

Nonnatural protein–protein interaction-pair design by key residues grafting

Sen Liu^{*†}, Shiyong Liu^{*†}, Xiaolei Zhu^{*}, Huanhuan Liang^{*}, Aoneng Cao^{*}, Zhijie Chang[‡], and Luhua Lai^{*†§}

^{*}Beijing National Laboratory for Molecular Sciences, State Key Laboratory for Structural Chemistry of Unstable and Stable Species, College of Chemistry and Molecular Engineering, Peking University, Beijing 100871, China; [†]Center for Theoretical Biology, Peking University, Beijing 100871, China; and [‡]Department of Biological Sciences and Biotechnology, School of Medicine, Tsinghua University, Beijing 100084, China

Edited by Stephen L. Mayo, California Institute of Technology, Pasadena, CA, and approved January 22, 2007 (received for review July 21, 2006)

Protein–protein interface design is one of the most exciting fields in protein science; however, designing nonnatural protein–protein interaction pairs remains difficult. In this article we report a *de novo* design of a nonnatural protein–protein interaction pair by scanning the Protein Data Bank for suitable scaffold proteins that can be used for grafting key interaction residues and can form stable complexes with the target protein after additional mutations. Using our design algorithm, an unrelated protein, rat PLC δ_1 -PH (pleckstrin homology domain of phospholipase C- δ_1), was successfully designed to bind the human erythropoietin receptor (EPOR) after grafting the key interaction residues of human erythropoietin binding to EPOR. The designed mutants of rat PLC δ_1 -PH were expressed and purified to test their binding affinities with EPOR. A designed triple mutation of PLC δ_1 -PH (ERPH1) was found to bind EPOR with high affinity (K_D of 24 nM and an IC_{50} of 5.7 μ M) both *in vitro* and in a cell-based assay, respectively, although the WT PLC δ_1 -PH did not show any detectable binding under the assay conditions. The *in vitro* binding affinities of the PLC δ_1 -PH mutants correlate qualitatively to the computational binding affinities, validating the design and the protein–protein interaction model. The successful practice of finding a proper protein scaffold and making it bind with EPOR demonstrates a prospective application in protein engineering targeting protein–protein interfaces.

de novo design of protein–protein interaction pair | erythropoietin | functional site grafting | key residue at interface

Protein–protein interaction is critical in many biological processes ranging from cell differentiation to apoptosis; however, little is known about the general principles governing the specificity and binding affinity of protein–protein interactions. Recent developments in structural bioinformatics have contributed greatly to our understanding of protein–protein interactions (1–6). Several groups have addressed the feasibility of computational redesign of protein–protein interactions by using native or homologous protein–protein interfaces (7–10). Because different protein backbones can perform similar functions (11), in the current study we explored the feasibility of designing nonnatural protein–protein interaction pairs using known protein scaffolds.

To reconstruct the function of a protein, one of the common strategies is grafting, i.e., transferring the functional epitopes from one protein to another. The critical step for protein grafting is to find suitable sites for functional epitope transfer. To this purpose, several computational methods have been developed including a geometric hashing paradigm (12), FITSITE (13), GRAFTER (14), and DEZYMER (15, 16). We have developed a strategy for protein–protein interface redesign by grafting discontinuous interaction epitopes to nonhomologous proteins (17–19). The erythropoietin (EPO)–EPO receptor (EPOR) system was used as an example for nonnatural protein–protein interaction-pair design with the purpose of finding scaffold proteins that can host the functional epitopes of EPO and thus can bind with EPOR after proper mutations.

Human EPO (hEPO), a potent regulator of erythropoiesis, is a glycoprotein hormone primarily produced by the kidney. Biochemical studies showed that EPO contains two binding sites for the EPOR, a high-affinity binding site with a K_D of \approx 1 nM and a low-affinity site of 1 μ M (20). Mutational analysis suggested that residues 11–15, 44–51, 100–108, and 147–151 on EPO are important for erythropoiesis activity (21). The crystal structure of the EPO–EPOR complex has been determined at a resolution of 1.9 \AA [Protein Data Bank (PDB) ID code 1EER] (22). This structure demonstrates that Phe-48, Asn-147, Arg-143, and Arg-150 of EPO are among the key residues for binding to EPOR (20–22).

We have taken the primary EPO–EPOR binding site as our target of protein–protein interface redesign. After scanning the PDB, rat pleckstrin homology (PH) domain of phospholipase C- δ_1 (PLC δ_1 -PH) was found to accommodate well the key interaction residues in EPO and at the same time can form a good interface with EPOR. The designed mutants of rat PLC δ_1 -PH showed high binding affinity with EPOR *in vitro* and significant biological activity in a cell-based assay. The current study presents a successful example for *de novo* nonnatural protein–protein interaction-pair design.

Results

Computational Design Results. Based on the experimental mutation data (20–22) and the crystal structure of the EPO–EPOR complex (22), three key residues in EPO, Phe-48, Asn-147, and Arg-150 at the high-affinity binding site of EPO–EPOR complex, were selected and used in scaffold protein searching. By using the crystal structure of the EPO–EPOR complex (PDB ID code 1EER) as the template structure, a search of PDB yielded 1,756 potential scaffold proteins. After filtering for an acceptable rms difference ($<1.0 \text{\AA}$), 813 solutions that correspond to 534 scaffold proteins were obtained. These solutions were further screened for good shape complementarity, appropriate atomic packing density, and large enough buried accessible surface area (ASA) ($>1,200 \text{\AA}^2$). The top 15 solutions that meet all of the above criteria and high shape complementarity scores, and only one suggested additional mutation beyond the three key residues discussed above, are listed in supporting information (SI) Table 2.

