*Protein Engineering, Design & Selection*, 2016, vol. 29 no. 1, pp. 1–9
doi: 10.1093/protein/gzv055
Advance Access Publication Date: 20 October 2015
Original Article


Original Article

# Engineering a genetically encoded competitive inhibitor of the KEAP1–NRF2 interaction via structure-based design and phage display

**Gurkan Guntas[1], Steven M. Lewis[1], Kathleen M. Mulvaney[2,3], Erica W. Cloer[2,3], Ashutosh Tripathy[1], Thomas R. Lane[1], Michael B. Major[2,3], and Brian Kuhlman[2,3,*]**

[1]Department of Biochemistry and Biophysics, [2]Department of Cell Biology and Physiology, and [3]Lineberger Comprehensive Cancer Center, University of North Carolina, 120 Mason Farm Road, Genetic Medicine Building 3010, Chapel Hill, NC 27599-7260, USA

*To whom correspondence should be addressed. E-mail: bkuhlman@email.unc.edu
Edited by Shohei Koide

Received 21 September 2015; Revised 21 September 2015; Accepted 24 September 2015

## Abstract

In its basal state, KEAP1 binds the transcription factor NRF2 ($K_d$ = 5 nM) and promotes its degradation by ubiquitylation. Changes in the redox environment lead to modification of key cysteines within KEAP1, resulting in NRF2 protein accumulation and the transcription of genes important for restoring the cellular redox state. Using phage display and a computational loop grafting protocol, we engineered a monobody (R1) that is a potent competitive inhibitor of the KEAP1–NRF2 interaction. R1 bound to KEAP1 with a $K_d$ of 300 pM and in human cells freed NRF2 from KEAP1 resulting in activation of the NRF2 promoter. Unlike cysteine-reactive small molecules that lack protein specificity, R1 is a genetically encoded, reversible inhibitor designed specifically for KEAP1. R1 should prove useful for studying the role of the KEAP1–NRF2 interaction in several disease states. The structure-based phage display strategy employed here is a general approach for engineering high-affinity binders that compete with naturally occurring interactions.

**Key words:** KEAP1, NRF2, protein design, redox regulation, Rosetta molecular modeling program

## Introduction

NRF2 (nuclear factor erythroid 2-related factor 2) is a transcription factor that serves as a key modulator of cellular oxidative environment (Nguyen *et al.*, 2009). Under basal conditions, NRF2 binds the Kelch domain of homodimeric KEAP1 (Kelch-like ECH-associated protein 1) through its high-affinity ($K_d$ 5 nM) and low-affinity motifs ($K_d$ 1 μM) (Lo *et al.*, 2006; Tong *et al.*, 2006). KEAP1, via its BTB domain, also binds CUL3 and recruits NRF2 to the CUL3-RING E3 ligase complex for ubiquitylation and subsequent degradation (Fig. 1A) (Nguyen *et al.*, 2009). Under oxidative stress, cysteine residues on KEAP1 are modified resulting in inhibition of NRF2 ubiquitylation (Taguchi *et al.*, 2011). As a result, NRF2 accumulates and translocates to the nucleus where it initiates transcription of a myriad of cytoprotective genes.

Activation and deactivation of NRF2 have been implicated in a variety of disease states. Activation of the NRF2 defense response has been shown to protect against neurodegenerative diseases, inflammation, cardiovascular disease, diabetes and the initiation of cancer (DeNicola *et al.*, 2011). However, it has also been shown that NRF2 activation can protect cancer cells from oxidative stress, chemotherapeutic agents and radiotherapy (Ohta *et al.*, 2008; Hayes *et al.*, 2010; DeNicola *et al.*, 2011). Indeed, a variety of loss-of-function mutations in KEAP1 that lead to NRF2 activation have been discovered in human cancers (Hast *et al.*, 2014).

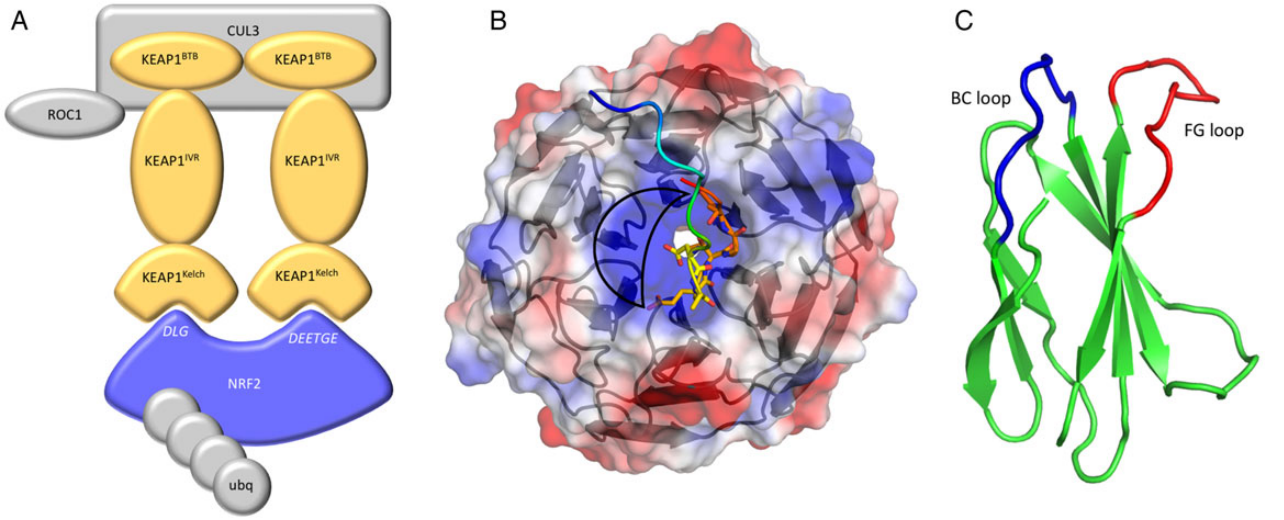


**Fig. 1** The KEAP1–NRF2 interaction. (**A**) NRF2 binds to the Kelch domain of homodimeric KEAP1 via high-affinity (DEETGE) and low-affinity binding motifs (DLG). KEAP1 also binds to the E3 ubiquitin ligase, Cullin3 (CUL3), which leads to ubiquitylation of NRF2. (**B**) A crystal structure of the KEAP1 Kelch domain bound to a 16mer peptide from NRF2, drawn from PDB 2FLU (Lo *et al.*, 2006). The binding motif is highlighted in sticks. The KEAP1 surface is colored according to electrostatic charge. A black highlight shows the unfilled portions of the pocket targeted with the non-anchor loop by the AnchoredDesign protocol. (**C**) A crystal structure of 10th type III domain of human fibronectin with the FG loop colored as red (used as the anchor loop in our design procedure) and the BC loop.

