evaluate binding free energies. *J. Am. Chem. Soc.* **121**, 8133–8143.
- Huo, S., Massova, I. & Kollman, P. A. (2002). Computational alanine scanning of the 1:1 human growth hormone-receptor complex. *J. Comput. Chem.* **23**, 15–27.
- Zoete, V. & Michielin, O. (2007). Comparison between computational alanine scanning and per-residue binding free energy decomposition for protein-protein association using MM-GBSA: application to the TCR-p-MHC complex. *Proteins: Struct. Funct. Bioinform.* **67**, 1026–1047.
- Zoete, V., Meuwly, M. & Karplus, M. (2005). Study of the insulin dimerization: binding free energy calculations and per-residue free energy decomposition. *Proteins: Struct. Funct. Bioinform.* **61**, 79–93.
- Gohlke, H., Kiel, C. & Case, D. A. (2003). Insights into protein-protein binding by binding free energy calculation and free energy decomposition for the Ras-Raf and Ras-RalGDS complexes. *J. Mol. Biol.* **330**, 891–913.
- Gohlke, H., Kiel, C. & Case, D. A. (2004). Change in protein flexibility upon complex formation: analysis of Ras-Raf using molecular dynamics and a molecular framework approach. *Proteins: Struct. Funct. Bioinform.* **56**, 322–337.
- Brandsdal, B. O., Åqvist, J. & Smalås, A. O. (2001). Computational analysis of binding of P1 variants to trypsin. *Protein Sci.* **10**, 1584–1595.
- Elcock, A. H., Sept, D. & McCammon, J. A. (2001). Computer simulation of protein-protein interactions. *J. Phys. Chem. B*, **105**, 1504–1518.
- Wang, W. & Kollman, P. A. (2000). Free energy calculations on dimer stability of the HIV protease using molecular dynamics and a continuum solvent model. *J. Mol. Biol.* **303**, 567–582.
- Muegge, I., Schweins, T. & Warshel, A. (1998). Electrostatic contributions to protein-protein binding affinities: application to Rap/Raf interaction. *Proteins: Struct. Funct. Genet.* **30**, 407–423.
- Gohlke, H., Kiel, C. & Case, D. A. (2004). Converging free energy estimates: MM-PB(GB)SA studies on the protein-protein complex Ras-Raf. *J. Comput. Chem.* **25**, 238–250.
- Beveridge, D. L. & DiCapua, F. M. (1989). Free energy via molecular simulation: applications to chemical and biomolecular systems. *Annu. Rev. Biophys. Biophys. Chem.* **18**, 431–492.
- Jorgensen, W. L. (1989). Free energy calculations: a breakthrough for modeling organic chemistry in solution. *Acc. Chem. Res.* **22**, 184–189.
- Kollman, P. (1993). Free energy calculations: applications to chemical and biochemical phenomena. *Chem. Rev.* **93**, 2395–2417.
- Srinivasan, J., Cheatham, T. E., Cieplak, P., Kollman, P. A. & Case, D. A. (1998). Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate-DNA helices. *J. Am. Chem. Soc.* **120**, 9401–9409.
- Kollman, P. A., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L. *et al.* (2000). Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.* **33**, 889–897.
- Duan, Y., Wu, C., Chowdhury, S., Lee, M. C., Xiong, G., Zhang, W. *et al.* (2003). A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **24**, 1999–2012.
- Lee, M. C. & Duan, Y. (2004). Distinguish protein decoys by using a scoring function based on a new AMBER force field, short molecular dynamics simulations, and the generalized born solvent model. *Proteins: Struct. Funct. Bioinform.* **55**, 620–634.
- Wang, J., Cieplak, P. & Kollman, P. A. (2000). How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **21**, 1049–1074.
- Sanner, M. F., Olson, A. J. & Spehner, J.-C. (1996). Reduced surface: an efficient way to compute molecular surfaces. *Biopolymers*, **38**, 305–320.
- Sitkoff, D., Sharp, K. A. & Honig, B. (1994). Accurate calculation of hydration free energies using macroscopic solvent models. *J. Phys. Chem.* **98**, 1978–1988.
- Weiser, J., Shenkin, P. S. & Still, W. C. (1999). Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO). *J. Comput. Chem.* **20**, 217–230.
- Gilson, M. K. & Honig, B. (1988). Calculation of the total electrostatic energy of a macromolecular system: solvation energies, binding energies, and conformational analysis. *Proteins: Struct. Funct. Genet.* **4**, 7–18.
- Honig, B. & Nicholls, A. (1995). Classical electrostatics in biology and chemistry. *Science*, **268**, 1144–1149.
- McQuarrie, D. A. (1976). *Statistical Mechanics*. Harper and Row, New York.
- Tsui, V. & Case, D. A. (2000). Theory and applications of the generalized born solvation model in macromolecular simulations. *Biopolymers*, **56**, 275–291.
- Onufriev, A., Bashford, D. & Case, D. A. (2000). Modification of the generalized Born model suitable for macromolecules. *J. Phys. Chem. B*, **104**, 3712–3720.
