Journal of Structural Biology 185 (2014) 168–177

Contents lists available at SciVerse ScienceDirect

Journal of Structural Biology

journal homepage: www.elsevier.com/locate/yjsbi

Affinity maturation of a computationally designed binding protein affords a functional but disordered polypeptide

Maren Butz, Peter Kast, Donald Hilvert *

Laboratory of Organic Chemistry, ETH Zurich, 8093 Zurich, Switzerland

ARTICLE INFO

Article history:
Available online 26 March 2013

Keywords:
Computational design
Directed evolution
Intrinsically disordered proteins
Molten globule
Protein–protein interactions
Protein–receptor interface
Yeast surface display

ABSTRACT

Computational methods have been recently applied to the design of protein–protein interfaces. Using this approach, a 61 amino acid long protein called Spider Roll was engineered to recognize the kinase domain of the human p21-activated kinase 1 (PAK1) with good specificity but modest affinity (K_D = 100 μM). Here we show that this artificial protein can be optimized by yeast surface display and fluorescence-activated cell sorting. After three rounds of mutagenesis and screening, a diverse set of tighter binding variants was obtained. A representative binder, MSR7, has a >10^2-fold higher affinity for PAK1 when displayed on yeast and a 6 to 11-fold advantage when produced free in solution. In contrast to the starting Spider Roll protein, however, MSR7 unexpectedly exhibits characteristics typical of partially disordered proteins, including lower α-helical content, non-cooperative thermal denaturation, and NMR data showing peak broadening and poor signal dispersion. Although conformational disorder is increasingly recognized as an important property of proteins involved in cellular signaling and regulation, it is poorly modeled by current computational methods. Explicit consideration of structural flexibility may improve future protein designs and provide deeper insight into molecular events at protein–protein interfaces.

1. Introduction

Protein–protein interactions, ubiquitous in living organisms, promote essential biological processes, including cell adhesion, immune recognition, signal transduction, molecular transport and catalysis. Aberrant protein–protein interactions, in contrast, may lead to disease (Ryan and Matthews, 2005). There is consequently considerable interest in understanding the physical principles governing molecular recognition in supramolecular protein assemblies (Jones and Thornton, 1996) and in applying them to the rational design of artificial protein binding pairs for diagnostic and therapeutic applications (Chen and Keating, 2012).

Mutagenesis of protein interfaces is a powerful tool for probing factors that control protein–protein specificity (Bonsor and Sunberg, 2011; Thom et al., 2006). This methodological approach can also be utilized to alter the binding specificity of existing protein interfaces. In this case, rather than testing individual mutants one by one, interesting variants are often selected from large protein libraries and then optimized by iterative cycles of random mutagenesis and screening. For this type of directed evolution experiment, genotype and phenotype are frequently linked by phage (Smith and Petrenko, 1997), cell-surface (Boder and Wittrup, 1997) or ribosome (Mattheokis et al, 1994; Lipovsek and Plückthun, 2004) display. Selection and affinity maturation of many protein-binding antibodies and antibody surrogates, such as lipocalins and ankyrin repeat proteins, attests to the practicality of these methods for tailoring protein–protein interactions (Boersma and Plückthun, 2011; Gebauer and Skerra, 2009; Winter et al., 1994).

As an alternative to experimental evolution, recent advances in computation are being increasingly exploited for the rational (re)-design of protein interfaces (Chen and Keating, 2012). Because of the huge number of combinatorial possibilities, exploring protein sequence space represents a daunting task. To address this challenge, efficient search algorithms have been developed to identify and rank combinations of amino acids able to form protein surfaces of defined shape and chemical composition (Chen and Keating, 2012). While still imperfect, these programs have been successfully used to alter the binding specificities of native protein interfaces (Grigoryan et al., 2009; Kortemme et al., 2004; Potapov et al., 2008) and to repurpose individual protein scaffolds for selective binding of arbitrary targets (Huang et al., 2007; Jha et al., 2010; Liu et al., 2007; Roberts et al., 2012). By introducing complementary changes into the partner protein as well, it has been possible to generate heterodimeric complexes with enhanced affinities (Karanicolas et al., 2011).

