



Published in final edited form as:

J Mol Biol. 2010 July 9; 400(2): 257–270. doi:10.1016/j.jmb.2010.05.006.

Computational design of a PAK1 binding protein

Ramesh K. Jha¹, Andrew Leaver-Fay¹, Shuangye Yin¹, Yibing Wu^{2,3,4}, Glenn L. Butterfoss^{1,5}, Thomas Szyperski^{2,3}, Nikolay V. Dokholyan¹, and Brian Kuhlman^{1,*}

¹Department of Biochemistry and Biophysics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-7260

²Department of Chemistry, University at Buffalo, The State University of New York, Buffalo, NY 14260-3000

³Northeast Structural Genomics Consortium

Abstract

We describe a computational protocol, called DDMI, for redesigning scaffold proteins to bind to a specified region on a target protein. The DDMI protocol is implemented within the Rosetta molecular modeling program and uses rigid-body docking, sequence design, and gradient-based minimization of backbone and side chain torsion angles to design low energy interfaces between the scaffold and target protein. Iterative rounds of sequence design and conformational optimization were needed to produce models that have calculated binding energies that are similar to binding energies calculated for native complexes. We also show that additional conformation sampling with molecular dynamics can be iterated with sequence design to further lower the computed energy of the designed complexes. To experimentally test the DDMI protocol we redesigned the human hyperplastic discs protein to bind to the kinase domain of p21-activated kinase 1 (PAK1). Six designs were experimentally characterized. Two of the designs aggregated and were not characterized further. Of the remaining four designs, three bound to the PAK1 with affinities tighter than 350 μM . The tightest binding design, named Spider Roll, bound with an affinity of 100 μM . NMR-based structure prediction of Spider Roll based on backbone and $^{13}\text{C}\beta$ chemical shifts using the program CS-ROSETTA indicated that the architecture of human hyperplastic discs protein is preserved. Mutagenesis studies confirmed that Spider Roll binds the target patch on PAK1. Additionally, Spider Roll binds to full length PAK1 in its activated state, but does not bind PAK1 when it forms an auto-inhibited conformation that blocks the Spider Roll target site. Subsequent NMR characterization of the binding of Spider Roll to PAK1 revealed a comparably small binding 'on-rate' constant ($\ll 10^5 \text{ M}^{-1} \text{ s}^{-1}$). The ability to rationally design the site of novel protein-protein interactions is an important step towards creating new proteins that are useful as therapeutics or molecular probes.

Keywords

Computational protein design; protein-protein interactions; protein docking; Rosetta molecular modeling program; NMR; CS-ROSETTA

© 2010 Elsevier Ltd. All rights reserved

*corresponding author bkuhlman@email.unc.edu.

⁴Present address: Department of Biochemistry and Biophysics, University of Pennsylvania, Philadelphia, Pennsylvania 19104-6059

⁵Present address: Center for Genomics and Systems Biology, New York University, New York, New York 10003

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Introduction

Protein-protein interactions (PPI) are indispensable for life and irregularities in PPI are implicated in many pathological conditions. The rational design of PPIs is a rigorous test of our understanding of molecular recognition and accurate design strategies should allow for the creation of novel protein therapeutics, diagnostics and research tools. Recently there has been considerable success in the computational redesign of protein binding affinities and specificities.^{1; 2; 3} In these studies, rotamer and sequence optimization protocols have been used to identify amino acids that form good packing interactions, electrostatic interactions and hydrogen bonds at target interfaces. In general, these simulations begin with a high-resolution crystal structure of the target interaction. Considerably more difficult is the design of protein interactions for which there is no starting structure. There have been impressive results in the design of new coiled-coils, but these studies rely on known patterns of recognition between coiled-coils.^{4; 5; 6} The rational design of novel interfaces between arbitrarily chosen proteins remains largely an unsolved problem.

Recent successes in directed evolution of PPI indicate that even fairly rigid protein scaffolds can be remodeled to bind new target proteins. Ribosome display has been used to design ankyrin repeat proteins that bind with high affinity to maltose binding protein and aminoglycoside phosphotransferase.^{7; 8} Crystal structures of the complexes show only small changes in the conformations of each protein when they dock together. These results suggest a minimal protocol for computer-based interface design: dock the scaffold on to the target protein and then redesign the amino acids on the surface of the scaffold to form favorable interactions with the target. There are many ways that two proteins can be brought together and some orientations are likely to be more designable than others. The challenge is that before redesigning the surface of the scaffold in the presence of the target, it is difficult to determine which docked orientation will provide the lowest energy interactions. Huang and Mayo used a reduced representation of amino acid side chains and a fast Fourier-transform based docking algorithm to find orientations and positions that maximize potential interactions with the target without bringing the proteins too close together.⁹ They used this strategy to redesign the $\beta 1$ domain of streptococcal protein G to form a novel heterodimer with a binding affinity of ~ 300 μM .¹⁰ In their study, the sequences of both sides of the protein interface were optimized and only one docked conformation was explicitly evaluated with protein design simulations.

We have developed a strategy for interface design, called DDMI for dock, design and minimize interface, which is based on the hypothesis that protein interfaces can be designed with minimal changes to their backbone conformations and therefore focuses its sampling on the myriad of possible docked conformations. This protocol builds on Rosetta's existing fixed-backbone design subroutine¹¹, that, when given a docked conformation for the scaffold and target backbones, searches through side chain sequence- and conformation-spaces to produce a low-energy sequence for the docked conformation. Since we cannot know before design begins which docked conformations will lead to good interfaces, we must sample many docked conformations. We therefore begin DDMI with a randomization of the scaffold protein's position and orientation and a stochastic, low resolution ("centroid mode"), rigid-body docking of the two proteins.¹² Because the low-resolution score function is so coarse, independent randomization and docking trajectories are very unlikely to yield the same docked conformation. The initial phase is merely producing a rough guess at how the backbone of the two proteins might orient with respect to each other. After the docking phase completes, DDMI iterates between rounds of design and gradient-based minimization to settle into a low-energy sequence for the scaffold protein. Independent DDMI trajectories settle into dissimilar regions of conformation space, so we typically simulate tens- to hundreds-of-thousands design trajectories. Here we use the DDMI protocol to design an interaction between the hyperplastic discs protein (HYP) and the kinase domain of p21-activated kinase 1 (PAK1).

Results

The model system

We chose kinase domain of p-21 activated kinase 1 (PAK1) as our 'target' protein. The full length PAK1 (PAK1-fl) is a multi-domain protein that can switch between an inactive and active conformation. In the inactivated state, the auto-inhibitory domain of PAK1-fl binds with the kinase domain of PAK1. In the activated state, the auto-inhibitory domain is unfolded by accessory factors and no longer interacts with the kinase domain.¹³ For *de novo* interface design we targeted a region of the PAK1 that is exposed when the auto-inhibitory domain releases. This is an attractive binding site because it is a known region of protein-protein interaction and binders that target this region will be sensitive to the activation state of PAK1, potentially providing a tool for sensing or controlling PAK1 activity. The 'target' patch is a hydrophobic cleft in the C-terminal domain of the kinase domain between two α -helices (α EF and α G). The auto-inhibitory domain of PAK1-fl forms a small helical bundle that inserts in the cleft. As a design scaffold for targeting the PAK1 kinase domain, we used a small helical bundle protein, the Hyperplastic discs protein (HYP, PDB ID: 1I2T),¹⁴ that is similar in size to the auto-inhibitory domain and can fit in the target cleft.

In a preliminary set of HYP designs we found that it was prone to aggregation when redesigned to bind PAK1. Because we are targeting a hydrophobic cleft on PAK1, the redesigned scaffolds typically have additional hydrophobic groups on their surface. To increase the baseline solubility of HYP we selected positions away from the target interface for mutagenesis to polar residues. The mutations, G26E, L37E and L38N are presented in some of the designs discussed here. Additionally, we introduced the mutation A15C to HYP to allow for conjugation of the fluorophore Bodipy for measuring changes in fluorescence polarization upon binding to PAK1. Circular dichroism spectra indicate that these mutations do not perturb the helical structure of HYP (Supplementary figure S1).

Interface design protocol, DDMI

To redesign HYP to bind the PAK1 we developed an interface design protocol, called DDMI, within the Rosetta molecular modeling suite^{15; 16} (Figure 1). The DDMI protocol took as inputs 1) a pdb file containing both the scaffold and target proteins, in an undocked conformation, and 2) a set of constraints, described below. As output, a single DDMI trajectory would produce a single docked conformation of the target and scaffold where the scaffold's residues at the interface had been redesigned to favorably interact with the target's interface residues. The first stage of the protocol was a randomization of the scaffold protein followed by a low resolution rigid-body docking of the two proteins.¹² The goal of this stage was to find a plausible docked-conformation for the two proteins so that they can be designed to bind in this conformation. We filter at this stage any docked conformations that show backbone collision which the design phase would be unable to remove. To bias sampling of the scaffold's conformation to bind against the target cleft, square-well constraints were added to the energy function to reward the burial of specified residues at the interface: residues were rewarded if they were within 7 Å of any residue on the opposite chain. These constraints were removed after the initial stage of docking. Figure 2 (a) shows the representative sampling of structures created during 'dock' stage of the protocol. DDMI preserved the scaffold's original sequence during this phase. Though this sequence would have been unlikely to form favorable interactions (even at low resolution) with residues on the target protein, the constraints ensured that the desired regions of the scaffold's surface were positioned near the target.

After docking, iterative rounds of sequence design and structure optimization with Rosetta's all-atom energy function¹⁶ were used to find low energy sequence structure pairs. Sequence design was performed with simulated annealing and a rotamer-based representation of the

amino acid side chains. Structure optimization was performed using gradient-based minimization of the rigid-body degrees of freedoms orienting the two chains¹⁷ as well as the backbone and side chain torsion angles of the residues at the interface. In the early rounds of design and minimization, DDMI down-weighted the repulsive component of the Lennard-Jones potential and added coordinate constraints for the backbone C α atoms to prevent the fledgling interface from “exploding”: the typical binding energy for the interfaces that resulted from the first few iterations was positive., so that, without the coordinate constraints, the minimizer displaced the scaffold from the target, preventing the design of any interface. In each iteration, DDMI decreased the weight on the coordinate constraints and increased the weight on the Lennard-Jones repulsive term. We found that after eight rounds of sequence design and structure minimization most trajectories had converged on a local minimum (Figure 2 b).