We applied the following rules to select appropriate proteins for experimental studies. The proteins should be stable for mutational

Author contributions: Sen Liu and Shiyong Liu contributed equally to this work; Z.C. and L.L. designed research; Sen Liu, Shiyong Liu, X.Z., and H.L. performed research; Sen Liu, Shiyong Liu, A.C., Z.C., and L.L. analyzed data; and Sen Liu, Shiyong Liu, and L.L. wrote the paper. The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Abbreviations: EPO, erythropoietin; hEPO, human EPO; EPOR, EPO receptor; PLC δ_1 , phospholipase C- δ_1 ; PH, pleckstrin homology; PLC δ_1 -PH, PH domain of PLC δ_1 ; hEPOsR, extracellular domain of human EPOR; SPR, surface plasmon resonance; EMP, EPO-mimetic peptide; PDB, Protein Data Bank; ASA, accessible surface area.

[§]To whom correspondence should be addressed. E-mail: lhlai@pku.edu.cn.

This article contains supporting information online at www.pnas.org/cgi/content/full/0606198104/DC1.

© 2007 by The National Academy of Sciences of the USA

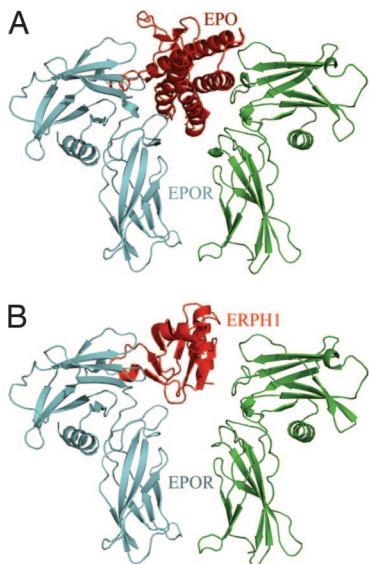


Fig. 1. Comparison of the designed ERPH1-EPOR complex model with the EPO-EPOR complex structure. (A) The structure of EPO-EPOR complex structure (PDB ID code 1EER). Red, EPO; green and cyan, EPOR dimer. (B) The designed ERPH1-EPOR complex model. Red, EPO; green and cyan, EPOR dimer. Images were created by using PyMOL (48).

studies and suitable for expression in *Escherichia coli*, should not contain metal ions to avoid probable cell toxicity, should have no more than four cysteine residues or two pairs of disulfide bonds, and should be of human origin or highly homologous to ensure the safety of probable clinic applications. After eliminating the proteins that failed to meet these criteria, PLC δ_1 -PH (PDB ID code 1MAI) was selected for further experimental studies.

Analysis of the ERPH1-EPOR Interface. ERPH1 is one of the designed mutants of PLC δ_1 -PH with the three key residue mutations E63F, D47N, and K49R. The ERPH1-EPOR complex has a buried surface area of 1,825.5 Å² and a buried hydrophobic surface area of 976.7 Å² at the interface. These values are comparable with that of the EPO-EPOR binding site (with a total buried surface area of 1,946.8 Å² and a buried hydrophobic surface area of 1,052.2 Å²). Fig. 1 A and B compares the overall complex structure of EPO-EPOR and the model of ERPH-EPOR. ERPH1 interacts mainly with the strong binding site in EPOR dimer. It is interesting that EPO and PLC δ_1 -PH have very different protein structures: EPO molecule forms a four-helix bundle, whereas PLC δ_1 -PH adopts a mixed helix and sheet structure. The designed interface contains strong H-bond interactions and good hydrophobic packing similar to those in the primary binding site of EPO-EPOR complex (Fig. 2). Just like the interactions at the EPO-EPOR interface, the newly mutated Phe-63 in ERPH1 forms a hydrophobic cluster with Phe-93 and Phe-205 in EPOR, the side chain amide of Asn-47 in ERPH1 forms H-bonds with the side chain of His-114 in EPOR, and the side chain of Arg-49 in ERPH1 forms H-bonds with the side chain of Glu-117 and the backbone carbonyl oxygen of Pro-203 in EPOR.

To calculate the electrostatic potentials of ERPH1 and EPOR, the online PDB2PQR server (<http://agave.wustl.edu/pdb2pqr/server.html>) was used to prepare the input file for APBS (23). A strong negative electrostatic potential was observed at the EPOR binding site, whereas the ERPH1 binding site showed a positive electrostatic potential similar to that of EPOR (SI Fig. 6).

To assess the potential binding abilities of the single, double, and triple mutants of PLC δ_1 -PH with EPOR, we calculated the binding free energies for the mutants using a statistical potential (PMF-