* Corresponding author. Tel.: +41 44 632 3176; Fax: +41 44 632 1486.
E-mail addresses: mb879@cam.ac.uk (M. Butz), kast@org.chem.ethz.ch (P. Kast), hilvert@org.chem.ethz.ch (D. Hilvert).

1047-8477/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jsb.2013.03.008

Despite evident progress, the rational design of novel interfaces remains challenging because of the difficulty of quantitatively capturing all the subtle factors that relate sequence to function and cause tradeoffs between affinity and specificity. As a consequence, integrated approaches combining computational design with directed evolution may prove to be the most effective means of generating specific and high-affinity protein–protein complexes. Proteins possessing desired properties are first enriched using a computational step and then subsequently refined by mutagenesis and screening. As illustrated by the design of subnanomolar inhibitors of H1N1 influenza hemagglutinin (Fleishman et al., 2011), this strategy can, when successful, impart exceptionally high affinities.

The current study assesses an evolutionary approach for optimizing a computationally designed receptor for the kinase domain of p21-activated kinase 1 (PAK1). PAK1 regulates nuclear signaling and cytoskeleton reorganization (Bokoch, 2003), and proteins capable of differentiating its active and inactive conformational states might modulate many biological activities. The kinase is activated upon displacement of an auto-inhibitory domain that binds a hydrophobic patch on the kinase domain (Lei et al., 2000). Using a Rosetta-based molecular modeling program, Kuhlman and colleagues redesigned a small helical bundle protein – the 61 amino acid long hyperplastic discs protein HYP – to recognize this hydrophobic patch (Jha et al., 2010).

Computational redesign of HYP proceeded in several steps. Rigid-body docking provided the initial binding orientation of the scaffold relative to the PAK1 kinase domain (Gray et al., 2003). Iterative rounds of rotamer-based sequence design and conformational optimization (Rohl et al., 2004; Wang et al., 2007), followed by further molecular dynamics refinement (Ding et al., 2008), were then used to obtain candidate proteins with calculated binding energies similar to the binding energies of native complexes. The most successful design, dubbed Spider Roll, binds PAK1 with a K_D of 100 μM, preferentially recognizing the active form of the full-length kinase (Jha et al., 2010). NMR spectroscopic analysis confirmed that the designed protein preserves the original HYP helical bundle fold, and mutagenesis studies showed that it binds in the target cleft on the kinase. By optimizing the modest affinity of the Spider Roll protein for PAK1 through directed evolution, we hoped to explore how the design responds to evolutionary pressures.

2. Materials and methods

2.1. Bacterial and yeast strains

For general cloning and plasmid amplification the *Escherichia coli* strain XL1-Blue (*endA1*, *gyrA96* (Nal^R), *thi-1*, *recA1*, *relA1*, *lac*, *glnV44*, F' [::Tn*10* (Tet^R), *proAB*$^+$, *lacI*^{q}, Δ*(lacZ)*M15], *hsdR17* ($r_K^-m_K^+$); Stratagene) was used (Bullock et al., 1987). Yeast cell surface display was performed with *Saccharomyces cerevisiae* strain EBY100 (*MATa, GAL1-AGA1*::*URA3*, *ura3–52*, *trp1* (Trp^-), *leu2Δ1*, *his3Δ200*, *pep4*::*HIS3*, *prb1Δ1.6R*, *can1*, *GAL*) (Boder and Wittrup, 1997; Siegel, 2009). Spider Roll and PAK1 variants were overproduced in *E. coli* strain BL21 (DE3) (F^-, *ompT, gal, [dmc], [lon], hsdS*$_B$ *(*$r_B^-m_B^-$*)*, λ (DE3) [(UV5 P_{lac} expressed) T7 RNApol, *imm21,* Δ*nin5, Sam7 (*int^-*)*]) transformed with pLysS (Cam^R) (Studier et al., 1990).

2.2. DNA manipulations

Molecular cloning was performed according to standard procedures (Sambrook and Russell, 2001). Oligonucleotides were synthesized and purified by Microsynth AG (Balgach, Switzerland). DNA sequencing was performed on an ABI PRIZM 3100-Avant Genetic Analyzer (PE Applied Biosystems) using chain termination chemistry (Sanger et al., 1977) with the BigDye Terminator Cycle Sequencing Kit v3.1.