Global sampling of conformational space (within the target constraints) was achieved by performing independent trajectories that effectively start from uniquely docked complexes. We performed >1 million independent DDMI trajectories with HYP and PAK1. To assess the quality of our designs, we compared our models against 43 naturally occurring protein-protein interfaces with high-resolution crystal structures (2.3 Å or less, Supplementary table S1). We observed three key differences between native interfaces and designed interfaces. In the native interfaces, the minimized Rosetta energies correlated strongly with the interface sizes (Supplementary figure S2). The average *binding energy density*, defined as the binding energy (in Rosetta Energy Units, REU) per buried surface area (Å²) was $-0.013 \text{ REU}/\text{Å}^2$ for the native interfaces (Table 1). Binding energy density was a potent discriminator of designed and natural interfaces; most DDMI trajectories resulted in interfaces with poor binding energy densities. We also observed that native structures had on average 4 unsatisfied hydrogen bonds at their interfaces; whereas DDMI models often contained many more. Finally, we observed that naturally occurring interfaces were packed more tightly than those produced by DDMI as measured by the SASApack score. The SASApack score in Rosetta is derived from examining the difference in the molecular surface areas accessible to a 0.5 Å radius probe and accessible to a 1.4 Å radius probe.¹⁸ This difference indicates surface area on the protein that is not in contact with either water or other protein atoms, and hence reflects the presence of voids that are too small to be filled with water. The score is normalized by the average surface-area difference observed in a large set of crystal structures. A negative SASApack score indicates better packing than crystal structures, a positive SASApack score indicates worse packing. The average SASApack score for the set of native interfaces was -1.39 ± 1.29 , whereas many of the designs had positive SASApack scores.

These observations motivated the set of filters we incorporated into the DDMI protocol. DDMI discarded designs if their binding-energy density was higher than $-0.01 \text{ REU}/\text{Å}^2$ and if they buried 4 or more polar groups which lacked hydrogen bonding partners. Final selection of the design models was made based on the SASApack score of less than 2.0 and a minimum buried surface area of 700 Å². Satisfying all four criteria required considerable sampling; less than 1% of the DDMI trajectories passed these filters. The interfaces that were left were of moderate size (900 Å² – 1700 Å²).

Four designs with favorable values for all of the evaluation metrics were selected for experimental validation: “0233”, “Spider Roll”, “1212” and “3533”. Each of them had the c-terminal helix of the scaffold interacting with the ‘hotspot’ cleft in PAK1. The main interacting residues on PAK1 were L470, L473 and Y474. In some cases V436, R438 and R471 were also contributing to the interaction. On the scaffold side, the mutations were mainly concentrated on the helix IV (Figure 3). Residues involved in interaction were mostly hydrophobic consisting of one or two aromatic amino acids forming the centre of interactions. Design model 1212 had

the greatest number of polar residues at the interface, and design model s032 (introduced below) had the fewest polar amino acids at the interface (Table 2).

Discrete molecular dynamics (DMD) to sample more backbone conformations

To study the effect of additional backbone flexibility on the design process, we used discrete molecular dynamics^{19; 20; 21} to sample conformational space near one of the designed complexes (Spider Roll). From the initial design model, we first performed short DMD simulations and picked ten structure snapshots from the simulation trajectory. Next, we applied the iterative sequence design and minimization phase (DMI phase) from the DDMI protocol to these ten snapshots to generate ten new designs, from which the one with the best Rosetta energies was selected for the next round of DMD simulation. These DMD+DMI iterations were repeated 100 times at decreasing DMD simulation temperature to identify low energy designs. The 1000 designs that resulted were filtered using the same criteria as for the DDMI protocol, and of the remaining structures, the top 10 by total energy were visually examined. This complete protocol was executed twice, and from the top 10 in each execution, we selected one design for experimental validation. The designs were named “s032” and “s037”. The backbone RMSD between s032 and Spider Roll when the target protein was superimposed was 5.36 Å while between s037 and Spider Roll the deviation was 4.53 Å. These RMSDs reflect a substantial shift in the docked conformation of these designs against the target, as well as a significant change to the starting backbone geometries. The two designs moved 1.28 Å and 1.52 Å in C α RMSD away from the starting Spider Roll backbone. In contrast, gradient-based minimization in the standard DDMI protocol produced designs with C α RMSDs to the starting HYP scaffold of under 0.3 Å. The sequences of the two new designs differ significantly from the initial Spider Roll design. Out of the 21 residue sites being designed, the numbers of mutated amino acids are 12 and 18 for s032 and s037, respectively.

Binding measurements and 'hotspot' mapping

The six designs were expressed as MBP fusions. All of them expressed in the soluble fraction of *E. coli* lysate, but 3533 and s032 aggregated when MBP was removed with TEV protease. No further studies were performed with 3533 and s032. The circular dichroism (CD) spectra of the designs indicate that the proteins are helical and all of the designs exhibited cooperative thermal melts as monitored by the CD signal at 222 nm (Supplementary figure S1). The four soluble designs were labeled with the fluorescence probe Bodipy, and fluorescence polarization was used to monitor binding to PAK1. Spider Roll showed the best binding affinity ($K_d = 100 \mu\text{M}$) while design model s037 which was derived from Spider Roll using DMD bound with a dissociation constant of 160 μM . Model 1212 bound with an affinity of 330 μM and model 0233 failed to show any conclusive binding with PAK1.

To probe if Spider Roll was interacting with PAK1 as designed, we mutated residues on both sides of the interface that were predicted to contribute to binding (Figure 4). Consistent with the design model, the PAK1 mutations L473A and L470E each destabilized binding by over 1.4 kcal/mol. R471A and Y474A also destabilized binding by 0.35 and 0.85 kcal/mol respectively. R438, predicted to form a hydrogen bond with backbone carbonyls at the C-terminus of Spider Roll, did not have any effect on binding affinity when mutated to alanine. Mutations to residues on Spider Roll that were designed to form interactions with PAK1 also weakened binding. The point mutations F59A and G56E weakened binding by 0.55 kcal/mol. Y52A and A55E mutations had only subtle effect with a binding energy cost of approximately 0.4 kcal/mol. A more dramatic decrease in binding was observed by combining F59A with G56E (1.4 kcal/mol). I58A mutation, which was predicted to contact PAK1 at the edge of the interface, had very minimal effect on binding. The design scaffold (with solubilizing mutations G26E, L37E and L38N) had basal level of affinity for PAK1. Taken together the mutational

data indicates that Spider Roll interacts with the target patch on PAK1, and that helix IV of Spider Roll is involved in binding.

NMR-based structure prediction of Spider Roll

In order to structurally characterize Spider Roll as well as its binding to PAK1, we nominated Spider Roll mutant I58A, which expressed with a more than two-fold increases yield when compared with Spider Roll (see Methods), as a community outreach target of the Protein Structure Initiative (PSI) 2 and collaborated with the Northeast Structural Genomics Consortium (NESG: <http://www.nesg.org>) to obtain sequential polypeptide backbone and $^{13}\text{C}\beta$ chemical shift assignments. The 2D- ^{15}N , ^1H HSQC spectrum²² recorded for Spider Roll I58A shows favorable chemical shift dispersion and indicates that the protein is folded in solution (Supplementary figure S3). Assignment completeness of detectable peaks in the 2D- ^{15}N , ^1H HSQC spectrum was 91% (48/53). Polypeptide backbone and $^{13}\text{C}\beta$ chemical shift assignments (Supplementary table S2) were obtained for 50 residues and a total of 81% of the shifts assignable with the selected set (see Methods) of multidimensional NMR experiments (i.e. excluding the N-terminal $^{15}\text{NH}_3^+$, the three prolyl ^{15}N and the $^{13}\text{C}'$ shifts of residues preceding prolyl residues). The chemical shifts are in agreement with the location of α -helices in the X-ray crystal structure of the design scaffold protein (HYP), except for the last ~6 residues of helix IV (Supplementary figure S4; Supplementary table S3). The shifts were then used to predict the structure of Spider Roll I58A with the program CS-ROSETTA²³ (Supplementary figure S5). The CS-ROSETTA structure is very similar structure of design template protein HYP (backbone RMSD = 0.7 Å for helical residues; residues 1010–1022, 1026–1036, 1041–1049 and 1051–1065 for 1I2T and residues 2–14, 18–28, 33–41 and 43–57 for the current structure) which indicates that the redesign of HYP did not significantly affect the fold of the protein.

The chemical shift indices suggest that the last ~6 residues of the C-terminal helix (helix IV) of Spider Roll are frayed in solution. To further investigate the conformation of helix IV, we derived amide proton – amide proton upper distance limit constraints from 3D ^{15}N -resolved ^1H , ^1H -NOESY. The longer distances derived for the C-terminal segment of helix IV reflect weaker NOEs (the sequential NOEs between the last 6 residues either overlap or disappear), which are consistent with fraying of this segment. All of our designed interfaces include a fully intact C-terminal helix that makes close contact with PAK1. Fraying of the helix IV in the unbound state is not inconsistent with a fully folded helix in the bound state, but indicates that there will be an additional entropic penalty associated with binding. Hence, future improvement of the design of Spider Roll may focus on stabilizing the C-terminal segment of helix IV.