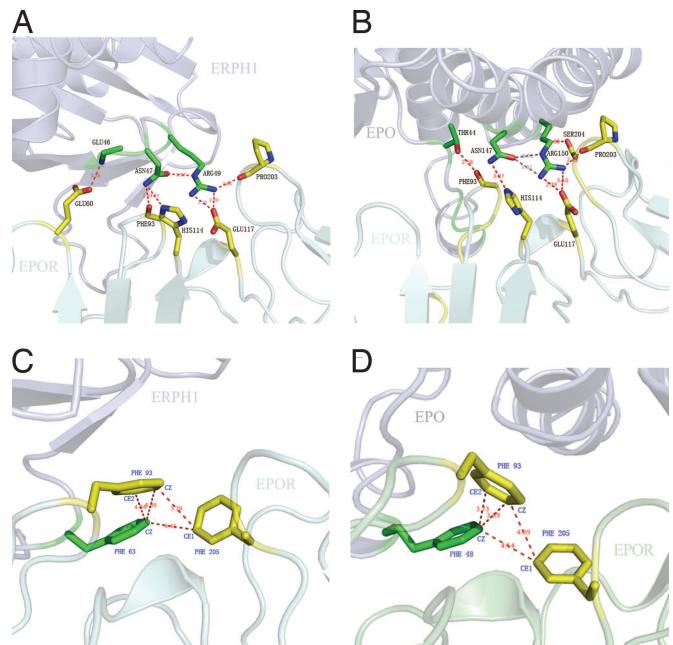


Fig. 2. Key interactions at the designed ERPH1-EPOR interface compared with those at the EPO-EPOR interface. (A) Hydrogen-bond interactions at the ERPH1-EPOR interface. (B) Hydrogen-bond interactions at the EPO-EPOR interface. (C) The hydrophobic cluster at the designed ERPH1-EPOR interface. (D) The hydrophobic cluster at the EPO-EPOR interface. ERPH1 is shown in gray ribbons, and EPOR is shown in pale green ribbons. The residues at the interfaces are shown in colored sticks.

Score) (24). The ranking of the calculated binding free energies was qualitatively consistent with that of the experimentally determined binding constants (Table 1).

In addition to the three key mutations in ERPH1, our calculation also suggested that Glu-46 be mutated to a smaller residue. The side chain of Glu-46 appeared to clash with Glu-60 of EPOR. A fourth mutation of E46A on ERPH1 was introduced to give ERPH2.

Secondary Structures of the PLC δ_1 -PH Mutants. PLC δ_1 -PH is a stable protein with well defined secondary and tertiary structures (25). Far-UVC circular dichroic spectra indicated that the eight mutants of PLC δ_1 -PH (ERPH1-8) maintained the original secondary structure of the WT protein (SI Fig. 7). This result indicates that mutagenesis did not change the rigid structural fold of PLC δ_1 -PH. The mutant proteins were expressed as N-terminal histidine fusion proteins in *E. coli*. Previous experiments have indicated that the His

Table 1. Equilibrium constants (K_D) of hEPO and ERPHs binding to hEPOsR as determined by SPR

Name	Mutation(s)	K_D , nM	ΔG , * kcal/mol
hEPO	—	0.13 ± 0.05	-21.08
PLC δ_1 -PH	—	ND	-11.63
ERPH1 ^a	E63F, D47N, K49R	24 ± 3	-13.42
ERPH2	E63F, D47N, K49R, E46A	26 ± 3	-13.26
ERPH3	E63F	240 ± 60	-12.96
ERPH4	E63F, D47N	29 ± 2	-13.54
ERPH5	E63F, K49R	69 ± 1	-12.91
ERPH6	D47N, K49R	500 ± 270	-12.10
ERPH7	D47N	ND	-12.21
ERPH8	K49R	ND	-11.59

ND, could not be determined under the experimental conditions because of very weak or undetectable binding.

*The binding free energies calculated by using PMFScore (24).

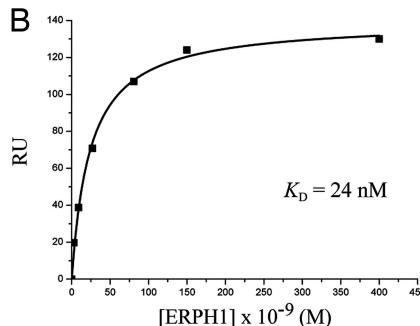
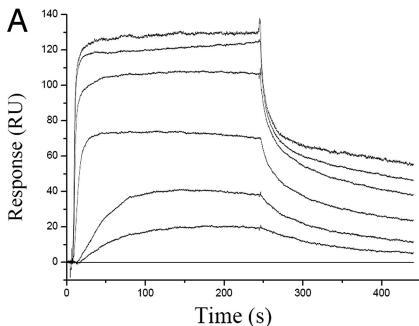


Fig. 3. K_D determination of ERPH1 binding to hEPOsR from *in vitro* SPR study. (A) Sensograms of the binding of increasing concentrations of ERPH1 to hEPOsR. The concentrations of ERPH1 used were 0, 3, 9, 27, 81, 150, and 400 nM from the bottom up. (B) K_D determination of the binding of ERPH1 to hEPOsR. Data points represent the equilibrium average response shown in A. The solid line represents the theoretical curve that was globally calculated from the steady-state fit model provided by BIAsimulation 4.1 software (Biacore).

tag did not affect the structure or function of WT PLC δ_1 -PH (26); thus, the tag was not removed in the following studies.

In Vitro Binding Affinity Measurement of ERPHs with the Extracellular Domain of Human EPOR (hEPOsR). The ERPH mutants were found to bind with hEPOsR at steady-state processes; therefore, the K_D values of ERPH binding to hEPOsR were calculated from the binding curves by using the steady-state model. The original sensogram of ERPH1 binding to hEPOsR (Fig. 3A) and the fitted curve generated by the BIAsimulation 4.1 software (Fig. 3B) yielded a fitted R_{max} of 140 RU, which agrees well with the theoretical value. The K_D values of hEPO and all mutants (ERPH1–8) binding to hEPOsR are listed in Table 1 (sensograms shown in SI Fig. 8). The K_D values of hEPO binding with hEPOsR measured in our surface plasmon resonance (SPR) experiment was 0.13 nM, which is consistent with published data ($K_D = 0.25$ nM) (27). The strongest binding mutant, ERPH1, has a binding affinity K_D of 24 nM. This value is ≈ 180 times less potent than native EPO. For the three single mutations, the binding of the E63F mutant (240 nM) was the strongest, and the binding of the other two mutants was not detectable under our experimental conditions. Comparison of the binding strengths of the double mutants showed that the mutants containing the E63F mutation (E63F-D47N, 29 nM; E63F-K49R, 69 nM) were stronger than the mutant without this change (D47N-K49R, 500 nM). Additional mutation of E46 to alanine did not contribute to binding.