Regulation of NRF2 via pharmaceuticals has been proposed for the treatment of several diseases (Magesh *et al.*, 2012). A NRF2-activating compound, dimethyl fumarate, was recently approved by the Federal Drug Administration for the treatment of multiple sclerosis (Moharregh-Khiabani *et al.*, 2009). Many NRF2-activating molecules, including dimethyl fumarate, function by covalently modifying cysteines on KEAP1, which leads to lower rates of NRF2 ubiquitylation (Zhang and Hannink, 2003; DeNicola *et al.*, 2011). These molecules are not specific for KEAP1, but rather non-discriminately react with activated cysteines on a variety of proteins via oxidation or alkylation reactions (Zhang and Hannink, 2003). Several chemical groups are adept at this reaction including Michael acceptors, oxidizable quinones, isothiocyanates and vivinal dimercaptans. Many of these groups are found in naturally occurring, plant-derived phytochemicals that have been shown to activate the NRF2 pathway (Hayes *et al.*, 2010). Although they are clearly useful for activating NRF2, these molecules do not provide precise tools for probing the KEAP1/NRF2 pathway. As they are reactive toward many proteins in a cell, observed phenotypes may be due to a variety of factors. To create a molecule that can more specifically probe the function of KEAP1, we engineered a protein that binds tightly to KEAP1 and competitively inhibits NRF2 binding.

Proteins have proved more effective at inhibiting protein–protein interactions than small molecules, as they are more adept at binding large surfaces (Arkin and Wells, 2004; Schön *et al.*, 2011). One approach for creating an inhibitor is to structurally mimic the naturally occurring interface that you are trying to inhibit. Most straightforwardly, this can be achieved by using a re-engineered version of one of the binding partners as the inhibitor. Typically, the natural binder is modified so that it no longer activates the endogenous signaling pathway. For instance, the extracellular portion of a cell surface receptor when expressed alone can bind signaling molecules and prevent them from activating the functional receptor (Peppel, 1991; Moreland, 1998). Directed evolution techniques like phage display or yeast display can be used to identify mutations that increase the affinity of the inhibitor for the target, and therefore allow the re-engineered version to outcompete the naturally occurring binding partner. However, in some cases, it may be advantageous to use an alternative protein as the scaffold for inhibitor design, for example, when there is a need for more improved expression, stability, solubility or pharmacokinetic properties.

When using unrelated protein scaffolds to design inhibitors, mimicry of the target interface can still be used to gain affinity and specificity (Richards *et al.*, 2003). Schief *et al.* have developed a computational grafting protocol that searches the protein database for proteins that can orient a set of surface residues in a manner that superimposes with residues from the target interface (Ofek *et al.*, 2010; Azoitei *et al.*, 2011). Using this approach, they have grafted antigen epitopes into scaffolding molecules in a manner that promotes binding to target antibodies and potentially function as vaccines. Recently, we developed a grafting protocol in the molecular modeling program Rosetta, called AnchoredDesign, for use with scaffolds that have surface exposed loops amenable to mutation (Lewis and Kuhlman, 2011). In this approach, a contiguous set of binding-critical interface residues from a binding partner of the target protein are inserted into one of the loops of the scaffold. These residues serve as an anchor point for the interaction. Surrounding residues and loops are further optimized using computational or screening techniques. Potential advantages of the anchored design approach versus screening with naïve loop libraries are that the binding site and orientation can be pre-specified and that it allows for creating libraries that are enriched in binders, increasing the probability that a tight binder will be identified.

The structure of the KEAP1–NRF2 interaction makes it well suited for the anchored design protocol. In the bound complex, a flexible region of NRF2 inserts into a large pocket on the Kelch domain of KEAP1 (Fig. 1B). Peptides derived from this region of NRF2 bind to KEAP1 with an affinity of ~100 nM (Chen *et al.*, 2011; Hancock *et al.*, 2013) and can serve as an excellent anchor for insertion into a scaffolding protein. As a scaffold, we used the 10th type III domain of human fibronectin (FN3), an IgG-like β-sandwich protein that has

three complementary determining region (CDR)-like flexible loops (BC, DE and FG loops) that are highly tolerant to mutagenesis (Batori *et al.*, 2002). Fn3 variants, often called monobodies, have been engineered to bind a wide array of target molecules using selection strategies such as phage display (Koide *et al.*, 2012). Unlike antibodies, the domain is monomeric and does not contain any disulfides and is therefore more stable in the reducing environment of a cell. To engineer a monobody that binds to KEAP1, we grafted a portion of a NRF2 peptide into the FG loop of FN3 and then used molecular modeling and phage display to optimize surrounding residues on the FG loop and the BC loop for enhanced affinity toward KEAP1. The tightest binder identified with these approaches binds to KEAP1 with an affinity of 300 pM, outcompetes wild-type NRF2 for binding to KEAP1, and activates NRF2-mediated transcription in living cells.

## Experimental procedures

### Computational design of anchor-FG loop and BC loops

Computational modeling of monobody–KEAP1 complexes proceeded in two stages. First, the Rosetta protocol AnchoredPDBCreator was used to create starting models for the design simulations (Lewis and Kuhlman, 2011). Second, flexible-loop design simulations were performed with Rosetta protocol AnchoredDesign to search for monobody loop sequences compatible with the inserted anchor conformation and KEAP1 binding.

AnchoredPDBCreator was used to create intermediate models containing structural data from two sources (Fig. 2, top). First, the KEAP1 Kelch domain with associated NRF2 peptide was sourced from PDB 2FLU, a 1.5 Å resolution structure of this complex (Lo *et al.*, 2006). Second, the Fn3 scaffold was sourced from PDB 1FNF, a 2 Å resolution structure containing the 7th through 10th type III domains of human fibronectin (Leahy *et al.*, 1996). AnchoredPDBCreator was used to combine the two structures such that loop residues of the Fn3 domain were overwritten with the conformation and identity of the NRF2-derived peptide. This chimera was then placed next to the KEAP1 Kelch domain such that the copied anchor residues were in their correct placement relative to the binding interface. Minor loop refinement of the non-anchor portion of the loop was then performed to resolve any major clashes between the Fn3 scaffold and KEAP1 (Fig. 2, top). The key data in this intermediate model are that the anchor and target structures maintain their crystallographic conformation even while the anchor has been embedded into a loop of the scaffold protein.

Given that we were performing flexible-backbone loop design, it was not obvious to us what loop lengths, anchor sequences or anchor placements within the loops are appropriate. We tested two different NRF2-derived anchor sequences, DEETGE and ETGE, in several loop placements and lengths (Supplementary Table SI).

In the second computational stage, design simulations were performed with Rosetta's AnchoredDesign protocol (Lewis and Kuhlman, 2011). AnchoredDesign implements the anchored design idea by remodeling both the sequence and conformation of surface loops of a scaffold protein while maintaining the embedded crystallographic anchor–target interface. Anchor residues are held fixed both internally and relative to the scaffold protein. Loop residues in the non-anchor loops are remodeled with Rosetta's implementations of the cyclic coordinate descent (Wang *et al.*, 2007) and kinematic loop closure (Mandell *et al.*, 2009) algorithms. Non-anchor residues in the anchor-containing loop can be remodeled with these same algorithms, provided that the internal degrees of freedom of the anchor residues are not visible as variable to the loop remodeling algorithms, and Rosetta's internal coordinate folding mechanism is modified appropriately (Wang *et al.*, 2007; Lewis and Kuhlman, 2011). Other residues near the interface are given the freedom to relax their side chain but not change sequence. Residues far from the interface are not modified.