NMR characterization of Spider Roll-PAK1 binding

The 2D- ^{15}N , ^1H HSQC spectrum²² was recorded for Spider Roll I58A and monitored as a function of PAK1 concentration. The spectra were recorded at three different molar ratios of Spider Roll I58A / PAK1: 231 μM / none, 210 μM / 120 μM , 148 μM / 487 μM . Unexpectedly, addition of PAK1 did not induce any perturbation of Spider Roll chemical shifts or introduce large broadening of Spider Roll resonances, but lead to a dramatic decrease of Spider Roll signal intensities. Specifically, at a PAK1 concentration of 487 μM , Spider Roll peak intensities were reduced to <5% of their starting values (Figure 5) However, no 'new' peaks emerged during the titration that could be attributed to Spider Roll bound to PAK1. This is likely due to the fact that the signals of the 43 kDa complex present at a concentration of only ~200 μM are too weak to be detected. Assuming for simplicity that both ^1H and ^{15}N line-widths scale linearly with molecular weight²², the S/N ratios for Spider Roll in complex with PAK1 are expected to be about 40-fold lower than for free Spider Roll.²⁴ Since the S/N ratios observed for free Spider Roll are in the range of ~20 in the 2D- ^{15}N , ^1H HSQC spectra, detection of the signal of bound Spider Roll is indeed not be expected.

Furthermore, Spider Roll may bind to PAK1 in different and slowly exchanging conformations, a phenomenon which would broaden resonance lines, thereby further impeding signal detection for bound Spider Roll. A third scenario which would manifest itself by lack of signals for the bound protein would be the formation of aggregates formed by the complex.

To check if the Spider Roll-PAK1 complex aggregates non-specifically, we performed size exclusion chromatography with the NMR sample and also ran an SDS page gel for an aliquot of the NMR sample. The SDS gel confirmed that the same amount of Spider Roll remained in solution as was initially added to the NMR sample, and the size exclusion chromatography indicated that the sample was not aggregating: the only two peaks in the chromatogram corresponded to monomeric PAK1 and Spider Roll, which is typically observed for proteins with micromolar binding affinities. Hence, these experiments suggest that the absence of NMR peaks from the bound state is indeed likely due to the slower overall rotational tumbling of the complex.

Our NMR data are consistent with the lifetime of the Spider Roll-PAK1 complex being much longer than the time required for signal detection ('NMR chemical shift time scale', around 0.1 s), that is, we obtain as an upper bound for the 'off-rate' $k_{\text{off}} \ll 10 \text{ s}^{-1}$. With the dissociation constant $K_D = k_{\text{off}}/k_{\text{on}} = 10^{-4} \text{ M}$, we then obtain for the 'on-rate' $k_{\text{on}} \ll 10^5 \text{ M}^{-1} \text{ s}^{-1}$. The on-rate constants for protein-protein binding can vary dramatically ($1 \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$ to $1 \times 10^9 \text{ M}^{-1} \text{ s}^{-1}$) but are often near $1 \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$.^{25; 26} The comparably small on-rate constant for Spider Roll could reflect the absence of an effective docking funnel or a conformational change that accompanies binding, such as folding of the end of the helix IV in Spider Roll.

To validate that the reductions of Spider Roll peak intensities in 2D- $[^{15}\text{N}, ^1\text{H}]$ HSQC are due to binding to the kinase domain, the NMR titration was repeated for PAK1 mutant L470E, which shows reduced affinity for Spider Roll when monitored using fluorescence polarization experiments. Consistently, the changes in peak volume were much smaller than those observed with wild type PAK1: at a PAK1 L470E concentration of 470 μM , the Spider Roll I58A peak volumes were still 40% of their original size (in contrast to <5% for WT PAK1) (Figure 5, Supplementary figure S6).

Does Spider Roll adopt multiple docked positions when binding PAK1?

Both the NMR data and mutational data indicate that Spider Roll binds the target cleft on PAK1, but they do not rule out the possibility that it can adopt alternative docked orientations when bound to PAK1. To further examine this possibility we used Rosetta to perform protein-protein docking simulations with Spider Roll and PAK1. In these simulations, Spider Roll was constrained to be near the target binding site, but was allowed to adopt alternative orientations relative to PAK1. Many independent trajectories were used to probe the energy landscape and the energies of the various models were plotted versus RMSD to the target conformation. We identified two clusters of low energy structures (Figure 6 a). The lowest energy cluster was centered on the design model, but the second cluster packed helix IV in a direction that was orthogonal to the design model (Figure 6 b, c). The mutational data does not strongly distinguish between the two alternatives. The mutations that have the strongest effect on binding energy are buried in both sets of models (Table 4) as calculated by the NACCESS program.²⁷

Spider Roll binds preferentially to the activated form of full-length PAK1

In its inactive form, full-length PAK1 (PAK-fl) forms a closed conformation in which an autoinhibitory domain binds to the same cleft in the kinase domain that we have targeted with Spider Roll. PAK1 can be opened by introducing mutations in the autoinhibitory elements (V127E, S144E) that weaken affinity for the kinase domain (unpublished data). Using fluorescence polarization we measured the affinity of Spider Roll for WT PAK-fl and PAK1-

fl with the mutations V127E and S144E. Spider Roll showed no binding with WT PAK1, but bound PAK1 V127E, S144E with an affinity of 200 μM (Figure 7).

Discussions

The computational design of novel protein-protein interfaces with high specificity and affinity remains an unsolved problem. Here we have explored the effectiveness of an interface design protocol that includes rigid-body docking, rotamer-based sequence design and minimization of side chain and backbone torsion angles. With this protocol we were able to design small interfaces between HYP and PAK1 ($\Delta\text{SASA} \sim 1000 \text{ \AA}^2$) that had calculated binding energies comparable to naturally occurring interfaces of the same size. Design models with interfaces larger than 1500 \AA^2 either exhibited inferior packing, buried too many unsatisfied hydrogen bonding groups or had significantly worse binding energy densities than native complexes. These results suggest that more extensive backbone sampling and/or screening of multiple scaffolds will be needed to build larger interfaces with tight packing and good hydrogen bonding.

Three of the six experimentally characterized designs bound PAK1. Spider Roll, which was the tightest binder, bound PAK1 with an affinity of 100 μM . For comparison, the truncated autoinhibitory domain (trunc-AID, residues 83–137) from PAK1 binds to the same target cleft with an affinity of 4 μM (unpublished data). Even though the AID is smaller (55 residues) than the HYP scaffold (61 residues), the interface between trunc-AID and PAK1 is larger than the Spider Roll interface, 1620 \AA^2 versus 970 \AA^2 . The calculated binding energy for the trunc-AID-PAK1 complex (after minimization with Rosetta) is -25 REU ($-0.016 \text{ REU} / \text{ \AA}^2$) compared to -15 REU ($-0.015 \text{ REU} / \text{ \AA}^2$) for Spider Roll. The more favorable energies achieved by the AID domain (experimental and computational) may reflect the evolution of this domain to have a backbone conformation suitable for binding the PAK1 kinase domain, while the backbone conformation of HYP has not been optimized for binding PAK1.

Mutation studies and NMR studies do not rule out the possibility that Spider Roll samples multiple docked positions when bound to PAK1. This may partially reflect the hydrophobic nature of the designed interface; the buried SASA is 82% nonpolar. For native protein-protein interfaces (Supplementary table S1) the average is 69% nonpolar (or 31% polar, Table 1). Polar interactions can increase specificity because of the strong distance and orientation dependence of hydrogen bond energies. The interfaces designed in this study were predisposed to be hydrophobic because of the hydrophobic nature of the target cleft; however, several polar residues on PAK1 surround the cleft. The AID domain forms hydrogen bonds with N468 and R471 on PAK1, while these residues are left solvent exposed in the Spider Roll design. Designing multiple hydrogen bonds across a protein interface is challenging because optimizing one bond with rigid-body or backbone perturbations is likely to disrupt the surrounding bonds.

Spider Roll binds to the open state of full length PAK1 but does not bind to the closed inactive form of the protein. Affinity reagents that distinguish the activation state of target molecules are useful starting points for creating biosensors. Typically the proteins are modified with fluorescent groups (chemically or genetically encoded) that change their fluorescent properties when the affinity reagent binds the target. Our results suggest that computational design may be one route for identifying novel affinity reagents. However, to have useful biosensors it will be necessary to have binding affinities tighter than 100 μM . A potential strategy for achieving this goal is to combine computational design with affinity maturation via techniques such as phage or yeast display. This strategy was recently used to improve the rate of catalysis by a computationally designed enzyme.²⁸

Materials and Methods

Computational methods

Rosetta—Our interface design protocol, DDMI, is described in the Results section and Fig 1. It was implemented within the Rosetta molecular modeling program.^{15; 16} A single trajectory consists of a 'dock' stage followed by several iterations of 'design' and 'minimization' stages. The 'dock' stage involves a low resolution rigid body Monte Carlo search,¹² where the side chain of each residue was represented as one bead placed at centroid position of the side chain. We applied constraints so that the 'scaffold' remains close to and packs one or more of L470, L473 and Y474 residues on the 'target'. A trajectory satisfying the constraints was designed with 8 iterations of sequence and backbone optimizations, similar to protocols described earlier.²⁹

Sequence optimization was done using Monte-Carlo simulated annealing protocol^{11; 30} which optimized the total energy of the complex. The all-atom energy function in Rosetta is a linear weighted sum of 12–6 Lennard-Jones potential, the Lazaridis-Karplus implicit solvation model³¹, an orientation-dependant hydrogen bonding potential³², backbone-dependent rotamer probabilities³³, a knowledge-based electrostatic energy³⁴, amino acid probabilities conditioned on ϕ and ψ space³⁵ and reference energies that approximate the unfolded state energy of an amino acid.¹¹

DDMI protocol can be found in Rosetta version 2.1 (<http://www.rosettacommons.org/>). The command line is:

```
rosetta.gcc -design -dock_des_min_inter -s input.pdb -linmem_ig -ex1 -ex2 -exOH  
-extrachi_cutoff 1 -no_his_his_pairE -series qq -protein input -chain _ -  
read_all_chains - multi_chain -randomize2 -resfile input.resfile -  
multi_cool_annealer -try_both_his_tautomers -set_interface_cutoff 7.0 -nstruct 100  
-ddmi_dUns_filter 4 -ddmi_dG_dSASA_ratio_filter_0.013
```

The design-minimize phase of DDMI protocol can be used on an already docked conformation. The command line is:

```
rosetta.gcc -design -design_min_inter -series qq -protein input -s input.pdb -chain  
_ -resfile input.resfile -no_his_his_pairE -read_all_chains -multi_chain -  
design_trials -extrachi_cutoff 1 -ex1 -ex2aro_only
```

Binding parameters calculation—The binding energy of a complex was calculated by subtracting the individual Rosetta energies of the chains from the total energy of the complex. Total buried surface area at the interface was similarly calculated by taking the difference in SASAs between the bound and unbound states. The number of buried unsatisfied hydrogen bonding groups was determined by counting the number of polar groups at the interface which were fully buried yet lacked hydrogen bonding partner. The side chain rotamers were left in the same conformations as in the complex when the calculations were made. The interface SASApack scores were reported as the average per-residue SASApack score for those residues within 5 Å of atoms on the opposite chain.