In Vivo ERPH1 Activity. According to the *in vitro* binding studies, the strongest binding mutant ERPH1 was further tested for its biolog-

ical activity in the EPO–EOPR signal cascade. We used the JAK2/STAT5 reporter system because EPO, by binding to EPOR on the surface of specific cells, activates STAT5 transcription activity (28). After transfection of pcDNA3.1_SP/ERPH1 (0 ng, 300 ng, and 600 ng) into 293T cells along with the luciferase reporter system, the ERPH1 protein with a secretory signal peptide was expressed in the presence of extrinsic hEPO protein for competition. Transcription activities as measured by luciferase activity and expression levels of ERPH1 were determined. As shown in Fig. 4, the luciferase activity decreased with the increased expression of ERPH1. This decrease may come from the competitive binding of ERPH1 to EPOR on the cell surface.

The inhibitory effect of ERPH1 on EPO activity was further studied by using purified ERPH1 protein (0–48.3 μ M) (Fig. 5). Luciferase activity decreased sharply as the concentration of ERPH1 was increased from 0.05 to 48.3 μ M, whereas the luciferase activity did not change obviously when adding 48.3 μ M WT PLC δ_1 -PH (Fig. 5A). Thus, the triple mutant ERPH1 can bind to EPOR and significantly inhibit EPO activity. From the dose-dependent com-

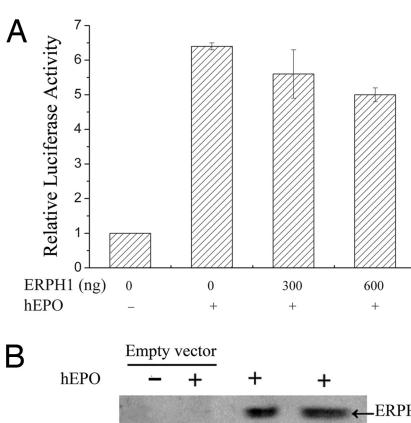


Fig. 4. Effects of ERPH1 on the EPO–EPOR signal pathway. (A) Dose-dependent inhibitory effects of ERPH1 on EPO–EPOR signal pathway. Different amounts of ERPH1 cDNA with a secretory signal peptide sequence were transfected into 293T cells. A total of 1.5 units (1.5 nM) of hEPO (Roche) was added as the competitor. Data were normalized by transfection with a blank pcDNA3.1_SP/myc-His A vector and are expressed as the mean \pm SD ($n = 3$). (B) The expression levels of ERPH1 protein detected by Western blotting.

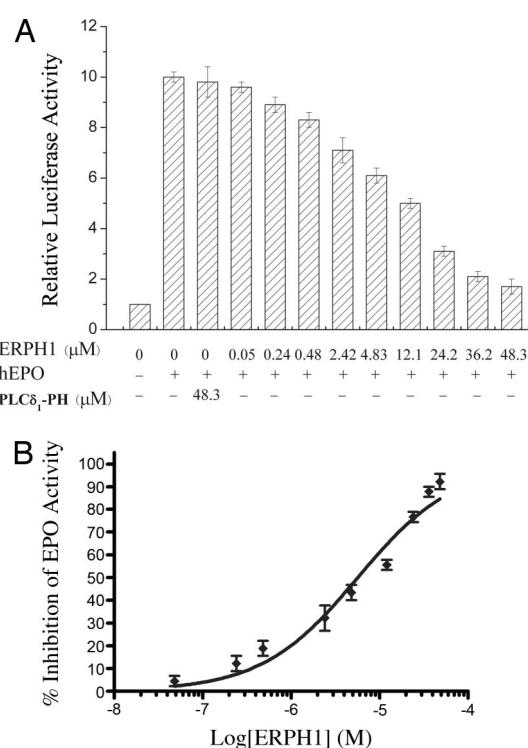


Fig. 5. Dose-dependent inhibitory effects of purified recombinant ERPH1 on EPO activity. (A) The relative luciferase activity when different concentrations of ERPH1 were added. (B) IC₅₀ plot of ERPH1. The IC₅₀ was determined to be 5.7 ± 1.2 μ M ($R^2 = 0.9695$).

petitive assay with purified ERPH1, the IC_{50} of ERPH1 on EPO activity was calculated to be $5.7 \pm 1.2 \mu\text{M}$ (Fig. 5B).

Discussion

In the past few years, rapid progress has been made in protein engineering from functional reconstruction to *de novo* design of monomeric proteins (29, 30); however, the rational design of protein–protein interacting pairs is still in its infancy. Recently, several groups suggested that it is feasible to computationally redesign protein–protein interfaces by using native or homologous protein pairs (7–10). For example, Reina *et al.* (7) computationally reengineered a PDZ domain to bind novel peptide target sequences, and the designed protein showed affinity of two orders of magnitude higher than that of the WT interacting partners (7). Shifman and Mayo (8, 31) reported specificity design based on a calmodulin–ligand complex. Havranek and Harbury (9) explicitly took negative design into account in protein–protein interface design. Kortemme *et al.* (10) redesigned a DNase–immunity protein interface to specifically bind a new pair of interacting partners stronger than their WT counterparts.

