nature
structural &
molecular biology

© 2004 Nature Publishing Group http://www.nature.com/natstructmolbiol

Computational redesign of protein-protein interaction specificity

Tanja Kortemme[1,4], Lukasz A Joachimiak[1,2,4], Alex N Bullock[1,3,4], Aaron D Schuler[2], Barry L Stoddard[2] & David Baker[1]

We developed a 'computational second-site suppressor' strategy to redesign specificity at a protein-protein interface and applied it to create new specifically interacting DNase-inhibitor protein pairs. We demonstrate that the designed switch in specificity holds in *in vitro* binding and functional assays. We also show that the designed interfaces are specific in the natural functional context in living cells, and present the first high-resolution X-ray crystallographic analysis of a computer-redesigned functional protein-protein interface with altered specificity. The approach should be applicable to the design of interacting protein pairs with novel specificities for delineating and re-engineering protein interaction networks in living cells.

Protein-mediated interactions in biological systems are used to organize the macromolecular complexes and networks responsible for regulation and complexity. Tools to alter and interfere rationally with protein interactions offer great promise for understanding and delineating these networks, but require a predictive description of the physical basis of affinity and specificity in protein interfaces. Simple general rules to identify protein recognition sites and predict energetic hotspots in protein complexes often fail[1], largely because of the extreme diversity in shape and chemical character of protein-protein interfaces[2]. Although physical models have been successfully used to rationalize energetically important interactions in protein-protein interfaces[3–7], a major challenge in understanding protein recognition is to develop computational methods that can capture the molecular basis of specificity: how do proteins discriminate their natural binding partners from many other possible ligands with similar sequences and structures?

Several computational methods to predict interaction specificity have been developed recently based on *in vitro* binding data on specific systems[8,9], evolutionary information[10] and empirical energy functions[11,12]. Complementing and extending these approaches, a stringent test of current understanding is the design of macromolecules with desired properties and novel specificities. There have been great successes in the computational design of monomeric proteins such as protein cores[13–15] (reviewed in ref. 16), metal-binding sites[17], complete proteins[18], folding mechanisms[19] and new topologies[20,21]. These rotamer search–based methods have been applied to the design of new protein-small molecule[22], coiled-coil[23,24] and protein-peptide interfaces[25,26] with enhanced[25,27] and novel specificities[24,26].

For interface design to become useful in manipulating complex protein-protein interaction networks *in vivo*, several conditions must be met. First, the methods must be applicable to large, often plastic protein-protein interfaces with nonlinear epitopes and conformationally coupled interactions between buried side chains, where the determinants of specificity are not obvious from simple geometric criteria. Second, structural information on the designed models is required to assess and improve the quality of the computational methodology. Third, the interface redesign must bring about a switch in specificity not only *in vitro* but also in the relevant biological process in the cellular context.

Here we describe the computational design and structural verification of new interacting protein-protein pairs that are functional and specific *in vivo* and *in vitro*. We introduce a computational second-site suppressor approach for protein-protein interface design that automatically identifies specificity-defining interface mutations by screening for disruptive interface mutations in one partner that can be compensated by alterations in the interacting second partner. The structural and functional analyses of the designed complexes identify interactions that can and cannot be modeled well using the current methodology.

RESULTS

The model system

We chose as our experimental system the complex between a bacterial nonspecific DNase (colicin E7) and its tightly bound inhibitor (immunity protein Im7). This protein-protein interaction is an attractive model for studying recognition specificity for several reasons: (i) the crystal structure of the colicin E7/Im7 complex has been

[1]Howard Hughes Medical Institute & Department of Biochemistry, Box 357350, University of Washington, Seattle, Washington 98195-7350, USA. [2]Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue N. A3-023, Seattle, Washington 98109, USA. [3]Present address: The Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK. [4]These authors contributed equally to this work. Correspondence should be addressed to D.B. (dabaker@u.washington.edu).