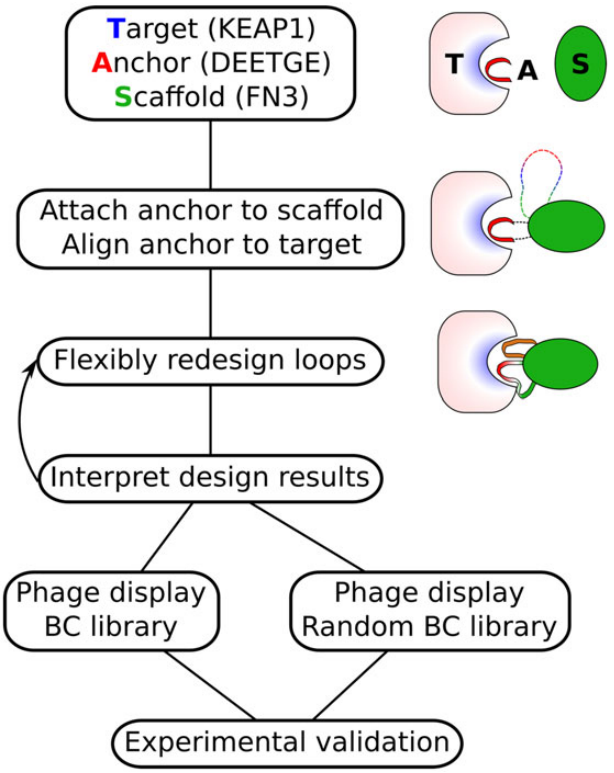


**Fig. 2** Protocol flow chart. This flow chart outlines the combined computational and experimental protocol used to identify tight binders to KEAP1. See also the Supplementary Methods for command lines. At the top are two steps performed inside the AnchoredPDBCreator protocol. In the middle are steps performed in the AnchoredDesign protocol. Note that the middle step loops on itself. Here, we performed two cycles of the AnchoredDesign modeling. The bottom portion represents data analysis and translation of modeling results into phage display libraries, followed by experimental validation.

An initial batch of models for each loop length and anchor placement in Supplementary Table SI were created. Each of these experiments produced ~1000–1500 models of varying sequences and conformations. The best-scoring and structurally interesting or distinct models from the first round were used as inputs for another round of modeling that focused on only small perturbations to the protein backbone (Fig. 2, middle portion). Top scoring sequences from this round were used to create the directed library described in results (Fig. 2, bottom portion). See the Supplementary Methods for the Rosetta command line options used for both rounds of modeling.

### Library construction

Both the directed-BC and randomized-BC libraries were constructed using Kunkel mutagenesis as fusions with the pIII protein from filamentous bacteriophage M13 and a DsbA signal sequence (Fellouse *et al.*, 2007; Wojcik *et al.*, 2010). Twenty-eight mutagenic oligonucleotides (Supplementary Table SII) encoding for the designed BC loops were pooled to have an equimolar solution of all 587 BC loop sequences. One hundred μM of either the oligonucleotide mixture for the directed-BC library or the single mutagenic oligonucleotide that

encodes for 7 NNK degenerate codons to create the randomized-BC library was 5′-phosphorylated with 20 units of T4 polynucleotide kinase in the presence of 1 mM ATP, 5 mM DTT, 10 mM $MgCl_2$ and 50 mM Tris–HCl pH 7.5. Phosphorylated oligos were annealed to 20 µg of uracil-incorporated single-stranded phagemid with 5:1 oligo:template ratio in the presence of 10 mM $MgCl_2$ and 50 mM Tris–HCl pH 7.5 by cooling from 90°C to 30°C over 50 min followed by incubation at least 5 min at 4°C. Annealed oligonucleotides were extended with 30 units of T7 DNA polymerase and the heteroduplex DNA was ligated with 2 kU of T4 DNA ligase by incubation at 20°C for 3 h in the presence of 0.85 mM each dNTP, 0.33 mM ATP, 5 mM DTT, 10 mM $MgCl_2$ and 50 mM Tris–HCl pH 7.5. Desalted reaction products were transformed into electrocompetent SS320 cells. Pooled transformed bacteria were expanded into 500 ml 2× TY medium supplemented with 100 µg/mL ampicillin and $10^{10}$ pfu/mL M13K07 helper phage. Phage library was produced by shaking (225 rpm) for 16 h at 37°C. Prior to superinfection with the helper phage, serial dilutions were plated to estimate library size. Phage was precipitated and quantified as described (Fellouse *et al.*, 2007).

### Library selections

Purified His-tagged KEAP1 was chemically biotinylated targeting three amine groups (N-terminus and two lysine residues) with 20-fold excess E-Z link Sulfo-NHS-LC-Biotin (Pierce 21335) in phosphate-buffered saline (PBS)/5 mM β-mercaptoethanol pH 7.4 buffer at 4°C. After 3 h, reaction products were desalted using a PD-10 column to remove excess biotin. Biotinylated protein was serially dialyzed twice against 500× volume of PBS/5 mM β-mercaptoethanol pH 7.4 buffer to remove trace biotin.

For the first three rounds of biopanning, streptavidin magnetic beads were pre-blocked overnight at 4°C with 0.5% w/v bovine serum albumin (BSA) in PBS, saturated with biotinylated KEAP1 (15–150 nM) and incubated with phage ($5 \times 10^{11}$–$10^{13}$) pre-adsorbed with pre-blocked, but uncoated beads. Beads were washed 10 times, and bound phage was eluted with 0.2 M glycine/HCl 5 mg/mL BSA pH 2.2 solution. For Rounds 4 thru 6, pre-adsorbed phage ($10^{10}$–$10^{11}$) was first incubated with biotinylated KEAP1 followed by capture with blocked streptavidin beads. Beads were washed 15–20 times before bound phage was eluted as above. After each round, log-phase SS320 cells were infected with the eluted phage followed by superinfection with the helper phage to amplify the phage for the next round.

### Protein expression and purification

All Fn3 variants were cloned into pGEX-4T-1 (GE Healthcare 28-9545-49) vector via *BamHI* and *XhoI* restriction sites. Proteins were expressed as a C-terminal fusion to glutathione-*S*-transferase (GST). Cells were lysed in PBS supplemented with 1 mM phenylmethylsulfonyl fluoride. GST-Fn3 fusions were purified from cleared lysates using pre-packed 5 ml glutathione columns (GE Healthcare). The eluate was treated for 2 days with 50 units of thrombin (Sigma) to cleave the GST away. The proteolyzed sample was concentrated and applied to Superdex-75 Hi-load 16/60 gel filtration chromatography column (GE Healthcare) to separate Fn3 from GST and thrombin. The free Fn3 variants were used in affinity and stability measurements, whereas GST-fused form was dialyzed twice against PBS to remove glutathione and used in pull-down experiments.