Because the topologies of natural proteins are limited (32) and similar functions can be performed by different proteins (11), the task of designing a nonnatural protein–protein interaction pair by using known protein structures is reasonable but extremely challenging. We have grafted noncontinuous epitopes of EPO to a small nonhomologous protein and made it bind with the EPOR. ERPH1 was successfully designed from the rigidly folded protein PLC δ_1 -PH to bind to its nonnatural partner EPOR with a high affinity *in vitro* as well as *in vivo*.

The PH domain is one of the most abundant domains found in the human genome (33). The PH domain was originally identified in 1993 as a stretch of 100–120 aa that appears twice in the platelet protein pleckstrin (34). Despite their low amino acid sequence identity, all PH domains with known structures have a core seven-stranded β -sandwich structure with one corner capped off by a C-terminal α -helix and another by three inter-strand loops (25). This “PH-domain superfold” is thought to be a particularly stable structure scaffold adapted to diverse functions (35). Of these PH domains, PLC δ_1 -PH is best characterized. The main function of PLC δ_1 -PH is to target PLC δ_1 to membranes via binding to membrane phosphoinositide (36).

The conserved fold of PLC δ_1 -PH was ideal for the rigid docking algorithm used to graft the three key residues from EPO, which tolerated the mutations and maintained its overall structure. Using the SPR assay, we demonstrated that WT PLC δ_1 -PH did not bind to hEPOsR at detectable levels; however, the designed mutants bound to hEPOsR. The binding affinities of the single and double mutants indicated that the three mutated residues contribute differently to the binding (E63F → D47N → K49R). This result is consistent with our analysis of the ERPH1–EPOR interface. The introduced aromatic ring of Phe at position 63 on ERPH1 forms a tight packing hydrophobic cluster with two critical residues, Phe-93 and Phe-205 of EPOR (37). Furthermore, the side chain amide of Asn-47 forms H-bonds with His-114 and Phe-93 in EPOR, whereas the two guanidine nitrogen atoms in the side chain of Arg-49 form H-bonds with Glu-117 and Pro-203 in EPOR. Mutation of Glu-63 to Phe significantly altered residue hydrophobicity, whereas mutation of Asp-47 to Asn and Lys-49 to Arg did not remarkably change residue property. Therefore, the interpretation of the contribution of these residues to the binding of EPOR is fairly straightforward. The triple mutation E63F, D47N, and K49R generated the highest affinitive mutant. An additional mutation, E46A, did not increase the binding affinity of the triple mutant. The Glu-46 may be accommodated well by the plasticity of the two proteins. The calculated binding-free energies were qualitatively consistent with the dissociation constants from the SPR experiments.

Recently, a series of synthetic peptides [EPO-mimetic peptides (EMPs)] were discovered to mimic EPO activity (38–40). A small peptide screened from random phage display peptide library inhibits the binding of ^{125}I -labeled EPO to the extracellular binding domain of EPOR (hEPOsR, 1–225 residues of EPOR) with an IC_{50} of $10 \mu\text{M}$. EMPS with higher binding affinities to hEPOsR were then designed based on the sequence of this peptide. The best candidate EMP (EMP1) had rather low IC_{50} values of $0.07 \mu\text{M}$ and $5 \mu\text{M}$, on TF-1 cells and hEPOsR beads, respectively. Our designed protein ERPH1 binds to hEPOsR with a dissociation constant of 24 nM, which is comparable to those selected from a large library, although the binding affinities of ERPH1 cannot be compared directly with EMP peptides as these values were obtained via different assays. As ERPH1 was designed from WT PLC δ_1 -PH with three residue mutations, its strong binding validates the design procedure. This high-affinity binding, however, does not necessarily indicate high EPO-mimetic activity because both proximity and orientation of the receptor molecules are important for bioactivity. For instance, one of the EMPS, EMP33, was an antagonist of EPO despite the fact that this protein bound EBP with reasonably high affinity (IC_{50} values of $15 \mu\text{M}$ and $50 \mu\text{M}$ on TF-1 cells and EBP beads, respectively) (41). In our cell-based assay, ERPH1 inhibited the activity of EPO, indicating that ERPH1 can bind EPOR in a cellular context. Because no EPO mimic activity was found, our designed ERPH1 behaves as an antagonist similar to EMP33, which binds to EPOR with high affinity but fails to produce active EPOR.

In the current study, we have successfully designed a nonnatural ERPH1–EPOR interaction pair by scanning the PDB for appropriate scaffold proteins and proper mutations. Why only a few mutations on PLC δ_1 -PH made it bind to its nonnatural partner EPOR is an interesting question. From our design procedure, we felt that key interaction residues contribute significantly. This is in accordance with hot-spots at protein–protein interface contributing significantly to the stability of the protein–protein complex (42). However, only the three key residues mutations on possible scaffold proteins will not guarantee strong protein–protein interactions. As a negative control, we grafted the three key residues of EPO onto the SH2 (Src homology) domain of the Nck2 adapter protein (PDB ID code 1Z3K), without other interface optimization except the key residues alignment. The SH2 domain has three appropriate positions for key residues grafting, but with a low complementarity score ($-3,906$) and bad steric collisions (SI Table 2). This implies that the interface of this SH2 domain is not appropriate for simple mutations to make it bind with EPOR. As confirmed by the SPR experiments, this SH2 domain mutant cannot bind to EPOR after three key residue grafting (SI Fig. 9).