Published online 21 March 2004; doi:10.1038/nsmb749

© 2004 Nature Publishing Group http://www.nature.com/natstructmolbiol

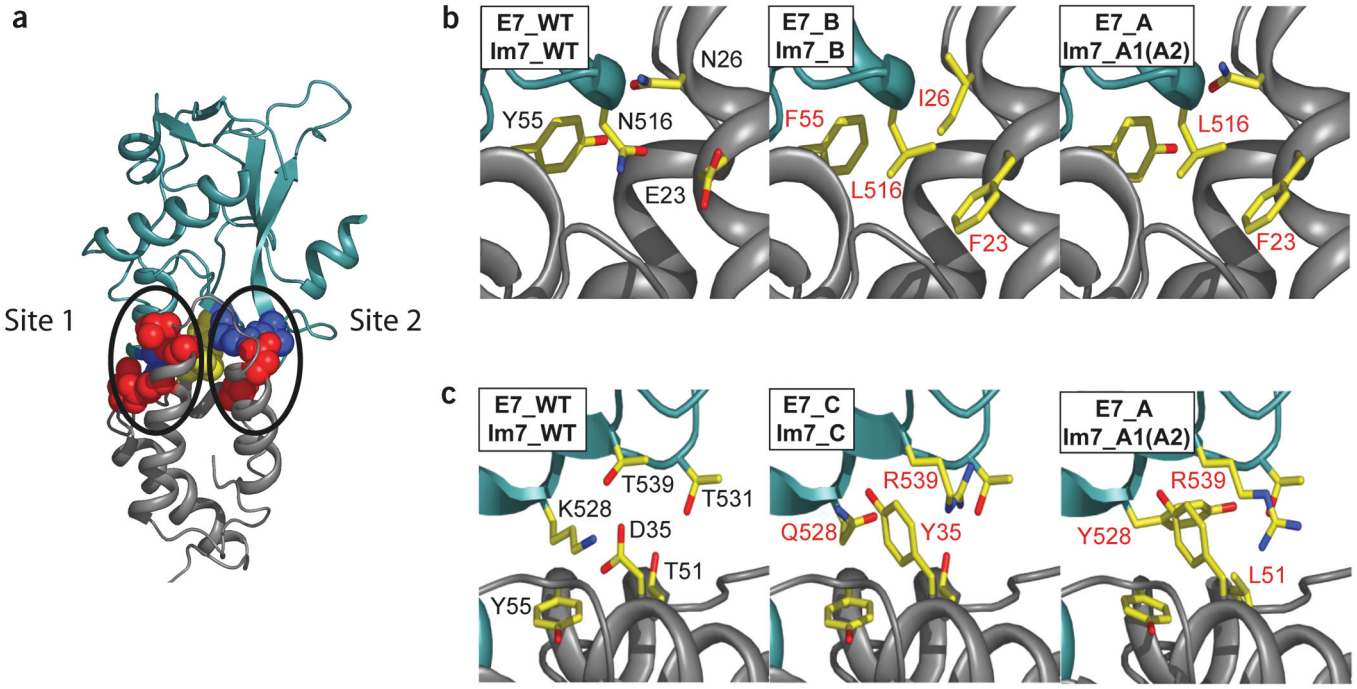

Figure 1 The DNase-immunity protein model system. The DNase polypeptide backbone is teal, the immunity protein backbone gray. (**a**) The wild-type E7 DNase–Im7 immunity protein complex, illustrating the design strategy. Residues participating in interactions at sites 1 and 2 are shown in space-filling representation (red, immunity protein; blue, DNase). The conserved tyrosine-tyrosine immunity protein motif in the center of the interface separating the two sites is yellow. (**b**,**c**) Comparison of the wild-type interface with structural models of the designed complexes illustrates a 'polarity switch' in site 1 (**b**) and a 'steric switch' in site 2 (**c**). (**b**) The polar interaction network around the wild-type Asn516 on the DNase is completely (E7_B/Im7_B) or partially (E7_A/Im7_A1 and E7_A/Im7_A2) replaced by hydrophobic residues, with similar sterics. (**c**) In the E7_C/Im7_C design, the wild-type interaction between Lys528 on the DNase and Asp35 on the immunity protein is substituted by a Gln528-Tyr35 pair. Tyr35 protrudes into the interface area, replacing the smaller wild-type aspartic acid residue to form a predicted hydrogen bond with the designed glutamine at position 528. Thr539, which makes a hydrogen bond with Lys528 in the wild-type structure, is now replaced by an arginine residue predicted to pack against and stabilize the designed tyrosine at position 35. E7_A/Im7_A1(A2) adopts a different packing arrangement at site 2: position 528 is now occupied by a tyrosine that seems too large to extend into the wild-type pocket. Instead, Tyr528 adopts a substantially different χ_1 angle and is now predicted to form intermolecular interactions to Tyr35 (packing) and Leu51 (packing), in addition to an intramolecular hydrogen bond with Thr531. Side chains in the designed regions are yellow and are labeled in red if the amino acid was mutated in the design in **b** and **c**. Figures were made using PyMOL (Delano Scientific).

solved to a resolution of 2.3 Å (ref. 28); (ii) the E7/Im7 pair belongs to a family of interacting bacterial proteins that are remarkably specific for their cognate partner proteins, and bind with high affinity (reviewed in ref. 28); (iii) the biological function of colicin proteins, which are cytotoxic in the absence of an intracellular cognate immunity protein, provides an assay to test altered specificities *in vivo*; two DNase-immunity protein pairs are functional and specific if the immunity proteins protect against cell killing by their cognate DNase but not by the noncognate DNase; (iv) interface substitutions are largely unrestricted as the binding interface is distant and distinct from the active site; (v) colicins with altered specificity may potentially be used as antibacterial agents[29,30].

The computational design strategy

We aimed to redesign the DNase-immunity protein interface to create new pairs of specifically interacting partners that interact more strongly with each other than with their wild-type counterparts. Optimizing all interface residues might result in a stable interface, but any change in specificity would be fortuitous. Such a procedure would probably produce sequences identical or close to the wild-type interface, as the high affinity of the colicin DNase–immunity protein complexes suggests a highly optimized interface. In fact, the assumption that on average sequences of proteins and protein interfaces are optimized, given their structures, has been used to parameterize and test energy functions used in computational protein design[31–33].

Our design strategy to create protein-protein pairs with new specificity thus required elements of positive and negative design: a computational second-site suppressor design strategy was developed to identify sequence perturbations in partner I in the complex that would destabilize the interface but could be compensated by redesigning the interacting residues on partner II. We applied the following general protocol: (i) interface residues for the initial perturbation in partner I were selected that are substantially buried in the interface and form side chain–side chain contacts across the interface. (ii) Computational screening was used to identify mutations at these selected positions (by making all possible single mutations in partner I and redesigning interacting residues on partner II) that maximized the computed free energy difference between the mutated partner I–wild-type partner II interface and the mutated partner I–redesigned partner II interface; (iii) amino acid changes at interface residues in both partners that destabilized the original complex and stabilized the designed complex were combined into complete redesigns comprising the whole interface, and the side chain conformations of all combinations were optimized; (iv) the binding free energies of the optimized combinations were computed, and the sequences with the highest predicted affinities were selected.

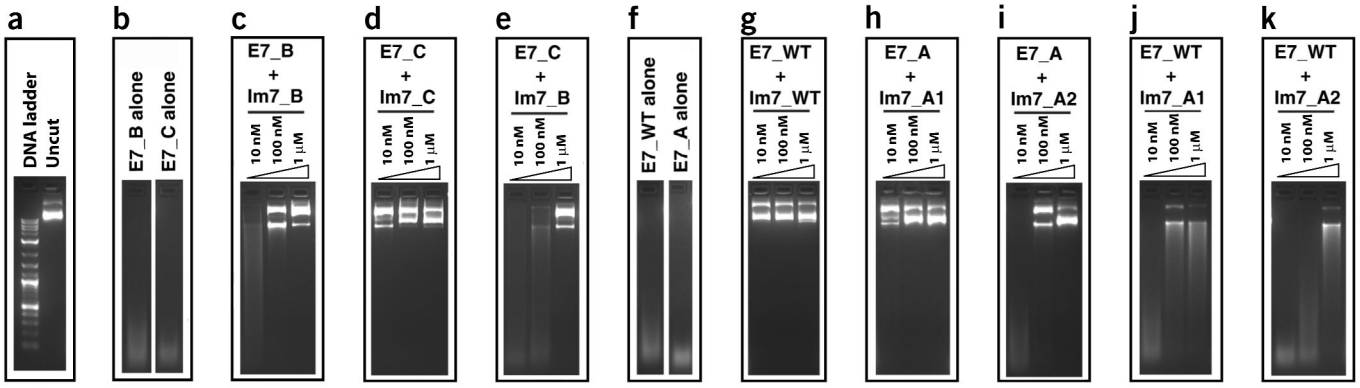

Figure 2 *In vitro* DNase activity assay. (**a**) 2-log DNA ladder (NEB) and uncut DNA. (**b–k**) Effective DNase inhibition by cognate E7_B/Im7_B (**c**), E7_C/Im7_C (**d**), E7_WT/Im7_WT (**g**), E7_A/Im7_A1 (**h**) and E7_A/Im7_A2 (**i**) complexes, but poor inhibition by the noncognate E7_C/Im7_B (**e**) and E7_WT/Im7_A1 (**j**) and E7_WT/Im7_A1 (**k**) complexes. Protection is indicated by the presence of undigested DNA with low mobility on the gel. (**b,f**) Control reactions with the different DNases but no immunity protein.

The mutations selected for the initial sequence perturbations were N516L, K528Q and K528Y in the DNase. These residues form two key interaction networks ('site 1' and 'site 2') that pack against either side of the central tyrosine-tyrosine motif that is strictly conserved in the naturally occurring immunity proteins (**Fig. 1a**). The redesign to compensate for the changes in the DNase suggested mutations at nine positions in the immunity protein; these mutations showed some sequence variation (**Table 1**, column 3) depending on how many residues near the initial perturbation were allowed to mutate (in the immunity protein) or change side chain rotamer (in the DNase). An enumeration of all combinations and ranking by binding free energy yielded the final designs (**Table 1**, columns 5–8). The final designed top-scoring sequences E7_A/Im7_A1 and E7_A/Im7_A2 differ at position 63 but were nearly equivalent in the design calculations. (All complex pairs are denoted with the DNase first and the immunity protein second; the wild-type complex is denoted E7_WT/Im7_WT below for consistency).