The Kelch domain of KEAP1 was expressed using a pET15b-based expression plasmid with a N-terminal His-tag followed by a thrombin cleavage site (Lo *et al.*, 2006). The expressed protein was purified by His-tag affinity chromatography followed by anion exchange. Finally, the protein was refined using Superdex-75 Hi-load 16/60 gel filtration chromatography.

For the co-purification of the Kelch domain in complex with R1, expression plasmids encoding the GST-R1 fusion and His-Kelch were co-transformed into BL21(DE3)pLysS cells. After an hour of recovery in super optimal broth medium, cells were transferred to a shake flask containing 30 ml of 2× TY/ampicillin medium. After 5 h, cells reached the log phase and were transferred to the 3-l LB/ampicillin medium. Following another 3 h of growth, cells were induced with isopropyl beta-D-1-thiogalactopyranside at 20°C for overnight expression. The GST-tagged R1 complexed with Kelch domain was first purified by glutathione affinity chromatography. The eluate was concentrated and resolved by Superdex 200 column. Fractions containing Kelch domain and GST-R1 fusion were treated with 50 units of thrombin to remove the His- and GST tags. The products were passed through glutathione column twice to remove GST and uncleaved GST-R1. The flow-through was resolved with Superdex-75 column to recover Kelch-R1 complex.

All proteins were purified to >90%. Concentrations were determined using $A_{280}$ readings and extinction coefficients predicted by ProtParam.

### Isothermal titration calorimetry

Purified Fn3 variants and Kelch domain were dialyzed in the same container against the ITC buffer (20 mM $KH_2PO_4$, 150 mM NaCl, 5 mM β-mercaptoethanol, pH 7). The post-dialysis buffer was also used as the diluent for all experiments. All titrations and measurements were performed using the MicroCal Auto-iTC system (GE Healthcare). Twenty microliters of 200–300 µM Fn3 variant is titrated in 10 successive injections to the cuvette containing 6–15 µM Kelch. Data were fitted to the one-site binding curve to derive the $K_d$ values.

### Competitive fluorescence polarization experiments

All fluorescence polarization (FP) experiments were conducted using a Jobin Yvon Horiba FluoroMax3 fluorescence spectrometer. TAMRA-labeled peptides (produced by UNC Peptide Core Facility) were excited with polarized light at 555 nm, and the polarization of emitted light was measured at 583 nm. Titrations were performed with 20 nM peptide and 350 nM KEAP1 in the starting solution; this represents ~80% bound peptide as determined by peptide-KEAP1 experiments. Monobody was then titrated into the cuvette and polarization measured after each addition. These data were fit for IC50 using a procedure reported by Lungu *et al.* (2012) and refined into $K_i$ with an Internet-based calculator (Nikolovska-Coleska *et al.*, 2004). Certain experiments violate the assumptions of Lungu *et al.*, and fits were generated from an iterative numeric fitting program adapted from Purbeck *et al.* (2010).

### Surface plasmon resonance

Surface plasmon resonance (SPR) experiments were performed using a BIORAD Proteon XPR36 biosensor. Fifty nM biotinylated KEAP1 (via lysine modification) was captured on a streptavidin-coated sensor chip, and sensorgrams were collected at varying concentrations of R1.

### Circular dichroism

Thermal melts were performed using a Jasco J-815 circular dichroism instrument, with 20 µM R1 or wild-type Fn3 in 20 mM $KH_2PO_4$ pH 7 buffer.

### Cell-based luciferase assays

R1 and FN3 wild type (WT) were gateway cloned into a mammalian expression CMV-Phage-Flag Gateway vector using pCR8/GW/TOPO TA Cloning Kit (Invitrogen). HEK293T cells were maintained in Dulbecco's modification of Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (FBS; Hyclone) in a 37°C humidified incubator with 5% $CO_2$. Transfection of expression plasmids was done with Lipofectamine 2000 (Invitrogen). HEK293T cells were transiently transfected in 48-well format with expression constructs, Flag-R1, Flag-Fn3 WT, FLAG-KEAP1, FLAG-NRF2, hQR41-ARE luciferase and a control plasmid containing *Renilla* luciferase driven by a constitutive cytomegalovirus (CMV) promoter. Approximately 24 h post-transfection, NRF2-mediated transcription was measured as the ratio of firefly to *Renilla* luciferase activity (Promega Dual-Luciferase Reporter Assay System). Values shown are the mean normalized relative light units (RLU) from one representative experiment of three biological replicate experiments.

### Western blotting and immunoprecipitations

HEK293T cells were lysed in 1% Triton buffer [1% Triton-X, 10% glycerol, 50 mM Tris–HCl, 150 mM NaCl, 2 mM EDTA supplemented with protease and phosphatase inhibitors (Thermo Scientific)] for 30 min on ice. Lysates were cleared by 10 min centrifugation at full speed. An input of 10% was taken, and the remaining lysate was divided into six immunoprecipitations (APs): beads only, 1 µg GST-Fn3 WT, or 0.1, 0.25, 0.5 or 1 µg of GST-R1 protein. APs were completed using 30 µl each of glutathione affinity beads (GE Healthcare), rotating for 1 h at 4C. Streptavidin affinity purifications (APs) were completed similarly to GST-APs, except using streptavidin beads (GE Healthcare) and these complexes were purified from HEK293T cells stably expressing SBPHA-KEAP1. APs were washed three times with lysis buffer, and bound proteins were eluted from the beads using SDS–PAGE sample buffer. Lysates were resolved on 4–12% SDS–PAGE gels (Invitrogen), transferred onto nitrocellulose membranes and probed overnight at 4°C with the following antibodies: KEAP1 (Protein Technologies, 10503-2-AP), GST (Cell Signaling, 2622), β-actin (Sigma, A2066), FLAG (Sigma, M2), NRF2 (Epitomics, 62352) and HO-1 (Abcam, 13243). Secondary HRP-conjugated antibodies (Jackson Laboratories) were applied at 1:10 000 dilution for 30 min at room temperature. Western blots were developed using the Pierce ECL western blotting reagent (Thermo Scientific). Data shown are representative experiments from 2 to 3 biological replicates.

## Results

The acidic NRF2 peptide (-DEETGE-) docks into a large positively charged pocket on KEAP1. It occupies roughly half of the volume of the pocket, while the rest of the space is filled with water (Fig. 1B) (Lo *et al.*, 2006). We hypothesized that we could create a KEAP1-binding monobody by building the DEETGE sequence into one loop of a monobody, and then use a neighboring loop in the monobody to fill the rest of the pocket on KEAP1. To determine which loops on the monobody are best positioned for this approach, we performed molecular docking experiments with the AnchoredDesign protocol in Rosetta. In this protocol, the anchor sequence, in our case the NRF2 peptide, is inserted into one of the monobody loops and the sequence and conformation of neighboring monobody loops are optimized for binding against the target (Fig. 2). Previous work has shown that the BC and FG on monobodies are tolerant to mutagenesis and that they form an effective surface for binding proteins (Koide *et al.*, 2012).