In addition to the appropriate sites for grafting the three key residues, we have used other criteria to select out good scaffold proteins from the PDB, which include shape complementarity, buried ASA, and atomic packing density at the designed interface. We believe that all these criteria contribute to the success of protein–protein interaction-pair design. These factors were shown to be important for distinguishing biological and nonbiological protein–protein interfaces (43). As shown by the electrostatic potentials of ERPH1 and EPOR, electrostatic interactions may be another factor that enhances the binding between PLC δ_1 -PH mutants and EPOR, which were believed to be crucial for protein–protein recognition (44). As to why a single mutation E63F of PLC δ_1 -PH made it bind to EPOR with a K_D of 240 nM, we believe that the WT protein already possesses good features for binding with EPOR, although its binding to EPOR was too weak to be detectable under our assay conditions. When mutating Glu-63 to Phe, the newly introduced Phe forms a strong hydrophobic cluster with other two Phes in EPOR, thus greatly increasing its binding ability with EPOR. This shows an advantage of our design procedure of using existing protein structures as a pool of possible scaffold proteins. When the number of known protein structures is bigger enough, we can expect to find a possible protein interaction

partner for any protein of interest. Thus, a similar algorithm might be useful in protein function prediction to search for possible protein binding partners based on 3D structures. The method reported here should be widely applicable to protein engineering targeting protein–protein interfaces.

Materials and Methods

Computational Grafting of the EPO Functional Epitopes. The residues Phe-48, Asn-147, and Arg-150 of EPO (20–22) were selected as the key residues for grafting. The EPO–EPOR complex structure (PDB ID code 1EER) was used as the template structure. The detailed algorithm of the grafting strategy has been described elsewhere (17–19). The following is a brief description.

We began by choosing the proteins with 100–200 residues in the PDB (version 2001-10-1) as candidates for EPO functional epitope grafting. These candidate proteins were used as scaffold proteins. Subsequently, if the combination of any three residues on a scaffold protein satisfies the geometric relationships of the C_{α} – C_{β} vectors of the three key residues in EPO, these three residues were defined as the key residues of the scaffold protein and were superimposed onto the three key residues of EPO by a set reduction algorithm. An interface was defined as a set of atoms on a protomer that loses at least 0.1 \AA^2 of ASA per atom upon binding to a partner and has an ASA of $<15 \text{ \AA}^2$. The solvent ASA of an atom or a residue was calculated by the program NACCESS (45) with a probe radius of 1.4 \AA . The interface area was measured by comparing the solvent ASA of the complex to that of its components defined by Eq. 1:

$$B = A_A + A_B - A_{AB}, \quad [1]$$

where A_{AB} is the solvent ASA of the complex and A_A and A_B are the solvent ASA of the dissociated components. B represents the buried surface area of the two proteins in contact. If the buried interface area of the computed protein–protein complex were $<1,200 \text{ \AA}^2$, the scaffold would be rejected.

Next, we mutated the three key residues of the scaffold protein identified in the first step to the corresponding residues of EPO. The shape complementarity score of the designed protein–protein interface of the scaffold protein with EPOR was evaluated by using grid-based molecular representations (18). The receptor protein was digitized onto a 3D grid with the grid step of 1 \AA . Grid nodes inside of the receptor protein were assigned a value of 1. Those grid nodes outside of the receptor protein were set to zero. The scaffold protein was projected onto a similar grid. The grid nodes occupied by the scaffold protein backbone atoms were assigned a value of -10 , and those occupied by side-chain atoms of scaffold protein were assigned a value of -2 . Two layers of grid nodes on the scaffold protein surface were assigned a value of 1, and others are set to zero. The geometric complementary score was calculated by accumulating the products of the numbers assigned to the two grids at each overlapping grid point.

The positions of the two proteins were then readjusted to ensure proper packing density, which was defined as the mean of atom packing densities at the interface. Most natural protein–protein complexes have a mean interface packing density in the range of 0.70–0.78 (19). Therefore, if the packing density of the designed interface was >0.78 , the scaffold protein was moved from the geometric center of the receptor to that of the scaffold protein in steps of 0.1 \AA until an acceptable value was reached.

The last step for the design process was to calculate the rms differences between the key residues of the scaffold protein and those of EPO. The scaffold protein was rejected if the rms value was $>2 \text{ \AA}$. Additionally, the scaffold protein was also rejected if any angle deviation between the C_{α} – C_{β} vectors of the scaffold protein and EPO was $>60^\circ$.

The above selection criteria include buried ASA at the interface, shape complementarity score, packing density, and rms differences between the key residues. After this step, the designed interfaces

will have reasonable interactions in addition to the key residue interactions. If other residues at the interface were found to clash with residues of the target protein, then further mutations would be done by using a side chain repacking calculation.

For the PLC δ_1 -PH–EPOR system, complex models of PLC δ_1 -PH single, double, triple, and quadruple mutants were built. After mutation of the key residues, conformations of other side chains at the interface were calculated by using the SCAP program (46). The complex models were then subject to energy minimization by using CHARMM (47).

Plasmids and Proteins. pGST4/PLC- δ_1 was a generous gift from Hitoshi Yagisawa (University of Hyogo, Japan). pETPHD was constructed by subcloning the DNA fragment encoding WT PLC δ_1 -PH (residues 1–140) from pGST4/PLC- δ_1 into the NdeI and EcoRI sites of the pET28a (Novagen, Madison, WI) vector. The mutants of WT PLC δ_1 -PH in pET28a were generated by cassette mutagenesis or Stratagene QuikChange (Stratagene, La Jolla, CA) site-directed mutagenesis and were named as shown in Table 1.