Table 1 Designed sequences

Position	Wild type	Favorable amino acids[a]	Best amino acids[b]	Designed sequences[c]			
				E7_A	E7_A	E7_B	E7_C
516	**N**	**L**[d]	**L**[d]	**L**	**L**	**L**	
528	**K**	**Q,Y**[d]	**Q,Y**[d]	**Y**	**Y**		**Q**
539[e]	T	K,R,F	K or R[d]	R	R		R
				Im7_A1	Im7_A2	Im7_B	Im7_C
23	E	F,Y	F or Y[f]	F	F	F	
26	N	I,Q,L,N	I or N[g]			I	
27	V	K,T	T	T	T		
31	D	S,N,D	D				
35	D	L,Y,W	Y	Y	Y		Y
51	T	P,I,L,R	L	L	L		
55	Y	F,Y	F or Y[g]			F	
56	Y	Y,W	Y or W[f]	W	W		
63	D	D,N	D or N[f]		N		

[a]Favorable amino acids in initial screening design runs. [b]Best amino acids from design runs ranked by binding energy. [c]Empty fields indicate wild-type residues. [d]Fixed in the designs as initial sequence perturbation of the DNase (bold). [e]As the wild-type Lys528 forms an intramolecular interaction with Thr539, this residue was redesigned together with Lys528. [f]Both choices of amino acids were essentially equivalent in the ranking by binding energy, and independent of the sequence choices at the other positions. [g]Covariation for positions 26 and 55: either Asn26-Tyr55 or Ile26-Phe55.

In addition to the extensive redesign of the entire interface in the E7_A/Im7_A1 and E7_A/Im7_A2 complexes, we also created two additional redesigned protein pairs (**Table 1**, columns 7 and 8) in which only one of the two sites described in the previous paragraph (**Fig. 1a**) was redesigned. In the first pair (redesign E7_B/Im7_B), a polar interaction network in site 1 was replaced by hydrophobic interactions (a 'polarity switch'). In the second pair (redesign E7_C/Im7_C), the polarity of the interface at site 2 was maintained but the steric packing of side chains was altered (a 'steric switch'). Comparison of the wild-type and structural models of the designed proteins (**Fig. 1**), shows the interaction networks formed around the initial sequence positions chosen for the sequence perturbations, the polarity switch around N516L (site 1, **Fig. 1b**) and the steric switch around K528Q or K528Y (site 2, **Fig 1c**).

Binding analysis of designed complexes

To test the designed specificity changes between the cognate and noncognate DNase/immunity protein complexes, we used an *in vitro* endonuclease digestion assay. We first investigated the extent of specificity between the 'polarity switch' at the redesigned site 1 (E7_B/Im7_B) and the 'steric switch' at the redesigned site 2 (E7_C/Im7_C). Both the Im7_B and Im7_C designed immunity proteins protect DNA from degradation by their cognate E7_B and E7_C DNases, respectively (**Fig. 2c,d**). Notably, the Im7_B is less effective in protecting against cleavage by the noncognate E7_C DNase (**Fig. 2e**). The redesigned proteins thus exhibit functional specificity. A similar specificity in the *in vitro* endonuclease digestion assay is observed for the E7_A/Im7_A1 and E7_A/Im7_A2 complexes, which combine a polarity switch in site 1 and a steric switch in site 2 (**Table 1** and **Fig. 1b,c**): the Im7_A1 and Im7_A2 immunity proteins protect more strongly against degradation by the E7_A DNase (**Fig. 2h,i**) than by the E7_WT DNase (**Fig. 2j,k**).

We then used surface plasmon resonance (SPR) measurements to compare the *in vitro* binding affinities of the cognate and non-

© 2004 Nature Publishing Group http://www.nature.com/natstructmolbiol

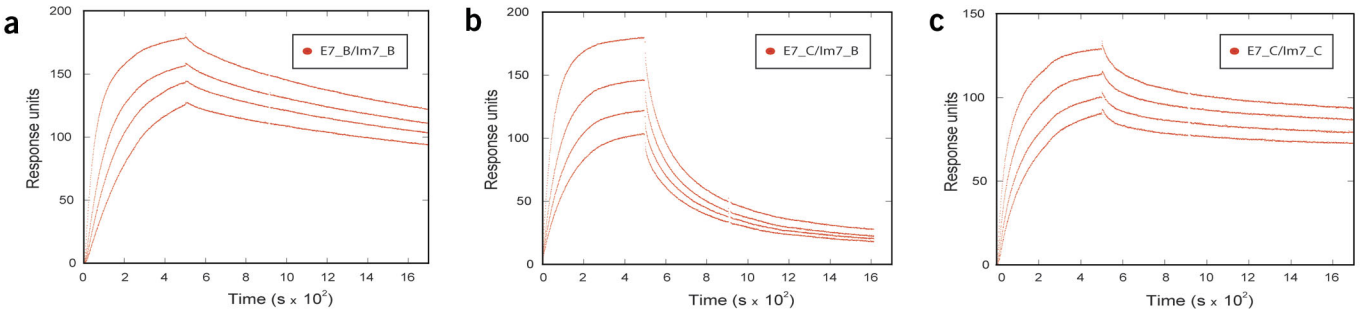

Figure 3 SPR sensograms. (**a**–**c**) A comparison of SPR sensograms for E7_B/Im7_B (**a**), E7_C/Im7_B (**b**) and E7_C/Im7_C (**c**) at concentrations of 100, 50, 30 and 20 nM immunity protein, demonstrating tight binding for the cognate E7_B/Im7_B and E7_C/Im7_C complexes and weaker binding for the noncognate E7_C/Im7_B complex.

© 2004 Nature Publishing Group http://www.nature.com/natstructmolbiol

cognate DNase-immunity protein pairs. SPR sensograms (**Fig. 3**) comparing the behavior of the cognate complexes E7_B/Im7_B and E7_C/Im7_C to the noncognate E7_C/Im7_B complex (a similar behavior was observed for the cognate E7_A/Im7_A2 and noncognate E7_A/Im7_WT complexes; see **Supplementary Fig. 1** online) confirm the results obtained in the DNase digestion assay (**Fig. 2**). The apparent association rate constants of the individual site design complexes are similar, ranging from 1×10^5 M^{-1} s^{-1} to 1.2×10^5 M^{-1} s^{-1} (**Table 2**; see Methods and **Supplementary Fig. 2** online for details of the SPR analysis). However, there are substantial differences in the dissociation rate; the noncognate E7_C/Im7_B complex dissociates much faster than the cognate E7_B/Im7_B and E7_C/Im7_C protein pairs (**Fig. 3**). The dissociation rates of the cognate complexes were too slow to measure accurately with SPR, as a substantial amount of the protein remained bound at the end of the experiment, whereas nearly complete dissociation was observed for the noncognate complex. Although the very slow cognate dissociation rates prevent an accurate quantitative analysis of the binding energy differences, the data suggest a substantial difference between the E7_C/Im7_B binding affinity and the cognate complex affinities.