Discrete molecular dynamics—DMD is a flavor of molecular dynamics simulation (MD) approach.^{19; 20; 21} Unlike traditional MD which uses continuous physical force field, in DMD, the interaction between two atoms is described by a simplified step-wise potential. For step-wise potentials, the derivations are zero between the potential steps, and the integrations can be ignored except when the two atoms go across the steps wall (collision events). The space and time of the collision events can be analytically calculated based on previous positions and velocities of the two atoms. The new velocities of the atoms after the collision can also be

calculated based on laws of conservation of energy, momentum, and angular momentum. The system evolves by calculation and sorting of further collision events. Compared with traditional MD that is driven by numerical integration over fine time steps, DMD is driven by collision events and is more efficient since it allows larger integration time steps in average. Extra efficiency is also gained by applying fast event sorting and updating algorithms. Although the actual speed up of DMD compared to traditional MD varies from system to system, it can reach 3–10 orders of magnitude.

Experimental methods

Plasmid constructs, gene synthesis and mutagenesis—Kinase dead mutant (K298R) of full length PAK1 (PAK1-fl) and PAK1 V127E/S144E (PAK-fl mutant) were cloned in pQE-80 L (Qiagen) vector. Kinase domain (K298R mutant), residues 250–545 from PAK-fl (here referred to as PAK1), was also cloned in pQE-80L vector with an extra sequence (MRGSHHHHHHGSDYDIPTTENLYFQC) in N-terminus. The PAK1 mutants were made by overlap extension using PCR.³⁶

Constructs for protein designs and Spider Roll mutants (from here referred as designed variants) were made as an MBP fusion with a TEV protease cleavage site and cloned in pQE-80L vector so that 6×His-tag remained at N-terminus after the expression. The genes for designed variants were synthesized using a gene synthesis protocol³⁷ where the codons were optimized for bacterial expression.

Expression and protein purification—Proteins were expressed in BL21(DE3) pLysS cells. The cells were grown up to OD₆₀₀ 0.6–0.8 and then induced with 0.3 mM IPTG and further grown at 25 °C for 6 h. Cells were then disrupted using sonication and resulting lysates were cleared by two rounds of centrifugation (18,000 × g) of 20 mins each. The supernatants of PAK1-fl and PAK1 variants were then purified using a prepacked Ni-NTA column (HisTrap, HP, GE Healthcare) followed by an anion exchange step (Source 15Q beads, GE Healthcare) and gel filtration chromatography (Superdex 75 or Superdex 200, GE Healthcare). Designed proteins (referred to as design variants) were purified using a Ni-NTA column followed by an overnight proteolysis using TEV protease and then again loading the product on the Ni-NTA column. The second Ni-NTA affinity step removed the 6×His-tagged MBP to give design variants (all with an extra GS sequence at N-terminus) in the flow through. The flow through was concentrated and loaded on gel filtration column (Superdex 75, GE Healthcare). All the design variants eluted at similar elution volume (size corresponding to monomer) and the appropriate fractions were combined and concentrated (Amicon Ultra, Millipore). Protein concentration was estimated using theoretical molar extinction coefficients and absorbance at 280 nm.

U-¹⁵N and *U*-¹⁵N, ¹³C-labeled Spider Roll I58A was expressed in M9 minimal media containing ¹⁵N-ammonium chloride (1 g/l). Glucose was replaced with ¹³C-glucose for ¹³C labeling. Cells were grown in LB broth upto OD₆₀₀ 0.6 to 0.8 and then spun down by centrifugation at 3000 rpm for 30 mins. The cell pellet was then resuspended in M9 minimal media, left for recovery for 20 mins under 250 rpm shaking and then induced with 0.3 mM IPTG. The expression was carried at 16 °C for 14 h.

Fluorescence polarization assay—Binding affinities was measured using fluorescence polarization.³⁸ The thiol-reactive fluorescent probe Bodipy(507/545)-iodoacetamide (Molecular Probes) was conjugated to design variants at the unique cysteine site. Design variants at a concentration range of 60–250 μM were buffer exchanged into 50 mM Tris-Cl, pH 7.5 using a PD10 desalting column (GE Healthcare) and spiked with 1 mM TCEP. A 20 mM stock solution of Bodipy(507/545-IA) suspended in dimethyl sulfoxide (DMSO) was

added drop-by-drop to 3–10 fold molar excess of the designed variants with constant mixing by inverting the tubes. The tubes were wrapped with aluminium foil and left overnight at 4 °C. 50 mM β-mercaptoethanol (BME) was used to quench the reaction. The mix was then centrifuged to pellet the unconjugated Bodipy and run over a PD10 desalting column (equilibrated with 50 mM Tris-Cl, pH 7.5 and 5 mM BME) to separate the conjugated design variants from free Bodipy. Bodipy labeled design variants were quantified using UV/Vis spectrophotometer (theoretical molar extinction coefficient at 280 nm for protein and molar extinction coefficient of 69,000 M⁻¹cm⁻¹ for the probe). A correction factor (Abs_{280 nm}/Abs_{580 nm}) of 0.03 was used to correct for the absorption of the conjugated probe at 280 nm. Typically probe conjugation efficiency (probe/protein) of 20–90 % was achieved for the design variants.

Fluorescence polarization assays were carried out on a Jobin Yvon Horiba Spex FluoroLog-3 instrument (Jobin Yvon Inc) performed in L-format with the excitation wavelength set at 508 nm and emission wavelength set at 545 nm. Bodipy-design variants (in 50 mM Tris-Cl, pH 7.5, 5 mM BME) at a final concentration of 5–10 μM and volume 180 μl were titrated with PAK1-fl or PAK1 variants (in 50 mM Tris-Cl, pH 7.5, 50 mM NaCl, 5 mM BME). Titrations were performed in 3×3-mm quartz cuvette. Slit width was adjusted to 3.5 nm to give a fluorescence intensity ~3.5 million counts per second. Two polarization readings consisting of 3 averaged measurements were collected for at least 15 concentrations of titrant. Data were averaged and any change in polarization occurring due to addition of buffer was subtracted. Data was then analyzed using a model for single site binding model according to the equation (1) which was incorporated in equation (2) to account for the observed polarization.

$$[A:B] = \frac{([A_t] + [B_t] + K_d) - \sqrt{([A_t] + [B_t] + K_d)^2 - 4[A_t] \cdot [B_t]}}{2} \quad (1)$$

$$P_{obs} = \frac{(P_{max} - P_0)[A:B]}{[A_t]} + P_0 \quad (2)$$

where [A:B] is the concentration of Bodipy labeled design variants and titrant (PAK1-fl or PAK1 variants) complex formed, [A_t] is the total concentration of Bodipy labeled design variants, [B_t] is the concentration of the titrant, K_d is the dissociation constant for the interaction, P₀ is the polarization in the absence of titrant, P_{max} is the maximum polarization when all Bodipy labeled design variants are bound to the titrant, and P_{obs} is the observed polarization at a given concentration of titrant. The data were fit according to equation (2) using non-linear regression with SigmaPlot software to obtain fitted parameters for K_d, P_{max} and P₀. Delta polarization (ΔPolarization) was calculated using equation (3) and fraction bound was calculated using equation (4).

$$\Delta Polarization = P_{obs} - P_{min} \quad (3)$$

$$\text{Fraction bound} = \frac{P_{obs} - P_{min}}{P_{max} - P_{min}} \quad (4)$$

where P_{max} and P_{min} are maximum and minimum polarization calculated from the fit. For weak-binding mutants that had relatively small changes in P_{obs}, the fit was forced to have the same P_{max} observed for the Spider Roll-PAK1 interaction (green circles in Figure 4 b). Inability

to achieve very high titrant concentrations prevented reaching saturation of binding sites (>10-fold than K_d).

NMR spectroscopy—As Spider Roll I58A mutant expressed at more than 2-fold higher yield when compared with Spider Roll and since the I58A mutation had a minimal effect on the binding affinity (Figure 4 b), we prepared U - ^{15}N and U - ^{15}N , ^{13}C -labeled NMR samples of Spider Roll I58A.

For assignment of polypeptide backbone and $^{13}\text{C}\beta$ chemical shifts, NMR spectra were recorded at 288 K on a Varian INOVA 600 spectrometer equipped with a cryogenic probe. 2D- ^{15}N , ^1H] HSQC (1 h measurement time) was acquired along with four through-bond correlated NMR experiments²² that is, HNCA (36 h), HN(CO)CA (36 h), HNCACB (42 h), CACB(CO)NH (42 h) and HNCO (9 h) for sequential resonance assignment. In addition, a 3D ^{15}N -resolved ^1H , ^1H] NOESY spectrum (42 hours) was acquired to confirm sequential resonance assignments and to derive backbone ^1HN - ^1HN distance constraints. All spectra were processed and analyzed using the program packages NMRPipe³⁹ and XEASY⁴⁰, respectively. Specifically, the titration of Spider Roll with PAK1 was monitored by recording 2D- ^{15}N , ^1H] HSQC spectra at 288 K on a Varian INOVA-700 MHz spectrophotometer equipped with a cryogenic probe. Isotope labeled proteins and kinase domain were buffer exchanged in 20 mM Phosphate buffer, pH 7.0, 50 mM NaCl and 1 mM DTT and 10% D_2O was added. Peak volumes and line-widths in 2D- ^{15}N , ^1H] HSQC were measured by using the 'peak detection module' of NMRDraw in NMRPipe. Signal-to-noise ratios (SNR) were calculated by dividing the volume by the noise level measured by the 'estimate noise module' of NMRpipe followed by division by the square root of the measurement time. To compare peak volume reductions arising from different degrees of Spider Roll-PAK1 complex formation, the thus obtained SNR per unit time was normalized to the volume measured in the absence of PAK1.