pcDNA3.1.SP/myc-His A was constructed from pcDNA3.1/myc-His A (Invitrogen, Carlsbad, CA) by adding a secretory signal peptide sequence (MSALLILALVGAAVA), which was derived from preprotrypsin, between the HindIII and KpnI sites. The stop codon between the restriction sites of BamHI and EcoRI was also removed. The pEF-FLAG-Stat5 vector containing mouse STAT5a (signal transducers and activators of transcription 5a) cDNA and the pMX-EPOR vector containing murine EPOR cDNA were kindly provided by Xin-yuan Fu (Indiana University School of Medicine, Indianapolis, IN). The pUC18-LHRE/TK vector containing lactogenic hormone responsive element (LHRE), which contains a STAT5 DNA-binding site, and a vector containing the thymidine kinase (TK) minimal promoter linked to a luciferase gene was generously provided by K. C. Leung (Garvan Institute of Medical Research, Sydney, Australia). pTK-RL (Promega, Madison, WI) was used to normalize transfection efficiency.

pcDNA3.1.SP/PHD was constructed by inserting the coding sequence of the PCR-amplified WT PLC δ_1 -PH into the BamHI and XbaI sites of pcDNA3.1_SP/myc-His A vector so that the C-terminal His tag and c-Myc epitope was fused in frame with WT PLC δ_1 -PH. The eight mutants in pcDNA3.1_SP/myc-His A were constructed from the mutants in pET28a.

Glycosylated hEPO that was used for the Biacore SPR assays was purchased from Bio-EPO Biotechnology (Beijing, China). The hEPO that was used for the cellular experiments was purchased from Roche. hEPOsR (carrier-free) was purchased from R & D Systems.

Protein Expression in *E. coli* and Purification. The detailed protein expression and purification procedure can be found in SI Methods.

CD. CD spectra were obtained by using a Jobin-Yvon CD6 spectropolarimeter (Longjumeau, France). Far-UV CD spectra from 190 nm to 260 nm (step size 0.5 nm; averaging time 3 sec) were the average of three scans determined at a protein concentration of $10.8 \mu\text{M}$ in 20 mM phosphate buffer (pH 7.0) and 1 mM DTT at 25°C in a 1-mm path-length cuvette. The spectra were corrected by subtracting a buffer blank.

SPR. The SPR experiments were performed with a Biacore 3000 instrument (Biacore, Uppsala, Sweden). Because immobilization deactivates part of the protein, 2,000 RU of hEPOsR in 100 mM sodium acetate (pH 4.0) was immobilized on a CM5 sensor chip via the primary amine groups according to the manufacturer's instructions. To subtract the RU resulting from nonspecific interactions and the bulk refractive index, a control flow cell was activated and blocked in the absence of protein. All binding experiments were performed in HBS-EP buffer (0.01 M Hepes/0.15 M NaCl, pH

7.4/3.7 mM EDTA/0.005% vol/vol Surfactant P20) at 25°C with a flow rate of 10 μ l/min. The chip surface was regenerated with 30 sec of 10 mM glycine-HCl (pH 2.2) followed by 30 sec of 10 mM glycine-NaOH (pH 10). Each analysis cycle consisted of (i) a 1-min stabilization time, (ii) injection of 40 μ l of protein over flow cells 1 and 2, (iii) 120-sec dissociation (buffer flow), (iv) regeneration, and (v) integrated fluidics cartridge wash. The signal was monitored by subtracting flow cell 1 from flow cell 2. Samples and a buffer blank were injected in a random order. The R_{\max} of WT PLC δ_1 -PH or ERPHs binding to the immobilized hEPOsR was computed by using Eq. 2 according to the Biacore 3000 manual:

$$R_{\max} = (M_r A / M_r L) \times R_I \times N, \quad [2]$$

where R_{\max} is the maximum response during injection (under saturating binding conditions). $M_r A$ and $M_r L$ are the molecular weights of the injected protein (analyte) and the immobilized material (ligand, which in this case is hEPOsR), respectively. R_I is the response unit from the active immobilized material, and N is the stoichiometry of the binding, which represents the number of molecules of analyte bound to each molecule of ligand. The amount of active hEPOsR gave a R_{\max} of 240 RU for glycosylated hEPO (SI Fig. 8). Assuming that WT PLC δ_1 -PH and ERPHs bound to the immobilized hEPOsR in a 1:1 ratio by using Eq. 2, the R_{\max} would be \approx 130 RU.

The observed dissociation binding constant, K_D , was calculated by using the average RU under steady-state conditions. Data were fitted globally by using the steady-state model provided by Biacore BIAsimulation 4.1 software.

Cell Culture, Transfection, and Luciferase Reporter Assays. 293T cells were grown in DMEM containing 10% FBS, glutamine, penicillin, and streptomycin (GIBCO). Cells were transfected with 25 ng of pEF-FLAG-Stat5a, 100 ng of pUC18-LHRE/TK, 250 ng of pMX-EPOR, 1.25 ng of pTK-RL, and 600 ng of pcDNA3.1-SP/myc-His A by using Tfx-20 (Promega) according to the manufacturer's recommendations. Twenty-four hours after transfection, cells were serum-starved for 12 h and then either stimulated with protein or serum-starved for an additional 6–8 h.