Tryptophan residues in the designed Im7_A1 and Im7_A2 inhibitor proteins allow binding to be measured by intrinsic fluorescence for the designs combining the polarity switch in site 1 and the steric switch in site 2 (**Fig. 4a** and **Supplementary Fig. 3** online), and yielded an independent set of rate constants for comparison with the SPR data (**Table 2**). As observed above, the majority of the affinity differences of cognate and noncognate colicin complexes are manifested in the dissociation rates (**Fig. 4**), as the apparent association rate constants are very similar, ranging from 1.4×10^5 M^{-1} s^{-1} to 3.6×10^5 M^{-1} s^{-1} in the SPR experiments and from 1×10^5 M^{-1} s^{-1} to 4.3×10^5 M^{-1} s^{-1} in the fluorescence experiments (**Table 2**). For the very tightly binding wild-type complex, the dissociation rates were too slow to measure using SPR and fluorescence, and for the E7_A/Im7_A1 interaction they were too slow to measure by SPR (although fitting the fluorescence dissociation traces yielded an estimate for the apparent dissociation rate constant of 6.2×10^{-5} s^{-1}, **Fig. 4b**; for details on the fitting see Methods and **Fig. 4a**). However, the Im7_A2 variant with an additional D63N mutation on the immunity protein showed faster dissociation rates in both the cognate E7_A/Im7_A2 and noncognate E7_WT/Im7_A2 complexes that were measurable by both SPR and fluorescence (**Table 2**, **Fig. 4c**, **Supplementary Fig. 1** online). A consistent and significant specificity switch was found using both methods (**Fig. 4d**, **Table 2**), with a difference in dissociation rate between the cognate and noncognate complexes E7_A/Im7_A2 and E7_WT/Im7_A2 of 17-fold and 12-fold in the fluorescence and SPR experiments, respectively. The cognate E7_A/Im7_A1 and the noncognate E7_WT/Im7_A1 complexes show a slightly larger speci-

Table 2 Kinetic data for the designed cognate and noncognate complexes

		$k_{app,on}$ (M^{-1} s^{-1}) ($\times 10^5$)	$k_{app,off}$ (s^{-1})	$K_{app,d}$ (nM)
E7_C/Im7_C, E7_B/Im7_B and E7_C/Im7_B binding pairs				
E7_C/Im7_C	SPR	1.20 ± 0.30 (5)	–	–
E7_B/Im7_B	SPR	1.01 ± 0.22 (4)	–	–
E7_C/Im7_B	SPR	1.05 ± 0.18 (5)	0.0034 ± 0.0013 (30)	32
E7_WT/Im7_WT(Y56W)[a], E7_WT/Im7_A1 and E7_A/Im7_A1 binding pairs				
E7_WT	SPR	3.65 (1)	–	–
Im7_WT(Y56W)[a]	Fluorescence	3.70 ± 0.063 (3)	–	–
E7_WT/Im7_A1	SPR	3.10 (1)	0.001800 ± 0.000100 (4)	5.8
	Fluorescence	4.30 ± 0.300 (3)	0.002500 ± 0.000138 (3)	5.8
E7_A/Im7_A1	SPR	1.82 (1)	–	–
	Fluorescence	–	0.000062 ± 0.000004 (2)	0.34[b]
E7_WT/Im7_WT, E7_WT/Im7_A2 and E7_A/Im7_A2 protein binding pairs				
E7_WT/Im7_WT	SPR	1.37 ± 0.40 (2)	–	–
	Fluorescence	–	–	–
E7_WT/Im7_A2	SPR	1.70 ± 0.25 (4)	0.0200 ± 0.00100 (16)	117
	Fluorescence	0.95 ± 0.10 (6)	0.0470 ± 0.00900 (5)	495
E7_A/Im7_A2	SPR	1.92 ± 0.83 (7)	0.0017 ± 0.00070 (28)	9
	Fluorescence	–	0.0027 ± 0.00015 (3)	14[b]

Where applicable, an apparent $K_{app,d}$ is noted given by the ratio of the apparent dissociation and association rate constants $k_{app,off}$ and $k_{app,on}$. SPR and fluorescence experiments were carried out in 50 mM MOPS, pH 7.5, but with different amounts of salt (200 mM NaCl and 600 mM NaCl, or 200 mM NaCl and 400 mM NaCl). Standard deviations are reported along with the number of independent measurements in parentheses for the $k_{app,on}$ and $k_{app,off}$. For association, each independent experiment includes traces at four or five different concentrations. [a]E7_WT/Im7_WT(Y56W) refers to mutant WT immunity protein containing a tryptophan at position 56. [b]Apparent $K_{app,d}$ calculated using the combination of SPR and fluorescence. The affinity of the interaction between the E7_WT and its cognate immunity protein has not been reported but the affinity of the similar E9_WT/Im9_WT complex has been measured at 10^{-14}–10^{-16} M (ref. 23).