We computationally tested 11 alternative loop lengths and anchor positions for the BC and FG loops, with anchor sequences of ETGE or DEETGE (Supplementary Table SI). A high-resolution crystal structure of the NRF2 peptide bound to KEAP1 (pdb code: 2FLU) was used as the starting point for the simulations. During these simulations, the coordinates and sequence of the anchor residues were held fixed, while surrounding residues in the anchor loop were allowed to sample alternative conformations and sequences in order to accommodate the anchor residues into the monobody scaffold. The position of the anchor residues was not varied with respect to the binding partner, KEAP1, but the rest of the monobody scaffold adopted different relative orientations with respect to the target as the connections between the anchor and monobody were perturbed. In this respect, the protocol is best described as a tethered-docking simulation.

For each of the alternative loop lengths and anchor positions in Supplementary Table SI, ~1000–1500 independent trajectories were performed to identify low-energy sequence/structure pairs. The resulting models were then ranked by the total energy of the complex, quality of contacts at the interface and loop metrics (Lewis and Kuhlman, 2011). Several dozen low-energy models were selected for further study. All of these models used an anchor inserted into the FG loop of FN3: although we tested BC loop insertions of an ETGE anchor, no models scored well after the first round in AnchoredDesign. These selected models each seeded another 1000–1500 AnchoredDesign runs aimed at generating smaller perturbations. We then examined the lowest scoring models from this second round of design to determine if there was an anchor position and loop sequence that was overrepresented in the lowest scoring models. We determined that one such sequence was yavRDEETGEFHWPis. Here, lowercase indicates constant Fn3 sequence for framing, the underline indicates the anchor and other capital letters indicate design positions. The post-anchor phenylalanine residue happens to be the NRF2 residue immediately following the ETGE sequence as well; this represents Rosetta's convergence onto biology but was indeed a designed position. The final designed proline residue is native to the scaffold, but again was allowed to vary during the design simulation. For reference, this anchor sequence is compatible with label BC plus 1, 6_77.83 in Supplementary Table SI. The anchor loop sequence requires a one-residue deletion in the Fn3 FG loop, and the accompanying low-energy BC loop sequences included a single insertion. Our database of models contained 7181 structures that converged to this particular anchor loop sequence, but with a diverse set of sequences and conformations for the BC loop. The DE loop did not make significant contact with KEAP1 in any of the low-energy designs.

The diverse set of BC sequences prompted us to design two combinatorial libraries from which we sought to select tight binders. In both libraries, the sequence of the FG loop was set to the design-obtained sequence RDEETGEFHWP. The first library was a directed library based on the low-energy models produced during the anchored design simulations. This directed library involved 587 different BC loop sequences and includes BC sequences from the top 80 computational models. The other 507 loop sequences in this library were highly similar to the top 80 designs (Supplementary Table SII). In parallel, we constructed a randomized library where seven positions in the BC loop were randomized using the degenerate codon NNK, for a theoretical size of $3.4 \times 10^{10}$ alternative DNA sequences. This library served as a control to gauge the level of modeling success and as another potential source of a potent competitive inhibitor. It is important to note that this library is still directed in the sense that we have fixed the anchor sequence and site. The experimental size of the randomized library was $\sim 2 \times 10^{10}$, and therefore, there was a

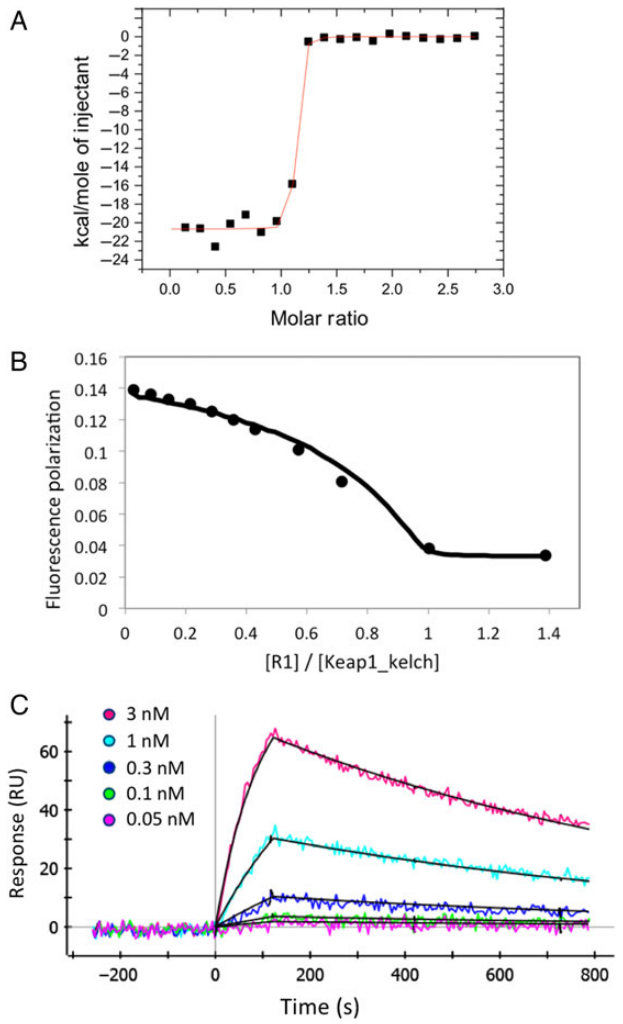


**Fig. 3** Biophysical characterization of R1 binding to KEAP1. (**A**) An ITC experiment titrating R1 into a solution of 10 μM KEAP1 Kelch domain. (**B**) A competitive FP assay titrating R1 into a sample containing the KEAP1 Kelch domain pre-incubated with a fluorescently labeled peptide. R1 binding dislodged the peptide and resulted in a decrease in FP. The data were fit assuming single-site binding (Purbeck *et al.*, 2010). (**C**) SPR data with KEAP1 Kelch domain immobilized on the sensor chip and R1 applied at the indicated concentrations.

~59% chance that a particular DNA sequence was sampled during selection. For both libraries, we used phage display to select for high-affinity binders as the FN3 scaffold has been shown to express robustly on the surface of phage. Following several rounds of biopanning, we sequenced clones from the directed library to assess library diversity. The library did not converge to a few clones, suggesting that the affinities of the winner clones were comparable, or that the selective pressure was not stringent relative to the affinity of many of the best binders in the pool. Among the selected clones from the directed library, we chose a few that were either exact BC sequences produced by the design simulations or were only a mutation or two away from a designed sequence (Supplementary Table SII). For the randomized library, we isolated a few winner clones that yielded high phage titers under highly stringent conditions indicative of high affinity for the target.