1. Brannetti B, Via A, Cestra G, Cesareni G, Helmer-Citterich M (2000) *J Mol Biol* 298:313–328.
2. Wollacott AM, Desjarlais JR (2001) *J Mol Biol* 313:317–342.
3. Aloy P, Russell RB (2002) *Proc Natl Acad Sci USA* 99:5896–5901.
4. Brinkworth RI, Breinl RA, Kobe B (2003) *Proc Natl Acad Sci USA* 100:74–79.
5. Li L, Shakhnovich EI, Mirny LA (2003) *Proc Natl Acad Sci USA* 100:4463–4468.
6. Reichmann D, Rahat O, Albeck S, Meged R, Dym O, Schreiber G (2005) *Proc Natl Acad Sci USA* 102:57–62.
7. Reina J, Lacroix E, Hobson SD, Fernandez-Ballester G, Rybin V, Schwab MS, Serrano L, Gonzalez C (2002) *Nat Struct Biol* 9:621–627.
8. Shifman JM, Mayo SL (2002) *J Mol Biol* 323:417–423.
9. Havranek JJ, Harbury PB (2003) *Nat Struct Biol* 10:45–52.
10. Kortemme T, Joachimiak LA, Bullock AN, Schuler AD, Stoddard BL, Baker D (2004) *Nat Struct Mol Biol* 11:371–379.
11. Keskin O, Nussinov R (2005) *Protein Eng Des Sel* 18:11–24.
12. Nussinov R, Wolfson HJ (1991) *Proc Natl Acad Sci USA* 88:10495–10499.
13. Hornischer K, Blocker H (1996) *Protein Eng* 9:931–939.
14. Hearst DP, Cohen FE (1994) *Protein Eng* 7:1411–1421.
15. Hellinga HW, Caradonna JP, Richards FM (1991) *J Mol Biol* 222:787–803.
16. Hellinga HW, Richards FM (1991) *J Mol Biol* 222:763–785.
17. Liang S, Li W, Xiao L, Wang J, Lai L (2000) *J Biomol Struct Dyn* 17:821–828.
18. Liang S, Liu Z, Li W, Ni L, Lai L (2000) *Biopolymers* 54:515–523.
19. Liang SD, Xiao L, Mao FL, Jiang L, Han YZ, Lai LH (2000) *Chin Sci Bull* 45:1707–1712.
20. Philo JS, Aoki KH, Arakawa T, Narhi LO, Wen J (1996) *Biochemistry* 35:1681–1691.
21. Elliott S, Lorenzini T, Chang D, Barzilay J, Delorme E (1997) *Blood* 89:493–502.
22. Syed RS, Reid SW, Li C, Cheetham JC, Aoki KH, Liu B, Zhan H, Osslund TD, Chirino AJ, Zhang J, et al. (1998) *Nature* 395:511–516.
23. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA (2001) *Proc Natl Acad Sci USA* 98:10037–10041.
24. Jiang L, Gao Y, Mao F, Liu Z, Lai L (2002) *Proteins* 46:190–196.
25. Ferguson KM, Lemmon MA, Schlessinger J, Sigler PB (1995) *Cell* 83:1037–1046.
26. Wang T, Pentyala S, Rebecchi MJ, Scarlata S (1999) *Biochemistry* 38:1517–1524.
27. Ahaded A, Winchenne JJ, Cartron JP, Lambin P, Lopez C (1999) *Prep Biochem Biotechnol* 29:163–176.
28. Richmond TD, Chohan M, Barber DL (2005) *Trends Cell Biol* 15:146–155.
29. Harbury PB, Plecs JJ, Tidor B, Alber T, Kim PS (1998) *Science* 282:1462–1467.
30. Kuhlman B, Dantas G, Ireton GC, Varani G, Stoddard BL, Baker D (2003) *Science* 302:1364–1368.
31. Shifman JM, Mayo SL (2003) *Proc Natl Acad Sci USA* 100:13274–13279.
32. Chothia C (1992) *Nature* 357:543–544.
33. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, et al. (2001) *Nature* 409:860–921.
34. Tyers M, Rachubinski RA, Stewart MI, Varrichio AM, Shorr RG, Haslam RJ, Harley CB (1988) *Nature* 333:470–473.
35. Blomberg N, Baraldi E, Nilges M, Saraste M (1999) *Trends Biochem Sci* 24:441–445.
36. Lemmon MA, Ferguson KM, O'Brien R, Sigler PB, Schlessinger J (1995) *Proc Natl Acad Sci USA* 92:10472–10476.
37. Middleton SA, Barbone FP, Johnson DL, Thurmond RL, You Y, McMahon FJ, Jin R, Livnah O, Tullai J, Farrell FX, et al. (1999) *J Biol Chem* 274:14163–14169.
38. Livnah O, Stura EA, Johnson DL, Middleton SA, Mulcahy LS, Wrighton NC, Dower WJ, Jolliffe LK, Wilson IA (1996) *Science* 273:464–471.
39. Wrighton NC, Farrell FX, Chang R, Kashyap AK, Barbone FP, Mulcahy LS, Johnson DL, Barrett RW, Jolliffe LK, Dower WJ (1996) *Science* 273:458–463.
40. Johnson DL, Farrell FX, Barbone FP, McMahon FJ, Tullai J, Hoey K, Livnah O, Wrighton NC, Middleton SA, Loughney DA, et al. (1998) *Biochemistry* 37:3699–3710.
41. Livnah O, Johnson DL, Stura EA, Farrell FX, Barbone FP, You Y, Liu KD, Goldsmith MA, He W, Krause CD, et al. (1998) *Nat Struct Biol* 5:993–1004.
42. Gao Y, Wang RX, Lai LH (2004) *J Mol Mod* 10:44–54.
43. Liu SY, Li QL, Lai LH (2006) *Proteins* 64:68–78.
44. Vizcarra CL, Mayo SL (2005) *Curr Opin Chem Biol* 9:622–626.
45. Hubbard SJ (1993) NACCESS Computer Program (Dept Biochem Mol Biol, Univ College London, London).
46. Xiang Z, Honig B (2001) *J Mol Biol* 311:421–430.
47. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M (1983) *J Comput Chem* 4:187–217.
48. DeLano WL, Ultsch MH, de Vos AM, Wells JA (2000) *Science* 287:1279–1283.