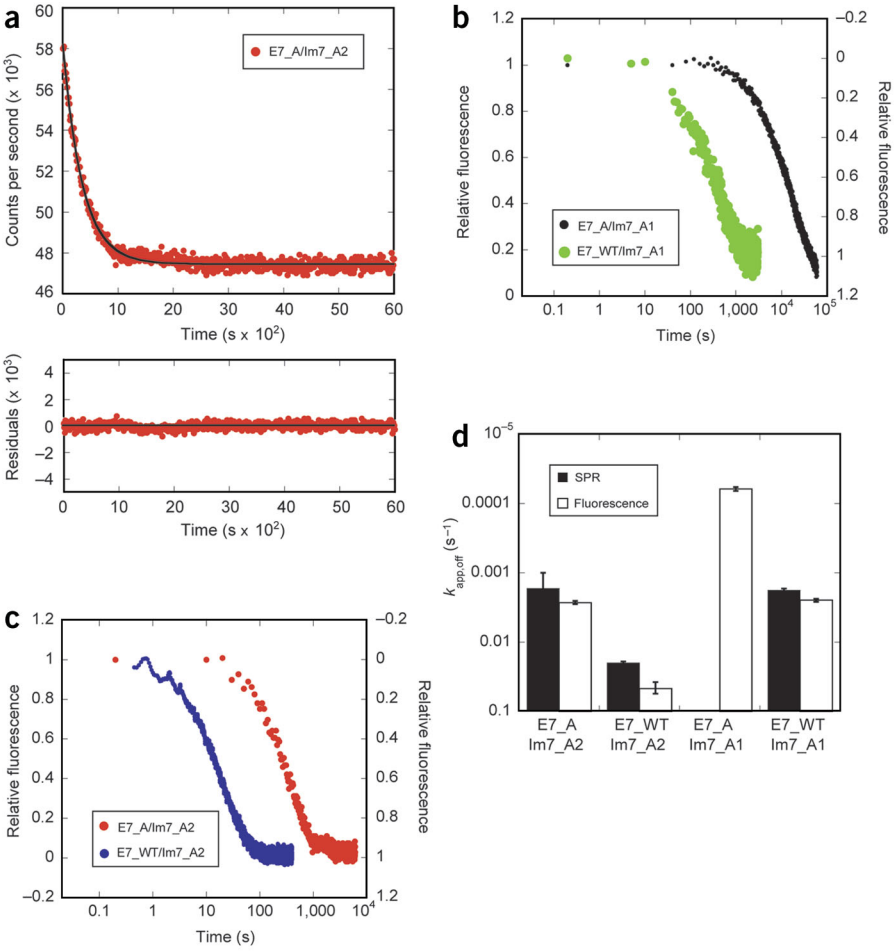

Figure 4 Intrinsic fluorescence. Dissociation rates measured by intrinsic fluorescence demonstrate a specificity switch between cognate and noncognate interactions for the E7_WT/Im7_WT, E7_A/Im7_A1 and E7_A/Im7_A2 complexes. The fluorescence change in binding is likely due to Trp56 introduced in the Im7_A1 and Im7_A2 immunity proteins, as the E7_WT/Im7_WT complex does not show a measurable tryptophan fluorescence change upon binding despite containing three tryptophan residues at positions 75, 464 and 500. (**a**) Example of a fluorescence trace monitoring the dissociation of the E7_A/Im7_A2 complex by competition (see Methods). The fit to a single-exponential rate equation is shown in black ($k_{app,off}$ = 0.0028 s^{-1}) and the residuals for this fit are below. (**b**) Dissociation traces for the E7_A/Im7_A1 cognate interaction (black) and the E7_WT/Im7_A1 noncognate interaction (green). (**c**) Dissociation traces for the E7_A/Im7_A2 cognate interaction (red) and the E7_WT/Im7_A2 noncognate interaction (blue). (**d**) Comparison of SPR (black) and fluorescence (white) apparent dissociation rate constants for cognate E7_A/Im7_A2 and E7_A/Im7_A1 complexes and noncognate E7_WT/Im7_A2 and E7_WT/Im7_A1 complexes. The single amino acid difference, D63N, between the Im7_A1 and Im7_A2 immunity proteins produces a substantial difference in affinity between the respective cognate complexes—the aspartate residue coordinates a water molecule involved in an extensive water network at the interface.

ficity switch with a substantial 40-fold difference between the apparent dissociation rate constants measured by fluorescence (**Fig. 4b**).

Whereas the E7_WT/Im7_A1 and E7_WT/Im7_A2 complexes show a clear specificity switch, the swapped complex E7_A/Im7_WT shows a very slow dissociation rate, similar to that of the wild-type E7_WT/Im7_WT complex, that is too slow to quantify (data not shown). This difference in behavior is not surprising as the noncognate E7_WT/Im7_A1 and E7_WT/Im7_A2 complexes each contain more mutations (five and six, respectively) than does the reverse E7_A/Im7_WT complex (three). Similarly, whereas the E7_C/Im7_B complex has five mutations compared with the E7_WT/Im7_WT complex and demonstrates the designed specificity switch, the E7_B/Im7_C complex has only two mutations and shows dissociation kinetics similar to those of the cognate complexes (data not shown).

In vivo function and specificity

A previous study[26] has used a two-hybrid assay to test *in vivo* binding of the designed molecules and demonstrated that the interactions are remarkably specific even in the context of a cell. Here we test directly whether the designed interface alters the specificity *in vivo* while preserving the biological function of the interaction. The *in vivo* assay of the specificity switch takes advantage of the natural function of the immunity protein to protect the cell from death by blocking the nuclease activity[34]. The equivalent experiment to the SPR-fluorescence measurements of E7_A/Im7_A1(A2) and E7_WT/Im7_A1(A2) binding is to compare the relative immunity of cells expressing Im7_A1 and Im7_A2 to E7_WT and E7_A toxin. Cells expressing the Im7_A1 or Im7_A2 immunity proteins are specifically protected against the action of the E7_A toxin, but remain susceptible to the E7_WT protein at higher concentrations (**Fig. 5a–c**). Thus, the designed Im7_A1 and Im7_A2 proteins not only are able to function in a cellular context, but also are remarkably specific for their designed partner protein. Very similar results were obtained with the individual site designs (**Fig. 5d–f**).

The rank orders of affinity differences of the cognate and noncognate complexes are consistent in the DNase digestion, SPR, fluorescence and *in vivo* assays. The apparent magnification of the *in vitro* specificity difference in the *in vivo* assay is consistent with findings that large changes in *in vivo* cytotoxic activity are associated with changes in *in vitro* binding affinity in the 10^{-10} to 10^{-7} M range (C. Kleanthous, University of York, York, UK) personal communication). There are also large uncertainties in the amount of toxin entering the cell (for example, 10–50 μM toxin may saturate the protein import machinery), the amount of immunity protein made in the cell and the amount of free toxin needed to kill the cell.

Structure of the E7_C/Im7_C complex

A comparison of the predicted and experimentally determined designed structures is essential to assess the accuracy of the design algorithm, to identify interactions that are poorly modeled in the algorithm and to suggest avenues for improvement. We obtained a crystal structure at a resolution of 2.1 Å of the E7_C/Im7_C complex (the E7_A/Im7_A1, E7_A/Im7_A2 and E7_B/Im7_B complexes did not crystallize under the conditions tested; a likely reason is that they all

© 2004 Nature Publishing Group http://www.nature.com/natstructmolbiol

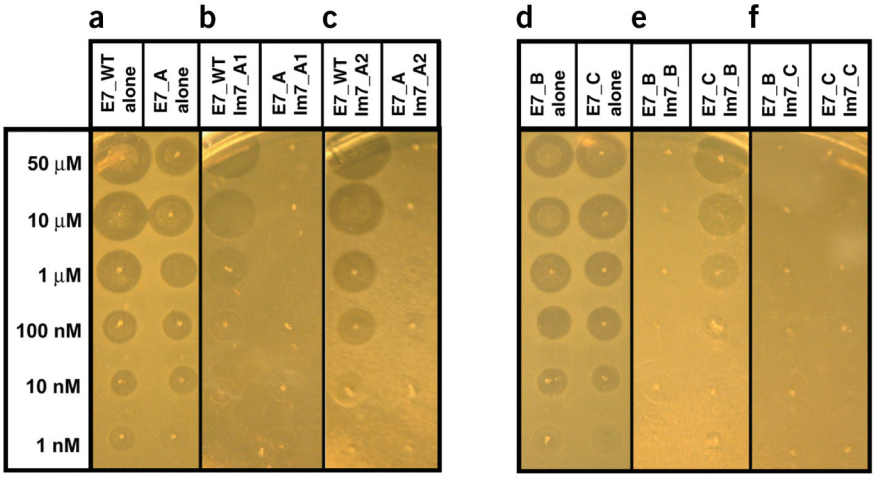

Figure 5 *In vivo* cell death assay. (**a**–**f**) Cells expressing Im7_A1 (**b**), Im7_A2 (**c**), Im7_B (**e**), Im7_C (**f**) or no immunity protein (**a**,**d**) as control were treated with different concentrations of E7_WT, E7_A, E7_B and E7_C. Clear zones appear when the cells are killed and demonstrate a lack of immunity protein protection against the imported DNase toxin. Im7_A1 and Im7_A2 can protect against E7_A but show only partial protection against E7_WT at low toxin concentration. Similar results are seen for the site designs B and C, with protection for the cognate E7_B/Im7_B and E7_C/Im7_C complexes, but cell death for the noncognate E7_C/Im7_B complex. The reciprocal noncognate complexes E7_A/Im7_WT (data not shown) and E7_B/Im7_C (**f**) show protection as they contain fewer mutations, as discussed in the text.