We purified the selected clones and measured affinity for the KEAP1 Kelch domain using isothermal titration calorimetry (ITC) (Fig. 3A) and a competitive FP assay where the FN3 variants competed with the fluorescently labeled NRF2 peptide for the Kelch domain of KEAP1 (Fig. 3B). This assay, in addition to quantifying affinity, is useful to validate that FN3 variants are binding the target-binding surface on KEAP1, and that binding should be competitive with NRF2 binding. For the affinity range of 0.1–1 μM, the two methods yielded comparable values. For tight binders, however, ITC was not as sensitive. In this case, we relied on the competitive FP assay to determine $K_d$ and also measured the binding affinity of the tightest binder with SPR (Fig. 3C).

The affinity we measured for the isolated NRF2 peptide for KEAP1 ($K_d$ = 141 nM) was consistent with published results (Hancock *et al.*, 2013). An FN3 construct, wtBC, with the computationally designed FG loop containing the DEETGE sequence and a naïve BC loop had 10-fold decreased affinity ($K_d$ = 1.56 uM) relative to the source peptide (Table I). We next examined the binding affinities of Fn3 variants selected from our directed and randomized libraries. The selected clones from the directed library showed modest improvements over our wtBC control, and had affinities ranging from 100 to 850 nM. To examine if the improvements in binding affinity were due to the computationally designed interactions in the BC loop, we performed alanine mutagenesis on BC loop residues of D1 and D4 that were predicted by the Rosetta modeling to form favorable interactions with the positively charged groove on KEAP1. Alanine mutagenesis did not significantly perturb binding affinities, suggesting that these designed contacts were not forming as designed or were not energetically favorable.

In contrast to the directed library, the two clones isolated from the randomized library showed dramatic increases in binding affinity. In particular, the clone R1, had a $K_d$ of 300 pM, which is 5000-fold tighter than our wtBC control 'starting point' and 15-fold tighter than the reported affinity for the interaction of full-length NRF2 for KEAP1 (5 nM) (Tong *et al.*, 2006). As expected for a binding affinity in this range, R1 and KEAP1 co-elute when run on a gel filtration column.

Encouraged by the tight binding between R1 and the Kelch domain, we next tested if (i) R1 binds KEAP1 in human cultured cell lines, (ii) R1 expression increases NRF2 protein levels and (iii) R1 activates NRF2-dependent transcription. First, protein lysate from HEK293T cells was spiked with different amounts of the GST-wt-Fn3 or GST-R1 recombinant proteins. GST pull-down with glutathione beads followed by western blot analysis revealed that R1 captured KEAP1, whereas Fn3 WT did not (Fig. 4A). Second, we expressed FLAG-tagged Fn3 WT or FLAG-R1 in HEK293T cells stably expressing SBP-tagged KEAP1 (Fig. 4B). Streptavidin AP of KEAP1 and subsequent western blot analysis revealed associated FLAG-R1 but not FLAG-Fn3 WT. These data suggest that R1 and KEAP1 co-complex in cells. To evaluate whether R1 expression increased NRF2 protein levels in cells, FLAG-R1 was overexpressed in HEK293T cells before western blot analysis; NRF2 levels were moderately and reproducibly increased by R1 but not FN3 WT (Fig. 4C). Last, we tested the impact of R1 expression on NRF2-dependent gene expression. FLAG-R1 expression promoted NRF2-dependent transcription of the hQR41 firefly luciferase reporter gene; this activation was dose dependent (Fig. 4D and E). Consistent with this observation, expression of FLAG-R1 in HEK293T cells increased the expression of heme oxygenase-1, HMOX1, a NRF2 target gene (Fig. 4B).

## Discussion

We tested two strategies for engineering the BC loop: design of specific loop sequences with Rosetta (587 loop sequences) and phage display

**Table I.** Binding affinities of engineered monobodies and peptides for KEAP1 as measured by ITC, FP or SPR

| | BC loop | FG loop | ITC | FP | SPR |
|---|---|---|---|---|---|
| Controls | | | $K_d$ (nM) | $K_d$ (nM) | $K_d$ (nM) |
| wt | APAVTV | TGRGDSPASSK | Undetectable | | |
| wtBC | APAVTV | RDEETGEFHWP | 1555 | | |
| Peptide | – | LDEETGEFL | (182) | 141 | (322) |
| Directed library | | | | | |
| D1 | AWSYDEV | RDEETGEFHWP | 104 | 201 | |
| D2 | GDLGTNT | RDEETGEFHWP | 116 | | |
| D3 | AWSYYEV | RDEETGEFHWP | 202 | | |
| D4 | GDGHEEI | RDEETGEFHWP | 849 | 690 | |
| D5 | GDHHEDI | RDEETGEFHWP | 328 | | |
| Randomized BC library | | | | | |
| R1 | RAYGYPS | RDEETGEFHWP | <3 | 0.3 | 0.30 |

For proteins D4 and D5, the underscored residue was in the directed library due to codon degeneracy and not the model database (see Supplementary Table SII). For the NRF2-derived peptide, ITC and SPR measurements (shown in parentheses) are from the literature (Tong *et al.*, 2006; Chen *et al.*, 2011).