Chemical shifts were deposited in the BioMagResBank (accession code: 16710).⁴¹

Circular dichroism (CD)—CD data were collected on a Jasco J-815 CD spectrometer. Far-UV CD scans (250–195 nm) were carried at a typical protein concentration of 20 μM in a 1 mm cuvette. The temperature was maintained at 20 $^\circ\text{C}$ using a Peltier device. Variable temperature scan was carried between 4–96 $^\circ\text{C}$ while measuring the ellipticity at 222 nm. All ellipticity data were corrected with a buffer blank and then converted to mean residue ellipticity.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

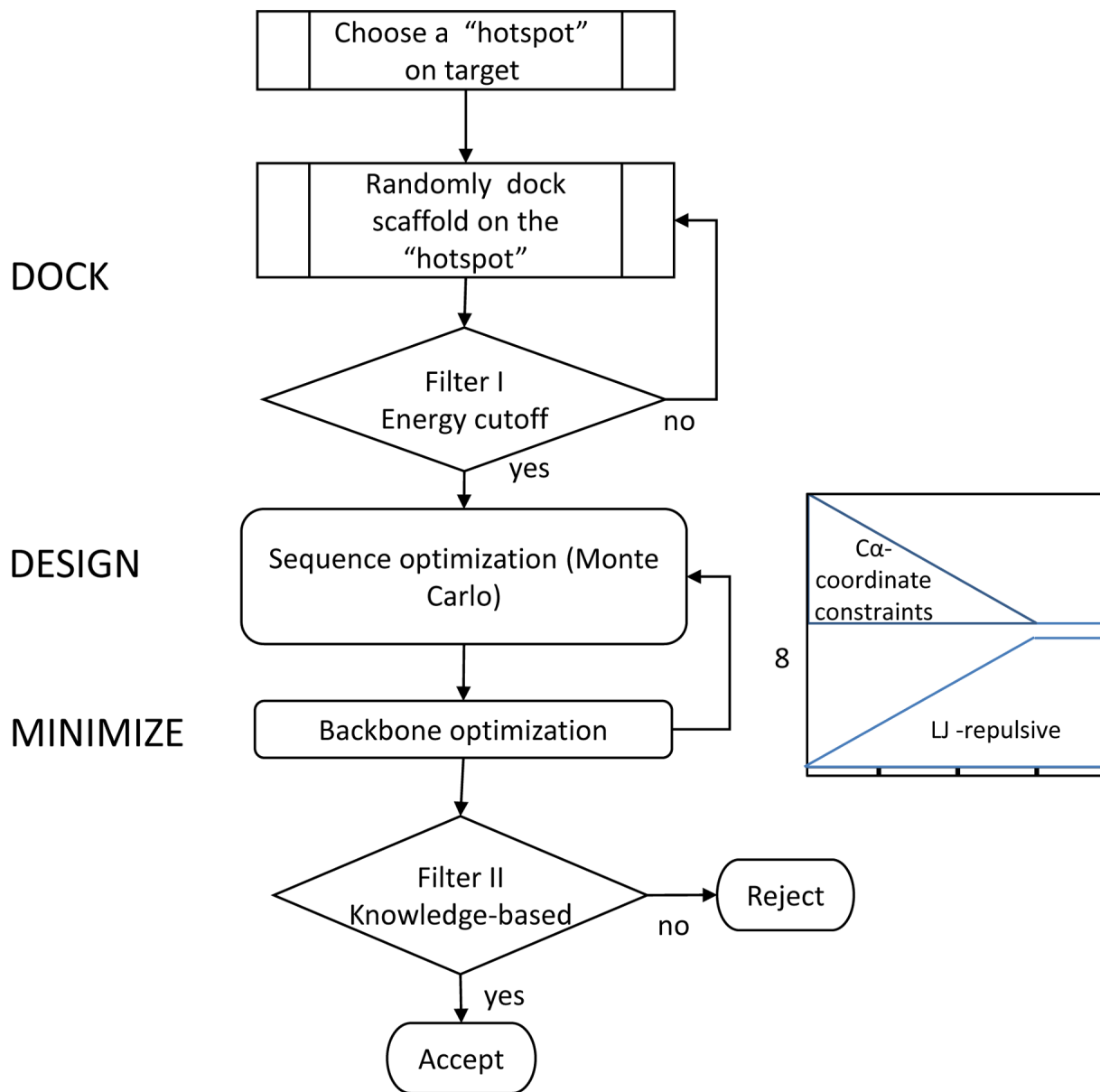
Computational services provided in part by RENCI (Renaissance Computing Institute), University of North Carolina at Chapel Hill. The work was supported by DARPA and NIH (to B.K.). This work is supported by the National Institute of Health Grant R01GM080742 and the ARRA supplement 3R01GM080742-03S1 (to N.V.D.) Spider Roll protein was selected as a community outreach target by the Northeast Structural Genomics Consortium (<http://www.nesg.org>; NESG ID OR24). We thank Dr. Ashutosh Tripathy from UNC Macromolecular Interactions Facility and Dr. Greg Young from UNC Biomolecular NMR Lab for their help during the experiments. We also thank Dr. Yi Wu and Dr. Klaus Hahn for PAK1 clones and useful discussions about PAK1.

References

1. Karanicolas J, Kuhlman B. Computational design of affinity and specificity at protein-protein interfaces. *Curr Opin Struct Biol* 2009;19:458–63. [PubMed: 19646858]
2. Kortemme T, Baker D. Computational design of protein-protein interactions. *Curr Opin Chem Biol* 2004;8:91–7. [PubMed: 15036162]

3. Mandell DJ, Kortemme T. Computer-aided design of functional protein interactions. *Nat Chem Biol* 2009;5:797–807. [PubMed: 19841629]
4. Grigoryan G, Reinke AW, Keating AE. Design of protein-interaction specificity gives selective bZIP-binding peptides. *Nature* 2009;458:859–64. [PubMed: 19370028]
5. Havranek JJ, Harbury PB. Automated design of specificity in molecular recognition. *Nat Struct Biol* 2003;10:45–52. [PubMed: 12459719]
6. Harbury PB, Plecs JJ, Tidor B, Alber T, Kim PS. High-resolution protein design with backbone freedom. *Science* 1998;282:1462–7. [PubMed: 9822371]
7. Binz HK, Amstutz P, Kohl A, Stumpp MT, Briand C, Forrer P, Grutter MG, Pluckthun A. High-affinity binders selected from designed ankyrin repeat protein libraries. *Nat Biotechnol* 2004;22:575–82. [PubMed: 15097997]
8. Kohl A, Amstutz P, Parizek P, Binz HK, Briand C, Capitani G, Forrer P, Pluckthun A, Grutter MG. Allosteric inhibition of aminoglycoside phosphotransferase by a designed ankyrin repeat protein. *Structure* 2005;13:1131–41. [PubMed: 16084385]
9. Huang PS, Love JJ, Mayo SL. Adaptation of a fast Fourier transform-based docking algorithm for protein design. *J Comput Chem* 2005;26:1222–32. [PubMed: 15962277]
10. Huang PS, Love JJ, Mayo SL. A de novo designed protein protein interface. *Protein Sci* 2007;16:2770–4. [PubMed: 18029425]
11. Kuhlman B, Baker D. Native protein sequences are close to optimal for their structures. *Proc Natl Acad Sci U S A* 2000;97:10383–8. [PubMed: 10984534]
12. Gray JJ, Moughon S, Wang C, Schueler-Furman O, Kuhlman B, Rohl CA, Baker D. Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. *J Mol Biol* 2003;331:281–99. [PubMed: 12875852]
13. Lei M, Lu W, Meng W, Parrini MC, Eck MJ, Mayer BJ, Harrison SC. Structure of PAK1 in an autoinhibited conformation reveals a multistage activation switch. *Cell* 2000;102:387–97. [PubMed: 10975528]
14. Deo RC, Sonenberg N, Burley SK. X-ray structure of the human hyperplastic discs protein: an ortholog of the C-terminal domain of poly(A)-binding protein. *Proc Natl Acad Sci U S A* 2001;98:4414–9. [PubMed: 11287654]
15. Das R, Baker D. Macromolecular modeling with rosetta. *Annu Rev Biochem* 2008;77:363–82. [PubMed: 18410248]
16. Rohl CA, Strauss CE, Misura KM, Baker D. Protein structure prediction using Rosetta. *Methods Enzymol* 2004;383:66–93. [PubMed: 15063647]
17. Wang C, Bradley P, Baker D. Protein-protein docking with backbone flexibility. *J Mol Biol* 2007;373:503–19. [PubMed: 17825317]
18. Leaver-Fay A, Butterfoss GL, Snoeyink J, Kuhlman B. Maintaining solvent accessible surface area under rotamer substitution for protein design. *J Comput Chem* 2007;28:1336–41. [PubMed: 17285560]
19. Dokholyan NV, Buldyrev SV, Stanley HE, Shakhnovich EI. Discrete molecular dynamics studies of the folding of a protein-like model. *Fold Des* 1998;3:577–87. [PubMed: 9889167]
20. Ding F, Dokholyan NV. Simple but predictive protein models. *Trends Biotechnol* 2005;23:450–5. [PubMed: 16038997]
21. Ding F, Tsao D, Nie H, Dokholyan NV. Ab initio folding of proteins with all-atom discrete molecular dynamics. *Structure* 2008;16:1010–8. [PubMed: 18611374]
22. Cavanagh, J.; Fairbrother, WJ.; Palmer, AG.; Rance, M.; Skegton, NJ. *Protein NMR Spectroscopy*. Academic Press; San Diego: 2007.
23. Shen Y, Lange O, Delaglio F, Rossi P, Aramini JM, Liu G, Eletsky A, Wu Y, Singarapu KK, Lemak A, Ignatchenko A, Arrowsmith CH, Szyperski T, Montelione GT, Baker D, Bax A. Consistent blind protein structure generation from NMR chemical shift data. *Proc Natl Acad Sci U S A* 2008;105:4685–90. [PubMed: 18326625]
24. Teng, Q. *Structural biology: Practical NMR applications*. Springer; 2005.
25. Gabdouliline RR, Wade RC. Biomolecular diffusional association. *Curr Opin Struct Biol* 2002;12:204–13. [PubMed: 11959498]