- Naim, M., Bhat, S., Rankin, K. N., Dennis, S., Chowdhury, S. F., Siddiqi, I. *et al.* (2007). Solvated interaction energy (SIE) for scoring protein-ligand binding affinities. 1. Exploring the parameter space. *J. Chem. Inf. Model.* **47**, 122–133.
- Purisima, E. O. & Nilar, S. H. (1995). A simple yet accurate boundary element method for continuum dielectric calculations. *J. Comput. Chem.* **16**, 681–689.
- Purisima, E. O. (1998). Fast summation boundary element method for calculating solvation free energies of macromolecules. *J. Comput. Chem.* **19**, 1494–1504.
- Rankin, K. N., Sulea, T. & Purisima, E. O. (2003). On the transferability of hydration-parametrized continuum electrostatics models to solvated binding calculations. *J. Comput. Chem.* **24**, 954–962.
- Rudolph, M. G., Linnemann, T., Grunewald, P., Wittinghofer, A., Vetter, I. R. & Herrmann, C. (2001). Thermodynamics of Ras/effector and Cdc42/effector interactions probed by isothermal titration calorimetry. *J. Biol. Chem.* **276**, 23914–23921.
- Kolch, W. (2005). Coordinating ERK/MAPK signaling through scaffolds and inhibitors. *Nat. Rev., Mol. Cell Biol.* **6**, 827–837.
- McKay, M. M. & Morrison, D. K. (2007). Integrating signals from RTKs to ERK/MAPK. *Oncogene*, **26**, 3113–3121.
- Dhillon, A. S., Hagan, S., Rath, O. & Kolch, W. (2007). MAP kinase signalling pathways in cancer. *Oncogene*, **26**, 3279–3290.

38. LoRusso, P. M., Adjei, A. A., Varterasian, M., Gadgeel, S., Reid, J., Mitchell, D. Y. *et al.* (2005). Phase I and pharmacodynamic study of the oral MEK inhibitor CI-1040 in patients with advanced malignancies. *J. Clin. Oncol.* **23**, 5281–5293.
39. Rahmani, M., Davis, E. M., Crabtree, T. R., Habibi, J. R., Nguyen, T. K., Dent, P. & Grant, S. (2007). The kinase inhibitor sorafenib induces cell death through a process involving induction of endoplasmic reticulum stress. *Mol. Cell. Biol.* **27**, 5499–5513.
40. Bukowski, R., Cella, D., Gondek, K. & Escudier, B. (2007). Effects of sorafenib on symptoms and quality of life: results from a large randomized placebo-controlled study in renal cancer. *Am. J. Clin. Oncol.* **30**, 220–227.
41. Kurzbauer, R., Teis, D., de Araujo, M. E. G., Maurer-Stroh, S., Eisenhaber, F., Bourenkov, G. P. *et al.* (2004). Crystal structure of the p14/MP1 scaffolding complex: how a twin couple attaches mitogen-activated protein kinase signaling to late endosomes. *Proc. Natl Acad. Sci. USA*, **101**, 10984–10989.
42. Lunin, V. V., Munger, C., Wagner, J., Ye, Z., Cygler, M. & Sacher, M. (2004). The structure of the MAPK scaffold, MP1, bound to its partner, p14: a complex with a critical role in endosomal MAP kinase signaling. *J. Biol. Chem.* **279**, 23422–23430.
43. Rekharsky, M. V., Mori, T., Yang, C., Ko, Y. H., Selvapalam, N., Kim, H. *et al.* (2007). A synthetic host-guest system achieves avidin-biotin affinity by overcoming enthalpy entropy compensation. *Proc. Natl Acad. Sci. USA*, **104**, 20737–20742.
44. Chen, W., Chang, C. E. & Gilson, M. K. (2004). Calculation of cyclodextrin binding affinities: energy, entropy, and implications for drug design. *Biophys. J.* **87**, 3035–3049.
45. Deng, Y. & Roux, B. (2006). Calculation of standard binding free energies: aromatic molecules in the T4 lysozyme L99A mutant. *J. Chem. Theory Comput.* **2**, 1255–1273.
46. Levy, R. M., Zhang, L. Y., Gallicchio, E. & Felts, A. K. (2003). On the nonpolar hydration free energy of proteins: surface area and continuum solvent models for the solute-solvent interaction energy. *J. Am. Chem. Soc.* **125**, 9523–9530.
47. Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M. *et al.* (2005). The Amber biomolecular simulation programs. *J. Comput. Chem.* **26**, 1668–1688.
48. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935.
49. Darden, T., York, D. & Pedersen, L. (1993). Particle mesh Ewald: an $N \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092.
50. Ryckaert, J. P., Ciccotti, G. & Berendsen, H. J. C. (1977). Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **23**, 327–341.
51. Bhat, S. & Purisima, E. O. (2006). Molecular surface generation using a variable-radius solvent probe. *Proteins: Struct. Funct. Bioinform.* **62**, 244–261.
52. Pflugrath, J. W. (1999). The finer things in X-ray diffraction data collection. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **55**, 1718–1725.
53. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H. *et al.* (2000). The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242.
54. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **60**, 2126–2132.
55. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **53**, 240–255.
56. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291.