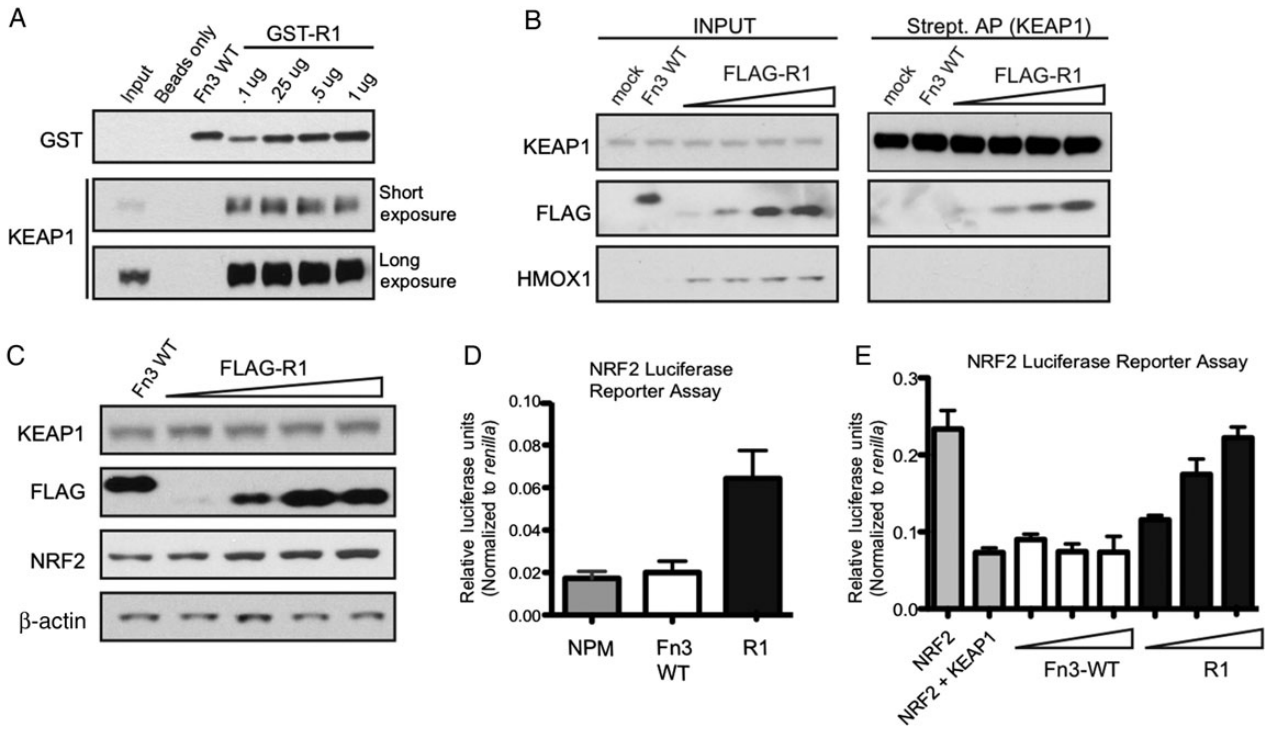


**Fig. 4** R1 binds to cellular KEAP1 and activates NRF2-dependent transcription. (**A**) HEK293T lysates were subjected to AP with recombinant GST-R1 at various concentrations (0.1–1.0 μg), GST-Wt Fn3 control (1.0 μg) or glutathione beads alone. Western blots demonstrate the amount of GST-tagged proteins and KEAP1 precipitated. (**B**) HEK293T cells stably expressing SBP-KEAP1 were transfected with FLAG-R1 at increasing DNA concentrations or with FLAG-Fn3 WT control and subjected to AP for KEAP1. Western blots show the amount of KEAP1-complexed R1. (**C**) HEK293T cells were transfected with FLAG-R1 at increasing concentrations of DNA or FLAG-Fn3 WT control before the western blot analysis. (**D** and **E**) HEK293T cells were transfected with the NRF2 transcriptional reporter (hQR41) and CMV-driven *Renilla* luciferase, with varying amounts of FLAG-WT Fn3 or Flag-R1 plasmid. Nucleophosmin expression served as a negative control. Data show the mean ± SD of relative luciferase units normalized to *Renilla*.

with degenerate codons placed at each position in the BC loop. Not surprisingly, the degenerate library produced the tightest binders as over 1 billion sequences were probed with this approach. However, it was disappointing that tighter binders were not identified among the Rosetta designed loops. The tightest binder from the Rosetta designed loops, D1, bound to Keap1 with an affinity similar to that observed for free peptide. Additionally, alanine scanning of the BC loop of D1 indicated that the side chains were not making strong favorable interactions with Keap1. This result highlights the challenge of performing accurate design and structure prediction with long loops. Nature frequently uses loops to mediate protein–protein interactions, but the most successful examples of computer-based design of protein–protein interactions have focused on interactions at more rigid secondary structural elements (Fleishman *et al.*, 2011; Stranges *et al.*, 2011). It is interesting that simply inserting the binding epitope into the FG loop led to ~10-fold drop in binding relative to the free peptide.

This suggests that either the FN3 scaffold creates some strain when the binding epitope adopts its bound conformation or that other elements of the scaffold, for instance, the BC-loop, are disfavoring binding.

When combined with phage display, the anchored design strategy was an effective approach for identifying tight binders against a specific binding site on KEAP1. The FN3 scaffold has three loops that are highly tolerant to mutagenesis. However, it is experimentally impossible to sample all permutations of amino acids on these loops. One strategy for handling this problem has been to design libraries that allow amino acids (i.e. A, S, Y, D) commonly observed in antibody CDR sequences (Fellouse *et al.*, 2007; Gilbreth *et al.*, 2008). It is possible to identify tight binders from such libraries, and the binders will often gravitate toward 'hot spots' on the target, but they are not guaranteed to bind the intended site on the target (DeLano, 2002). In these cases to identify competitive inhibitors, one may need to co-incubate the pool of binders with false targets lacking their proper interface residues while selecting for the true target, or use expensive semi-high-throughput screens that explicitly search for competitive binding. If the target and the competitor are of high affinity, it is a challenging task to design a single sequence or chemically diverse library that will yield both specific and tight binders that free the target from the competitor chemical. The anchored design strategy guarantees a biologically relevant mode of binding in addition to dramatically reducing the library complexity required to achieve competitive inhibition. In this particular case, exhaustive experimental search for the optimal seven-residue BC loop is nearly within the technical limits of combinatorial protein engineering, but searching exhaustively across multiple FN3 loops at once was neither possible nor necessary. The computational and selection portions of this method were quite complementary in that the random library found higher affinity binders than the computationally directed library, but the computationally identified FG loop sequence made construction of a random BC loop library enriched in high-affinity binders straightforward.

We have used anchored design to create a competitive inhibitor of KEAP1–NRF2 interaction. Full-length NRF2 binds KEAP1 via two non-contiguous epitopes of which the high-affinity epitope (-DEETGE-) is responsible for the majority of the binding energy to attain an overall dissociation constant of nearly 5 nM. In order to free NRF2 efficiently from KEAP1, we needed to design an FN3 variant with sub-nanomolar affinity. Rosetta's computational AnchoredDesign protocol coupled with phage display yielded protein R1 that bound KEAP1 with a $K_d$ of 300 pM. To our knowledge, this is the first sub-nanomolar peptide inhibitor of the KEAP1–NRF2 interaction (Chen *et al.*, 2011; Inoyama *et al.*, 2012; Steel *et al.*, 2012). Importantly, R1 is easily produced in bacteria, has a thermal stability comparable with wild-type Fn3 and co-purifies with KEAP1 when co-expressed. When spiked into mammalian cell extracts, R1 efficiently and specifically captured KEAP1 in a dose-dependent manner. When expressed in human cells, R1 induced higher NRF2 protein levels and activates NRF2-driven transcription.

As a high-affinity competitive inhibitor of the KEAP1–NRF2 interaction and strong NRF2 agonist, R1 offers a new molecular tool with translational potential. First, as a non-covalent KEAP1 inhibitor, R1 differs from cysteine-reactive small molecule activators. This promises increased specificity and diminished off-target effects. Second, R1 and other competitive inhibitors offer an elegant tool to further evaluate KEAP1 function. For example, in addition to NRF2, several additional proteins bind KEAP1 via an ETGE motif (i.e. WTX/FAM123B, DPP3, MCM3) (Hast *et al.*, 2013). It is probable that R1 will also block the binding of these proteins to KEAP1, the consequences of which are not yet known. Third, R1 and all competitive inhibitors offer preventative therapeutic opportunities, with the caveat that the intracellular delivery of proteins is still challenging. We found R1 to bind KEAP1 with picomolar affinity, thus strongly supporting future experiments that evaluate R1-driven cytoprotection. Fourth, R1 may also provide a means to correct KEAP1 mutations in cancer. Specifically, we previously demonstrated that ~50% of KEAP1 mutations found in lung squamous cell carcinoma bind NRF2, ubiquitylate NRF2, but fail to trigger the proteosomal degradation of NRF2 (Hast *et al.*, 2014). We hypothesize that these 'superbinder' mutations are defective in releasing ubiquitylated NRF2; this is somewhat consistent with the KEAP1 cycling model recently proposed (Baird *et al.*, 2013). Therefore, a competitive inhibitor like R1 might facilitate the dissociation of ubiquitylated NRF2 from the 'superbinder' KEAP1 mutations, restoring NRF2 degradation.