26. Alsallaq R, Zhou HX. Electrostatic rate enhancement and transient complex of protein-protein association. *Proteins* 2008;71:320–35. [PubMed: 17932929]
27. Hubbard, SJ.; Thornton, JM. 'NACCESS', Computer Program. Department of Biochemistry and Molecular Biology, University College London; 1993.
28. Rothlisberger D, Khersonsky O, Wollacott AM, Jiang L, DeChancie J, Betker J, Gallaher JL, Althoff EA, Zanghellini A, Dym O, Albeck S, Houk KN, Tawfik DS, Baker D. Kemp elimination catalysts by computational enzyme design. *Nature* 2008;453:190–5. [PubMed: 18354394]
29. Kuhlman B, Dantas G, Ireton GC, Varani G, Stoddard BL, Baker D. Design of a novel globular protein fold with atomic-level accuracy. *Science* 2003;302:1364–8. [PubMed: 14631033]
30. Leaver-Fay A, Snoeyink J, Kuhlman B. On-the-fly rotamer pair energy evaluation in protein design. *ISBRA* 2008:343–354.
31. Lazaridis T, Karplus M. Effective energy function for proteins in solution. *Proteins* 1999;35:133–52. [PubMed: 10223287]
32. Kortemme T, Morozov AV, Baker D. An orientation-dependent hydrogen bonding potential improves prediction of specificity and structure for proteins and protein-protein complexes. *J Mol Biol* 2003;326:1239–59. [PubMed: 12589766]
33. Dunbrack RL Jr, Cohen FE. Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Sci* 1997;6:1661–81. [PubMed: 9260279]
34. Simons KT, Ruczinski I, Kooperberg C, Fox BA, Bystroff C, Baker D. Improved recognition of native-like protein structures using a combination of sequence-dependent and sequence-independent features of proteins. *Proteins* 1999;34:82–95. [PubMed: 10336385]
35. Bowers PM, Strauss CE, Baker D. De novo protein structure determination using sparse NMR data. *J Biomol NMR* 2000;18:311–8. [PubMed: 11200525]
36. Ho SN, Hunt HD, Horton RM, Pullen JK, Pease LR. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* 1989;77:51–9. [PubMed: 2744487]
37. Stemmer WP, Cramer A, Ha KD, Brennan TM, Heyneker HL. Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene* 1995;164:49–53. [PubMed: 7590320]
38. Eletr ZM, Huang DT, Duda DM, Schulman BA, Kuhlman B. E2 conjugating enzymes must disengage from their E1 enzymes before E3-dependent ubiquitin and ubiquitin-like transfer. *Nat Struct Mol Biol* 2005;12:933–4. [PubMed: 16142244]
39. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 1995;6:277–93. [PubMed: 8520220]
40. Bartels C, Xia T-H, Billeter P, Guntert P, Wuthrich K. The program XEASY for computer-supported NMR spectral analysis of biological macromolecules. *J Biomolecular NMR* 1995;5:1–10.
41. Ulrich EL, Akutsu H, Doreleijers JF, Harano Y, Ioannidis YE, Lin J, Livny M, Mading S, Maziuk D, Miller Z, Nakatani E, Schulte CF, Tolmie DE, Kent Wenger R, Yao H, Markley JL. *BioMagResBank. Nucleic Acids Res* 2008;36:D402–8. [PubMed: 17984079]

**Figure 1.**

DDMI protocol for protein interface design. Each trajectory starts with rigid-body docking using a low resolution score function with additional constraints to direct predetermined residues to the interface. A docked conformation with a high score (binding energy >18 REU with a low resolution energy function¹⁶) was rejected (Filter I). When a trajectory passes the filter, it goes through 8 rounds of sequence optimization and backbone minimization during which LJ-repulsive are ramped upwards and coordinate constraints are ramped downwards for 6 rounds. During this process the Lennard-Jones repulsive term of the energy function is ramped from 0.85 to 1, whereas the coordinate constraint weight goes from 1 to 0. Designs are filtered before being output based on a binding energy density (<-0.01 REU/ \AA^2) and a maximum number of 4 unsatisfied polar groups at the interface (Filter II).

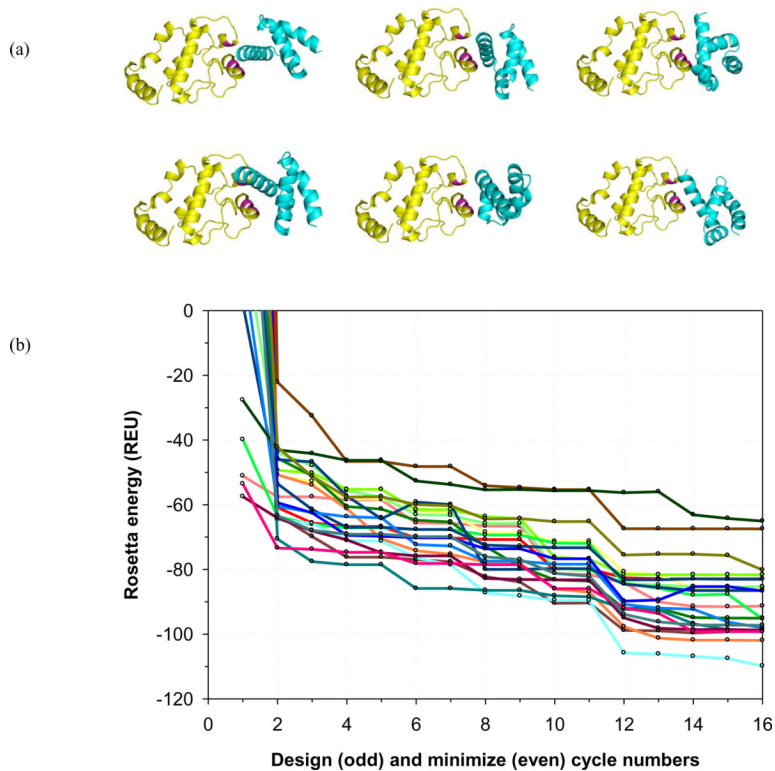


Figure 2. Conformational sampling in (a) 'dock' stage and (b) energy convergence during the 'design/minimize' stage. In (a) the region in 'magenta' depicts the 'hotspot' region on the target protein. In (b) each line represents an independent trajectory. A drop in energy is observed after cycles 11 and 12 during which the coordinate constraints are ramped to the final value and are sufficiently weak to let the proteins relieve the strain induced while holding the two partners next to each other.

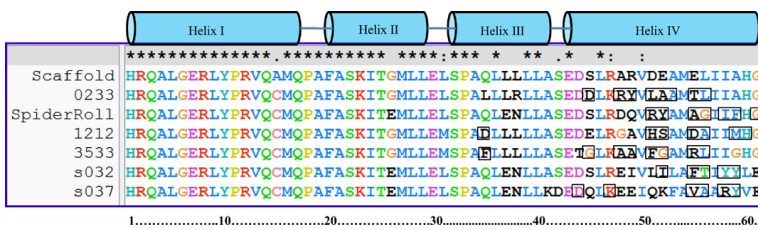
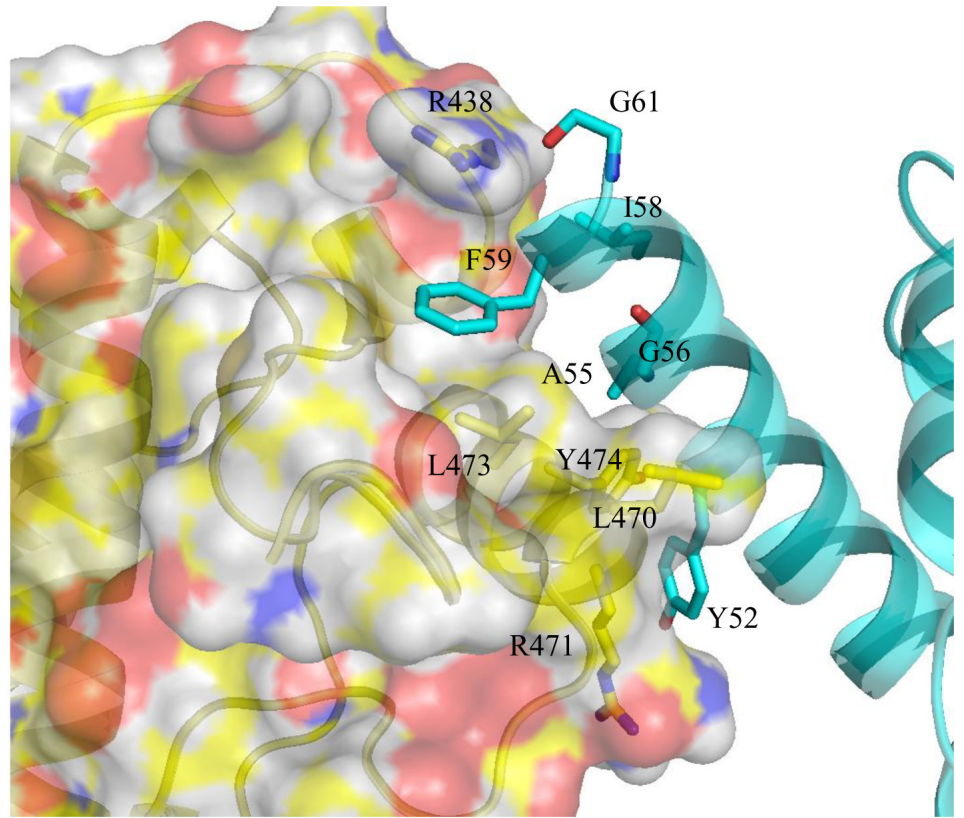
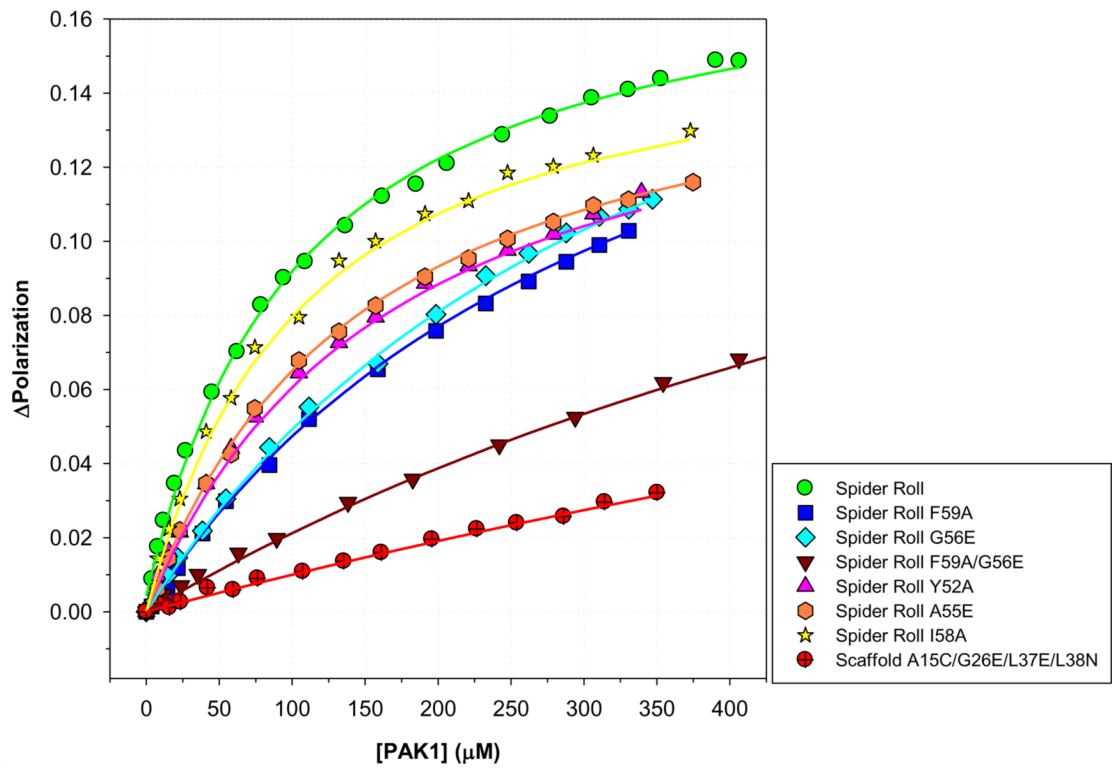