contain mutations in site 1 that are expected to alter crystal contacts). The electron density surrounding the redesigned residues is well defined (**Fig. 6a**). An overlay of the designed interface region in the experimental structure of the E7_C/Im7_C protein complex with the computational model shows good agreement (**Fig. 6b**): the all-atom r.m.s. deviation is 0.62 Å for all interface residues (defined as residues having at least one atom within 4 Å of at least one other atom on the interface partner). The protein backbone that was left unchanged in all design calculations has a Cα backbone r.m.s. deviation between the structure and the designed model of 0.5 Å (excluding the partly disordered loop, residues 465–473, which are distant from the interface region).

A detailed view of the side chain conformations and interactions of the three interface mutations, K528Q, T539R and D35Y (**Table 1**), in the E7_C/Im7_C complex identifies the hydrogen bond network (**Fig. 6c**). Notably, in the structure, the intermolecular hydrogen bond between Tyr35 and Gln528 is almost exactly as it is in the design model. The conformations of Tyr35 and Gln528 are predicted with great accuracy (χ angle deviations <16°), as is the geometry of the hydrogen bond between them. The surrounding residues in a 6-Å radius have φ, ψ and χ angles similar to those of the wild-type structure, suggesting that the introduced mutations strictly affect the targeted interactions. Previous computational protein designs that were structurally verified[14,16,18,20] have largely relied on hydrophobic packing interactions in the core[16]. The successful design of a new hydrogen bond between side chains in a buried protein-protein interface environment and the close structural agreement of the designed and experimentally observed interaction additionally validate our orientation-dependent hydrogen bonding potential[32]; this is important because of the considerable uncertainties in modeling polar and electrostatic interactions in proteins.

Differences between the design model and the actual structure are in large part due to bound water molecules not modeled in the design process. The predicted conformation of Arg539 in the model is different from the experimentally observed conformation because of a water molecule in the X-ray structure that bridges the guanidinium group of Arg539 and the backbone carbonyl of Tyr35; in the E7_C/Im7_C model (**Fig. 6c**, right) the water-bridged interaction is replaced by a direct hydrogen bond between the Arg539 guanidinium group and the Thr51 hydroxyl group (the crystal structure rotamer of Arg539 also has an unusual χ_1/χ_2 combination not present in the backbone-dependent rotamer library used here[27]). In addition to the water molecule coordinating Arg539, the X-ray structure of the E7_C/Im7_C complex reveals the presence of

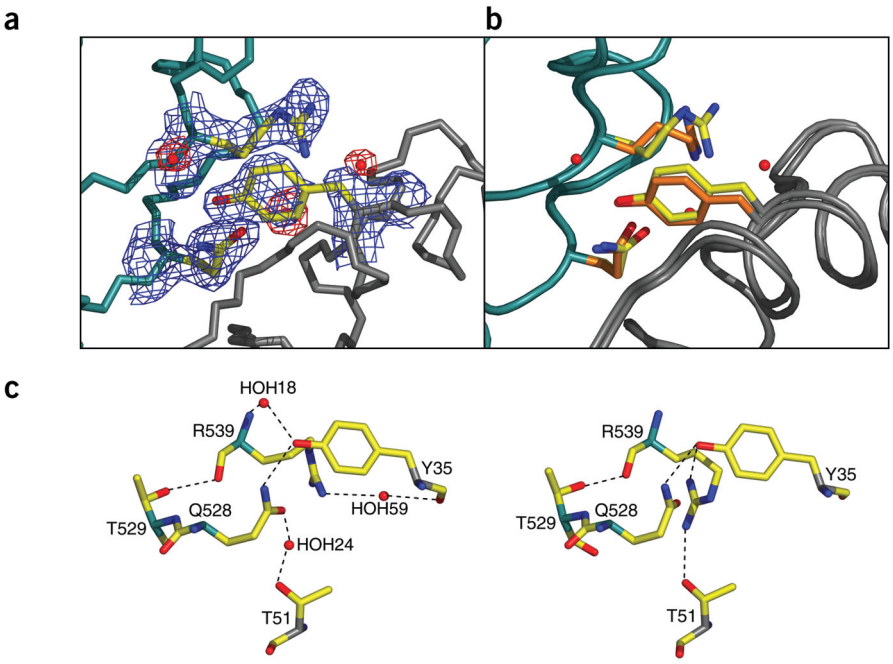

Figure 6 The crystal structure of the E7_C/Im7_C complex. The DNase backbone is teal, the immunity protein gray. (**a**) $2F_o - F_c$ density around the designed residues is contoured at 1.3 σ (blue) and the density around the waters is contoured at a 1 σ (red). The *B*-factors for the displayed water molecules are 29 Å^2 for water 24; 39 Å^2 for water 18; and 43 Å^2 for water 59 (the average *B*-factor for the waters in the structure is 36.85 Å^2). (**b**) Overlay of the model (orange side chains) with the experimentally determined structure (yellow side chains). (**c**) Hydrogen bonding patterns in the experimentally determined structure (left) and the designed model (right). DNase Cα carbons are teal, immunity Cα carbons are gray.

© 2004 Nature Publishing Group http://www.nature.com/natstructmolbiol

two other water molecules that bridge interactions across the interface in the redesigned region (**Fig. 6**). These observations suggest the importance of water-mediated interactions and the need to understand their role in stabilizing the interface (see below).