2.3. Plasmid construction

The Spider Roll gene on plasmid pQE-MBP-SR (Jha et al., 2010) was provided by Professor B. Kuhlman. For cell surface display, the gene was amplified by PCR using primers SRNdeInew (5′-TCT GCT AGC CGA TCC CAC CGC CAG G) and SRBamHI (5′-TTC GGA TCC GCC GTG AAA GAT GAT ACC AG), digested with *Nde*I and *Bam*HI, and ligated into the pCT vector (Boder and Wittrup, 1997), which had been digested with the same restriction enzymes, to give pCT-SR.

Plasmid pQE-MBP-SR was also used directly for protein production. It encodes a fusion between Spider Roll and maltose binding protein (MBP). The sequence ENLYFQG, which can be cleaved by TEV protease, was used as the linker at the N-terminus of Spider Roll, and a hexahistidine tag was appended to the N-terminus of MBP to facilitate purification. To generate the analogous construct for MSR7, pQE-MBP-MSR7, the MSR7 gene selected during directed evolution was amplified by PCR using the primers BamHIfwSR (5′-TGC TAG CGG ATC CCA CC) and SalIrvSR (5′-TTG GTC GAC TTA GCC GTG AAAG ATG ATA CC). The PCR fragment was digested with *Sal*I and *Bam*HI and ligated into the pQE-MBP (Jha et al., 2010) (*SalI*/*Bam*HI) acceptor. The primer pQE-H6-MBP-3′ (5′-GGC GGCA ACC GAG CGT TCT G) was used for sequencing.

2.4. Library construction

The Spider Roll gene was mutagenized by an error-prone polymerase chain reaction (epPCR) using the GeneMorph II Random Mutagenesis Kit (Stratagene) with primers YD-1 (5′-GGC AGC CCC ATA AAC ACA CAG TAT) and YD-12 (5′-GTA CGA GCT AAA AGT ACA GTG GGA). As templates for rounds one and two of directed evolution, pCT-SR and the plasmid pool from the first round of screening, pCT-LSRA4, were used, respectively, at a concentration of 0.1 ng/μl. Electrocompetent EBY100 yeast cells were cotransformed with the mutagenized Spider Roll gene and the *Nde*I/*Bam*HI fragment of the pCT vector (insert:vector = 10:1) as previously described (Colby et al., 2004). The first library, LSRA0, afforded 10^6 EBY100 transformants, whereas the second, LSRB0, gave 5×10^5 clones. For the third round of mutagenesis, genes selected in the previous sorting round were diversified by DNA shuffling. The clones were amplified by PCR with the YD-1 and YD-12 primers and digested with DNase I. The digestion reaction was quenched by addition of EDTA followed by heat inactivation of DNase I at 70 °C for 10 min. The DNA was precipitated with ethanol and purified on a 3% agarose gel. Fragments between 20–70 bp long were extracted and further purified using a NucleoSpin gel extraction kit (Macherey–Nagel). The DNA was concentrated and reassembled by a PCR-like reaction in the absence of primers [2 min at 95 °C; 40 × (0.5 min at 95 °C, 0.75 min at 45 °C, 3 min at 72 °C); 10 min at 72 °C]. Without further purification, a 6 μl aliquot from the reaction mixture was subjected to a PCR reaction using primers YD-1 and YD-12 to amplify the reassembled gene products [2 min at 95 °C; 40 × (0.5 min at 95 °C, 0.5 min at 50 °C, 1 min at 72 °C); 5 min at 72 °C]. EBY100 yeast cells were then transformed by electroporation as described above. The final library, LSRC0, consisted of 1.5×10^5 transformants.

2.5. Flow cytometry and fluorescence-activated cell sorting

Cells were incubated at pH 6 in 5 ml SD-CAA (20 g/L glucose, 6.7 g/L yeast nitrogen base, 5 g/L casamino acids, 10.19 g/L $Na_2HPO_4{\cdot}7H_2O$, 8.56 g/L $NaH_2PO_4{\cdot}H_2O$) for one day at 30 °C and 250 rpm. In this system, expression of the Spider Roll gene is under