Figure 3. Multiple sequence alignment of the design models selected for experimental validation. The interface residues in each case are highlighted.

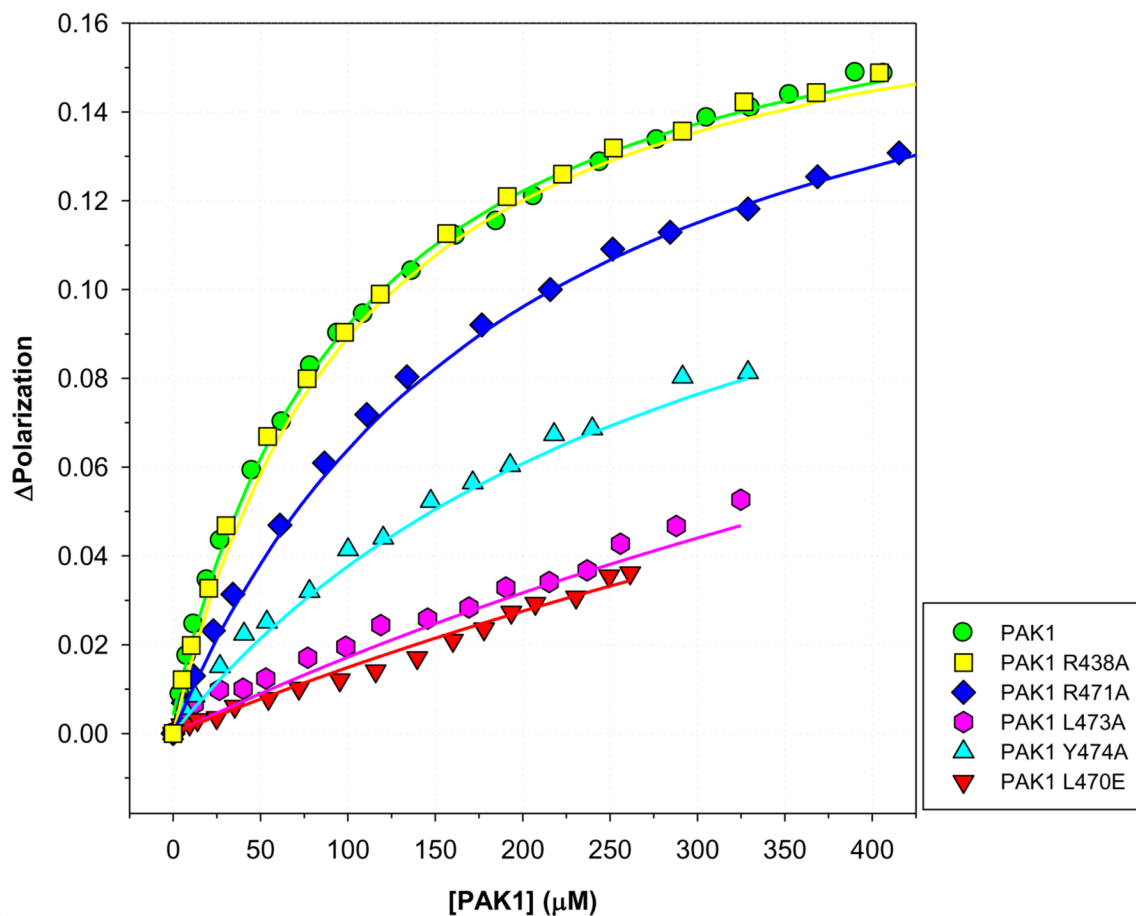
(a)



(b)



(c)

**Figure 4.**

(a) Spider Roll – PAK1 model showing the interface residues. On the PAK1 side (yellow) L470, R471, L473 and Y474 are the main residues involved in interaction. On the Spider Roll side the main contributions come from Y52, A55, G56, I58 and F59. G61 is involved in hydrogen bonding with R438. (b) Effects of Spider Roll mutations on Spider Roll – PAK1 interaction. Bodipy labeled mutants of Spider Roll were titrated with PAK1. None of the single mutation had a dramatic effect confirming that the interface on Spider Roll could be 'plastic' and able to resist mutation by changing orientation. HYP mutant (scaffold with A15C, G26E, L37E and L38N) did not show any binding. (c) Effects of PAK1 mutations on Spider Roll – PAK1 interaction. Bodipy labeled Spider Roll was titrated with mutants of PAK1. L470E and L473A mutations showed maximum disruption (more than 10 fold weakening in binding). R471A and Y474A had subtle effects while R438A did not show any effect on the interaction.

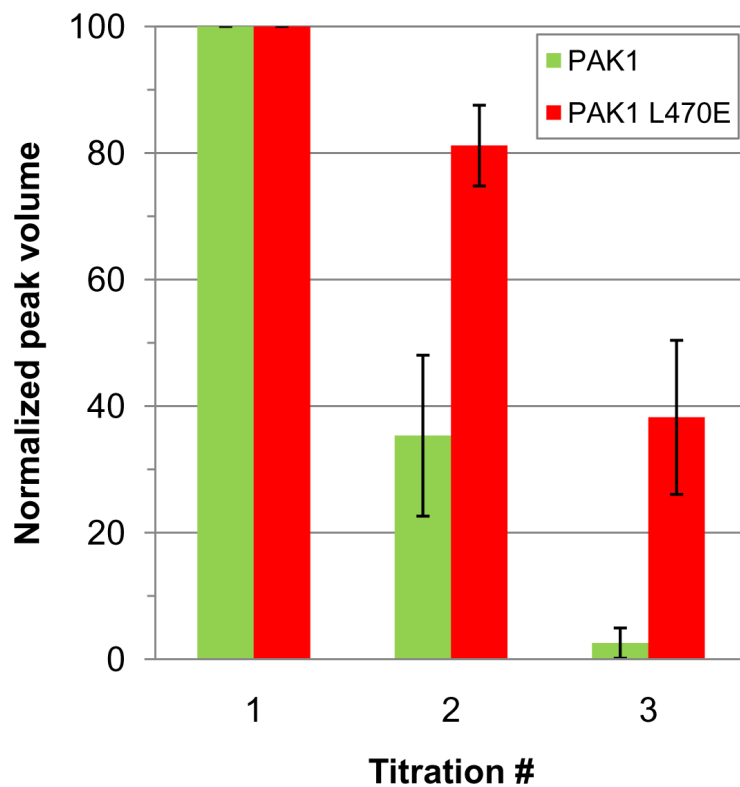


Figure 5. Normalized Spider Roll I58A HSQC peak volumes as a function of PAK1 concentration. Titration 1: 231 μM Spider Roll I58A with no PAK1; Titration 2: 210 μM Spider Roll I58A with 120 μM PAK1 (green bar) and 213 μM Spider Roll I58A with 124 μM PAK1 mutant (red bar); Titration 3: 148 μM Spider Roll I58A with 487 μM PAK1 (green bar) and 150 μM Spider Roll I58A with 490 μM PAK1 mutant (red bar). Positive and negative error bars are one standard deviation.