DISCUSSION

Four different colicin DNase-immunity proteins have been described that discriminate their cognate partners from noncognate homologs with great specificity[35]. Kleanthous and co-workers propose a 'dual-recognition mechanism' by which the immunity proteins discriminate among their target DNases: a conserved 'affinity site' in the immunity protein (Thr51, Asp52, Tyr55 and Tyr56 in WT_Im7) provides the majority of the binding affinity, and a variable 'specificity site' involving helix II (residues 34–45 in WT_Im7) mediates specificity[35,36]. The computational design of E7_A/Im7_A1 and E7_A/Im7_A2 is an additional solution to the specificity problem (see **Supplementary Fig. 4** online). It follows the dual-recognition strategy in that three of the four residues in the affinity site are unchanged or undergo conservative changes (Y56W), but differs in that specificity is achieved by changes that are not restricted to the specificity site. For example, position 516, a conserved polar residue in all naturally occurring sequences, is changed to a nonpolar residue in the design of site 1 (**Table 1**) and is compensated by mutation of the contacting residues, Glu23 and Asn26, which are located outside of the specificity site, to amino acids not found in the naturally occurring homologs. The structure of the E7_C/Im7_C complex suggests a solution to specific recognition at site 2, replacing the wild-type Asp35-Lys528 interaction with a hydrogen bond between the designed Tyr35 and Gln528 pair that formed as predicted (**Fig. 6**). Although position 35 is located in the immunity protein specificity site, the designed tyrosine residue is not found in any of the naturally occurring sequences. The differences between our designs and the naturally occurring variants suggest that evolution has sampled only a subset of the possible routes to specificity in this system (**Supplementary Fig. 4** online).

The biochemical and structural analyses of the designed protein pairs provide valuable data on the successes and failures of the methodology. We succeeded in designing functionally specific new protein pairs, and the experimental structure of the E7_C/Im7_C complex confirms the validity of the structural models created by the computational method (**Fig. 6**). The newly designed E7_A/Im7_A1 and E7_A/Im7_A2 complexes contain eight and nine mutations in the interface and show subnanomolar and low nanomolar binding, respectively; they have over an order of magnitude higher affinity that the noncognate E7_WT/Im7_A1 and E7_WT/Im7_A2 complexes. However, the designed cognate pairs have substantially lower affinity than the E7_WT/Im7_WT complex, and the experimentally measured affinity differences between the cognate and noncognate complexes are several orders of magnitude lower than the specificity differences observed between homologous members of the colicin family. The magnitude of these differences was not predicted by the design calculations, and suggests the important role of interactions mediated by interfacial water molecules[2] that are abundant in the colicin interfaces and are not explicitly modeled by the current energy function. This notion is supported by the presence of a new interfacial water network in the designed site in the E7_C/Im7_C complex (**Fig. 6c**) as well as conserved interfacial water networks in the structures of the E7_WT/Im7_WT, E9_WT/Im9_WT and E7_C/Im7_C complexes. Moreover, the observed destabilization of the E7_A/Im7_A1 complex by the D63N mutation was not predicted by the design energy function, and is probably due to a perturbation of the highly organized water network around Asp63 in the wild-type structure.

Interfacial water molecules increase the shape complementarity by filling gaps between imperfectly packed regions but also mediate many hydrogen bonds with the backbone and side chain polar groups or other bridging water molecules[2]. Our structural and biochemical analyses thus highlight the need for explicit modeling of water. Computationally, one way to achieve this is to model waters along with amino acid rotamers that are capable of forming hydrogen bonds. Including explicit water molecules in the simulations should substantially improve the correct prediction of interaction energies[37] and specificities that more closely match the naturally occurring specificities. Moreover, the structures of the homologous colicin cognate complexes E7_WT/Im7_WT and E9_WT/Im9_WT are related by a 19° rigid-body rotation of one of the partners versus the other[36]. Carrying out the design calculations using a family of complex backbone templates generated by small rigid-body perturbations should improve the predictive power of the methodology, and has been successful in the design of proteins that bind small molecules[22].

Conclusions

Despite the difficulties encountered in the modeling of protein-protein interaction specificity, our computational second-site suppressor redesign strategy, in conjunction with the current simple energy model including an explicit orientation-dependent hydrogen bonding potential, predicts new Im7 immunity protein sequences that discriminate *in vitro* and in the correct functional context *in vivo* between the designed and the wild-type partner E7 DNases. To our knowledge, the structure of the E7_C/Im7_C complex is the first high-resolution experimental validation of a computationally designed functional protein-protein interface with altered specificity, and provides valuable information on the strengths and limitations of the current energy function and protein representation used for protein design, as well as insights into the structural determinants of specificity. Similar and improved design strategies should be applicable to the design of stable and specific antagonists and protein pairs of novel specificity for the delineation and engineering of protein interaction networks in the cellular context.

METHODS

Computational protein design. The computational interface design procedure models amino acid side chains as rotamers in an all-atom representation (all heavy atoms and polar hydrogens) onto a fixed polypeptide template taken from the E7_WT/Im7_WT crystal structure[28] with polar hydrogens added as described[32]. Sequence positions were either designed (allowing rotamers for all 20 naturally occurring amino acids except cysteine using the backbone-dependent library compiled by Dunbrack[38] with additional rotamers for buried residues; positions to be designed were defined as described in Results), repacked (allowing all rotamers of the native amino acid type; this was done for residues directly contacting designed residues in a first interaction 'shell') or left unchanged in their native conformation (all other residues).

The general design strategy is described in Results. The DNase residue positions identified for the initial sequence perturbations were Asn516 and Lys528. At both sites, we computationally modeled single mutations at each position separately, and compared the predicted binding energies of each mutated DNase in complexes with (i) wild-type immunity protein, to estimate the destabilizing effect of the mutation on the wild-type interface (perturbed interface), and (ii) immunity protein, in which all interface residues were simultaneously redesigned (redesigned interface). In site 1, Asn516 interacts with several polar residues on the immunity protein (**Fig. 1b**), including several water molecules with low temperature factors (not shown). We sought to perturb this polar interaction network by inserting a large hydrophobic or aromatic residue (histidine, phenylalanine, tyrosine, tryptophan, isoleucine or leucine) at position 516. The computational screen identified leucine as the amino acid with a

© 2004 Nature Publishing Group http://www.nature.com/natstructmolbiol

© 2004 Nature Publishing Group http://www.nature.com/natstructmolbiol

sizeable difference in predicted binding energies of the perturbed and redesigned interfaces. For site 2, essentially all amino acids replacing Lys528 were predicted to destabilize the complex in comparison to the wild-type residue, with the largest difference found for phenylalanine. However, glutamine and tyrosine were predicted to yield the largest stabilizing effect upon redesign of the immunity partner protein, and were thus chosen as possible initial substitutions. To combine the immunity protein sequence changes obtained for the initial perturbations in sites 1 and 2 (**Table 1**, column 3) into a complete redesign comprising the whole interface, we enumerated combinations of these amino acids. For each sequence combination, we optimized the total energy of the complex using a Monte-Carlo simulated annealing protocol similar to that described[31]. The free energy function[31,32,39] is as described[7,32]. In a second step, we selected the best sequences based on their calculated binding energy computed as described[7] (the total energies of these sequences were also among the lowest sampled).

Construction of DNase-immunity protein designs. A plasmid for the wild-type E7/Im7 DNase-immunity construct ('pHBH,' a derivative of pQE30, Qiagen) was a gift from K. Chak (National Yang Ming University, Taipei, Taiwan). Designed constructs were cloned by standard methods. An enzymatically inactive variant of each design was also created by site-directed mutagenesis to introduce the DNase mutation H569A. These inactive constructs could be expressed with higher yields and were used for the SPR and fluorescence binding assays and the crystallography. The wild-type ColE7 plasmid containing the full-length DNase toxin (consistent of a receptor-binding 'R' domain, a translocation 'T' domain and the DNase domain) used in the *in vitro* DNA digestion assay and the *in vivo* plate assay was a gift from M. Riley (Yale University, New Haven). Variant constructs of the full-length toxin were cloned by standard procedures.