## Supplementary data

Supplementary data are available at *PEDS* online.

## Acknowledgements

The expression vector for KEAP1 was kindly provided by Mark Hannink, University of Missouri, and the pFN-OM6 vector used for phage display was provided by Sachdev Sidhu, University of Toronto. The hQR41-ARE luciferase reporter was a generous gift from Jeffrey Johnson, University of Wisconsin Madison.

## Funding

This work was supported by the National Institutes of Health (RO1-GM073960 to B.K.) and an American Cancer Society research scholar award to M.B.M. E.C. was supported by the National Center for Advancing Translational Sciences, National Institutes of Health, through Grant Award Number TL1TR001110.

## References

Arkin,M.R. and Wells,J.A. (2004) *Nat. Rev. Drug. Discov.*, **3**, 301–317.
Azoitei,M.L., *et al.* (2011) *Science*, **334**, 373–376.
Baird,L., Lleres,D., Swift,S. and Dinkova-Kostova,A.T. (2013) *Proc. Natl. Acad. Sci.*, **110**, 15259–15264.
Batori,V., Koide,A. and Koide,S. (2002) *Protein. Eng., Design Selection*, **15**, 1015–1020.
Chen,Y., Inoyama,D., Kong,A.N.T., Beamer,L.J. and Hu,L. (2011) *Chem. Biol. Drug. Des.*, **78**, 1014–1021.
DeLano,W.L. (2002) *Curr. Opin. Struct. Biol.*, **12**, 14–20.
DeNicola,G.M., *et al.* (2011) *Nature*, **475**, 106–109.
Fellouse,F.A., *et al.* (2007) *J. Mol. Biol.*, **373**, 924–940.
Fleishman,S.J., Whitehead,T.A., Ekiert,D.C., Dreyfus,C., Corn,J.E., Strauch,E. M., Wilson,I.A. and Baker,D. (2011) *Science*, **332**, 816–821.
Gilbreth,R.N., Esaki,K., Koide,A., Sidhu,S.S. and Koide,S. (2008) *J. Mol. Biol.*, **381**, 407–418.
Hancock,R., Schaap,M., Pfister,H. and Wells,G. (2013) *Org. Biomol. Chem.*, **11**, 3553–3557.
Hast,B.E., Goldfarb,D., Mulvaney,K.M., Hast,M.A., Siesser,P.F., Yan,F., Hayes,D.N. and Major,M.B. (2013) *Cancer. Res.*, **73**, 2199–2210.
Hast,B.E., *et al.* (2014) *Cancer. Res.*, **74**, 808–817.
Hayes,J.D., McMahon,M., Chowdhry,S. and Dinkova-Kostova,A.T. (2010) *Antioxid. Redox Signal.*, **13**, 1713–1748.
Inoyama,D., Chen,Y., Huang,X., Beamer,L.J., Kong,A.N.T. and Hu,L. (2012) *J. Biomol. Screen.*, **17**, 435–447.
Koide,S., Koide,A. and Lipovšek,D. (2012) *Protein Engineering for Therapeutics, Part B*, Academic Press, Waltham, Massachusetts, USA. pp. 135–156.
Leahy,D.J., Aukhil,I. and Erickson,H.P. (1996) *Cell*, **84**, 155–164.

Lewis,S.M. and Kuhlman,B.A. (2011) *PLoS ONE*, **6**, e20872.
Lo,S.C., Li,X., Henzl,M.T., Beamer,L.J. and Hannink,M. (2006) *EMBO J.*, **25**, 3605–3617.
Lungu,O.I., Hallett,R.A., Choi,E.J., Aiken,M.J., Hahn,K.M. and Kuhlman,B. (2012) *Chem. Biol.*, **19**, 507–517.
Magesh,S., Chen,Y. and Hu,L. (2012) *Med. Res. Rev.*, **32**, 687–726.
Mandell,D.J., Coutsias,E.A. and Kortemme,T. (2009) *Nat. Meth.*, **6**, 551–552.
Moharregh-Khiabani,D., Linker,R., Gold,R. and Stangel,M. (2009) *Curr. Neuropharmacol*, **7**, 60–64.
Moreland,L.W. (1998) *Rheum. Dis. Clin. North Am.*, **24**, 579–591.
Nguyen,T., Nioi,P. and Pickett,C.B. (2009) *J. Biol. Chem.*, **284**, 13291–13295.
Nikolovska-Coleska,Z., *et al.* (2004) *Anal. Biochem.*, **332**, 261–273.
Ofek,G., Guenaga,F.J., Schief,W.R., Skinner,J., Baker,D., Wyatt,R. and Kwong,P.D. (2010) *Proc. Natl. Acad. Sci.*, **107**, 17880–17887.
Ohta,T., *et al.* (2008) *Cancer. Res.*, **68**, 1303–1309.
Peppel,K. (1991) *J. Exp. Med.*, **174**, 1483–1489.
Purbeck,C., Eletr,Z.M. and Kuhlman,B. (2010) *Biochemistry*, **49**, 1361–1363.
Richards,J., Miller,M., Abend,J., Koide,A., Koide,S. and Dewhurst,S. (2003) *J. Mol. Biol.*, **326**, 1475–1488.
Schön,A., Lam,S.Y. and Freire,E. (2011) *Future Med. Chem.*, **3**, 1129–1137.
Steel,R., Cowan,J., Payerne,E., O'Connell,M.A. and Searcey,M. (2012) *ACS. Med. Chem. Lett.*, **3**, 407–410.
Stranges,P.B., Machius,M., Miley,M.J., Tripathy,A. and Kuhlman,B. (2011) *Proc. Natl. Acad. Sci.*, **108**, 20562–20567.
Taguchi,K., Motohashi,H. and Yamamoto,M. (2011) *Genes to Cells*, **16**, 123–140.
Tong,K.I., Katoh,Y., Kusunoki,H., Itoh,K., Tanaka,T. and Yamamoto,M. (2006) *Mol. Cell. Biol.*, **26**, 2887–2900.
Wang,C., Bradley,P. and Baker,D. (2007) *J. Mol. Biol.*, **373**, 503–519.
Wojcik,J., *et al.* (2010) *Nat. Struct. Mol. Biol.*, **17**, 519–527.
Zhang,D.D. and Hannink,M. (2003) *Mol. Cell. Biol.*, **23**, 8137–8151.