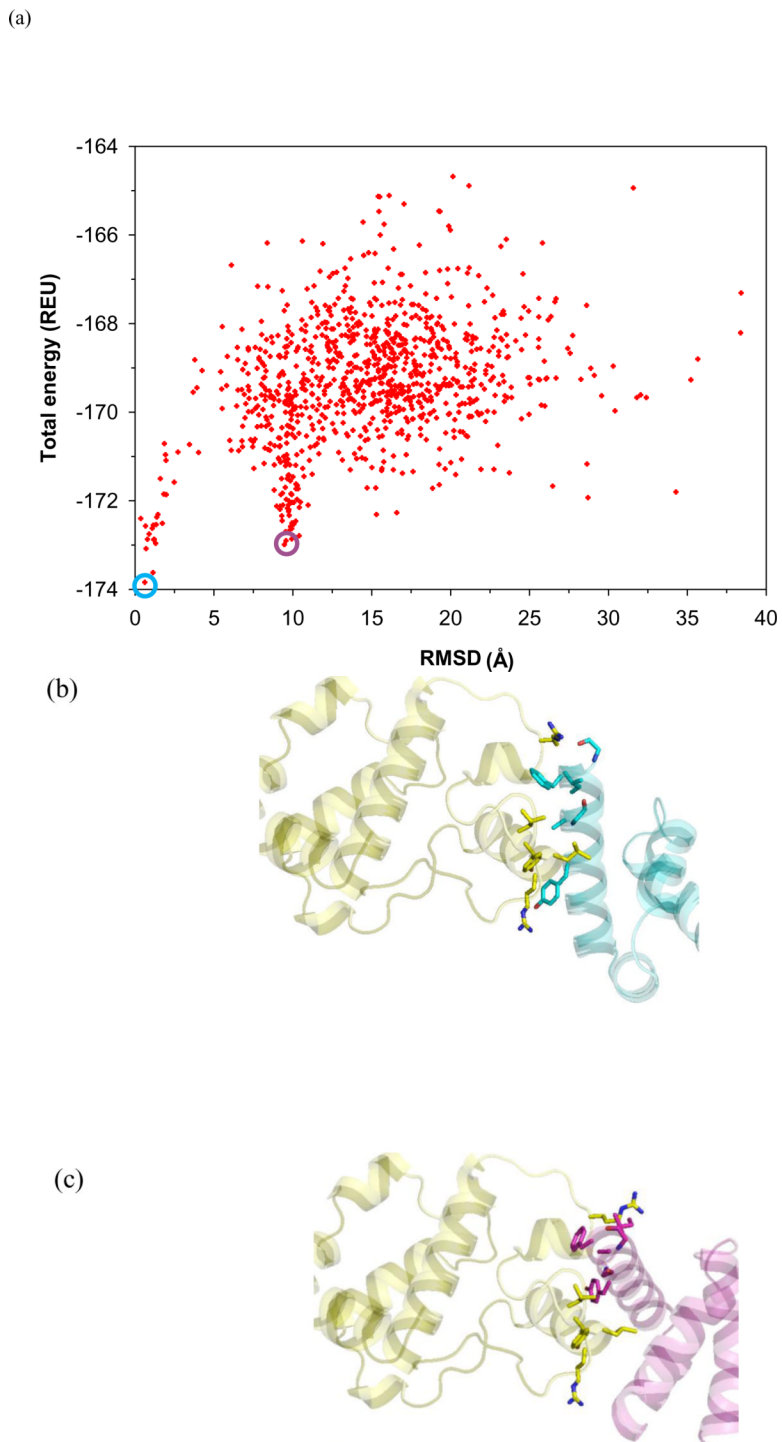


Figure 6.

Alternative docking orientations for Spider Roll (a) Each point represents a different model of the Spider Roll / PAK1 complex and derive from independent docking trajectories. Two alternative low energy states are observed, as shown by the two stems in the funnel. (b) Orientation of interface residues in the low RMSD confirmation (PAK1 in yellow and Spider Roll in cyan). (c) Orientation of the same residues in the high RMSD state (PAK1 in yellow and Spider Roll in magenta).

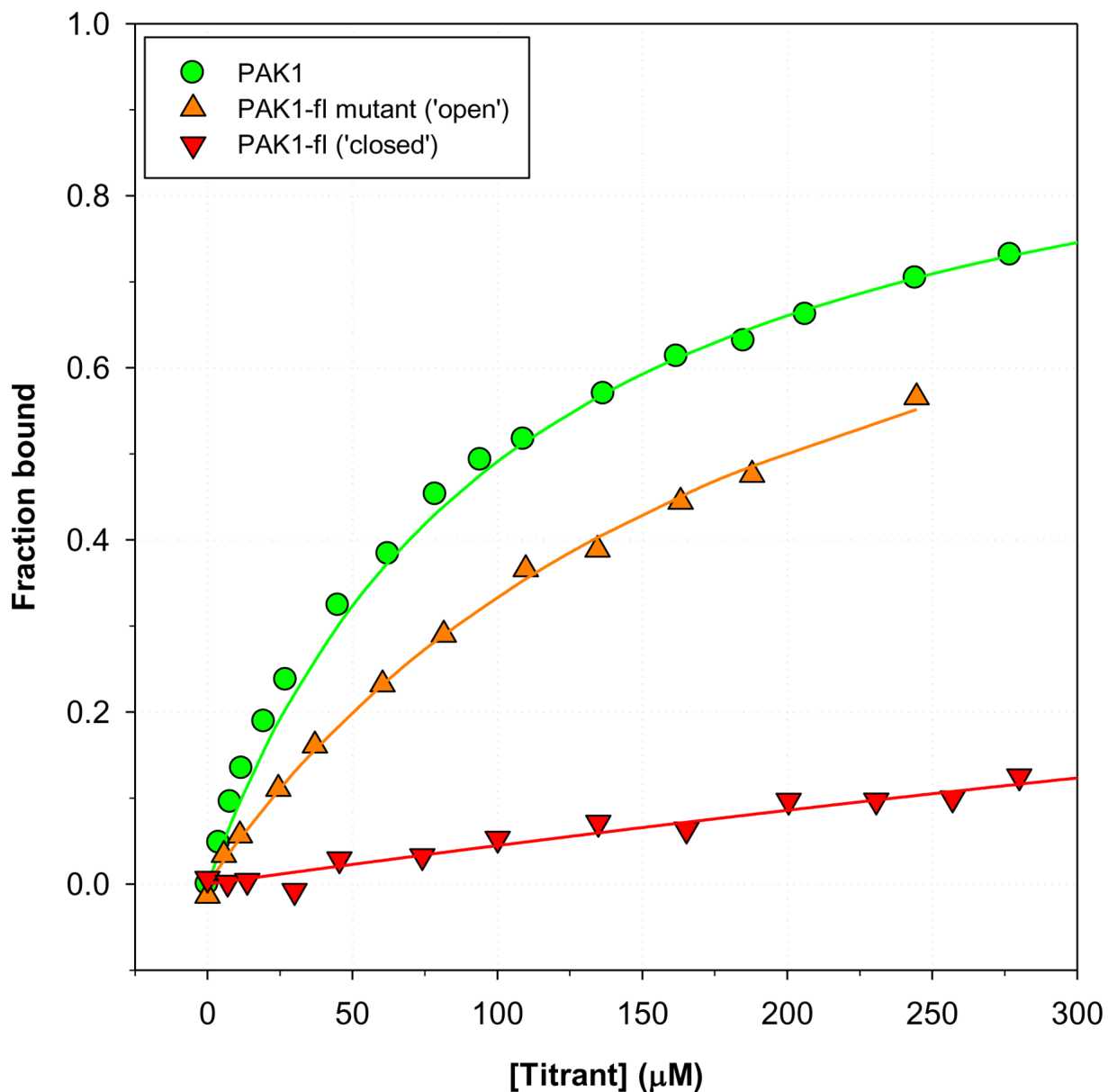


Figure 7. Spider Roll distinguishes between 'closed' and 'open' full-length PAK1 (PAK1-fl). Spider Roll binding titrations with PAK1 V127E/S144E mutant (PAK1-fl mutant, model for full length 'open' form) and PAK1-fl (inactive 'closed' form).

Table 1

Benchmark scores from native PDBs and the cutoffs used as Filter II during simulations

	Buried surface area (Å ²)	Polar surface area (Å ²)	Binding energy density (REU/Å ²)	Unsatisfied polar groups	SASApack score
Native complexes	2060 ± 675	637 ± 249	-0.013 ± 0.003	4.26 ± 2.44	-1.39 ± 1.29
Filter II	-	-	-0.01	4	-

Table 2

Computational scores of the selected design models and comparison with a native AIDkinase interaction in PAK1 full length.

Design	# of mutations on scaffold ^a	# of Dockdesign trajectories	Buried surface area (Å ²)	Polar buried surface area (Å ²)	Binding energy density (REU/Å ²)	# of unsatisfied polar groups	SASA pack score
0233	9	3000	1507	286	-0.0118	4	-0.5
Spider Roll	12	10000	966	174	-0.0152	0	-0.65
1212	10	20000	1500	383	-0.0134	0	2.31
3533	10		1681	323	-0.0129	3	2.79
s032^a	15	320	983	124	-0.0167	0	-3.26
s037^a	21	370	1207	288	-0.0173	0	0.72
Trunc-AID	-	-	1619	357	-0.0157	2	-0.027

^aThe models were generated using discrete molecular dynamics on Spider Roll for different backbone conformations and the binding interface was designed using DDMI.

Table 3

Biophysical characterization of the selected designs

Design	Gel Filtration	Far-UV CD Scan	Thermal denaturation	FP Assay: K_d (μ M)
0233	Monomer	Folded	Cooperative	Poor fit
Spider Roll	Monomer	Folded	Cooperative	100
1212	Monomer	N/D		330
3533	Aggregate			N/D
s032	Aggregate			N/D
s037	Monomer	Folded	Cooperative	160

Table 4

Buried solvent accessible surface area (SASA) of the interface residues in Spider Roll kinase complex in two alternate docked positions (Figure 6). NACCESS program²⁶ was used to calculate the absolute SASA of each interface residue in the complex and then subtracted from SASA of each residue in the independent chains to give buried SASA.

Interface residue	Low RMSD state (Å ²)	High RMSD state (Å ²)
R438	81	110
L470	104	135
R471	43	49
L473	50	61
Y474	98	67
Y52	102	125
A55	52	35
G56	22	16
I58	53	27
F59	135	113