Purification and separation of complexes. All computationally selected variants were transformed into SG13009 (pREP4) *Escherichia coli* (Qiagen), expressed in 2XTY and purified as described[18]. For SPR and fluorescence binding experiments, the complex was separated on a Ni-NTA column by eluting the immunity protein with 7 M Guanidine-HCl. The DNase was then eluted with an imidazole step gradient. For functional studies the full-length colicin was expressed from a pET28a vector in BL21(DE3)pLysS cells and purified as above.

Toxicity plate assay. Purified full-length DNase–immunity protein complexes (colicin toxin) were serially diluted in PBS buffer and 4 μl was spotted at each dilution onto freshly prepared bacterial lawns on LB agar. Plates were incubated overnight at 37 °C. Immunity was tested using transformants of the *E. coli* JM109 host strain harboring pQE30 plasmid containing a designed DNase H569A–immunity insert for expression of functional immunity protein, but an enzymatically inactive DNase. This and the *in vitro* DNase activity assay below use full-length toxin; all other experiments used just the DNase domain.

***In vitro* DNase activity.** Purified full-length DNase designs (2 nM) were equilibrated with immunity protein variants (10 nM–1 μM) for 10 min before the addition of 1 μg plasmid DNA (pCDNA3, Invitrogen). Digestion was allowed to proceed for 1 h at 37 °C in a final volume of 20 μl. Protein was buffered in 50 mM Tris, pH 7.5, 20 mM $MgCl_2$, 1 mM $NiCl_2$. Reactions were stopped with 0.1 M EDTA and visualized on a 1.5% (w/v) agarose gel using ethidium bromide staining.

SPR binding analysis. SPR measurements were taken with a BIAcore 2000 and 3000 biosensor (BIAcore) in buffer containing 50 mM MOPS, pH 7.5, 200 mM NaCl and 0.005% (v/v) surfactant P20. Protein concentrations were determined from the absorbance at 280 nm, using a calculated molar extinction coefficient (Protein Calculator, http://www.scripps.edu/~cdputnam/protcalc.html). DNase proteins were coupled to CM5 research-grade gold biosensor chips using amine-coupling chemistry. BSA or empty flow cells were used as concurrent negative controls. The immunity proteins were injected at 30 μl min^{-1} in a range of concentrations from 0.1 nM to 500 nM at 25 °C. See **Supplementary Figure 2** online for a more detailed discussion of the analysis of the SPR sensograms and resolution of associated problems.

Fluorescence binding analysis. Fast binding kinetics were monitored by fluorescence on a stopped-flow Biologic SFM4-QFM4 at 25 °C with an excitation wavelength of 280 nm using a 0.4 mm cuvette (FC-04) and a 324-nm cutoff filter. The longer-timescale experiments were done using manual mixing on a Spex Fluorolog 1681 0.22-m spectrometer with 280-nm excitation and 345-nm emission wavelengths.

Association kinetics were obtained using 1 μM DNase and five different concentrations of immunity protein (12, 10, 8, 6 and 4 μM) in the presence of 50 mM MOPS, pH 7.5, and either 600 mM NaCl or 200 mM NaCl and 400 mM NaI. Dissociation kinetics were acquired by competition experiments. A preformed 10 μM DNase–immunity protein complex was chased with a three-fold excess of Im7_WT, buffered in 50 mM MOPS, pH 7.5, 200 mM NaCl and 400 mM NaI or 600 mM NaI at 25 °C.

For the fast dissociation experiments the traces were fit to a double-exponential fit using an implementation of the Pade-Laplace algorithm in the Biokine software (Biologic). Of the two rates extracted from the fit, the faster rate constant (1×10^5 M^{-1} s^{-1}) was concentration-dependent and reflects the association of competitor immunity protein with DNase. The slower rate constant was concentration-independent and corresponds to the concentration-independent rate of dissociation of the initial complex; this was taken to be the dissociation rate constant. The traces acquired for the slower dissociation competition experiments were fit to single exponential curves using the Kaleidograph software package (Synergy). The rate constant was independent of the concentration of the excess Im7_WT competitor as expected for the dissociation rate constant; the faster association with the competitor observed in the stopped-flow experiments was not resolvable in these manual mixing experiments.

See **Supplementary Figure 3** online for a more detailed discussion of the methods used to determine and analyze the fluorescence binding traces.

Crystallization and data collection. Crystals of the E7_C/Im7_C cognate complex were grown at room temperature in hanging drops with 1 μl reservoir containing 18% (w/v) PEGMME 2000, 200 mM $(NH_4)_2SO_4$, 50 mM sodium acetate, pH 4.6, 25% (v/v) glycerol, 5% (v/v) DMSO, mixed with 1 μl of 25–30 mg ml^{-1} protein. Crystals were flash-frozen in liquid nitrogen. Diffraction data were recorded to a resolution of 2.1 Å at the Advanced Light Source beamline 5.0.1 (Berkeley, California, USA). For the resolution range (50–2.1Å) the data were 97.3% complete with an R_{merge} of 4.6%. There were 33,148 total reflections recorded of which 16,460 were unique. The intensities were integrated using DENZO and SCALEPACK[40] (see **Supplementary Table 1** online for complete statistics).

Data refinement and model building. The structure was solved via molecular replacement using EPMR[41] with the E7_WT/Im7_WT (PDB entry 7CEI) complex as the initial search model; the correlation coefficient for the solution was 52.1%. The structure was modeled in XtalView[42] and refined using CNS[43] with a 9.5% dataset for cross-validation. The final R_{work} and R_{free} for the E7_C/Im7_C complex were 24.2% and 27%, respectively (see **Supplementary Table 1** online for complete statistics).

Coordinates. The atomic coordinates of the E7_C/Im7_C complex have been deposited in the Protein Data Bank (accession code 1UJZ).

Note: Supplementary information is available on the Nature Structural & Molecular Biology website.

ACKNOWLEDGMENTS

We thank the members of the Baker lab for stimulating discussions, and J. Havranek for critical reading of the manuscript. This work was supported by a long-term fellowship from the Human Frontier Science Program (T.K.), a Wellcome Trust International Prize fellowship (A.N.B.), a US National Institutes of Health (NIH) training grant (L.A.J.) and a grant from the NIH (D.B.).

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Received 18 August 2003; accepted 23 February 2004
Published online at http://www.nature.com/natstructmolbiol/

1. Bogan, A.A. & Thorn, K.S. Anatomy of hot spots in protein interfaces. *J. Mol. Biol.* **280**, 1–9 (1998).
2. Conte, L.L., Chothia, C. & Janin, J. The atomic structure of protein-protein recognition sites. *J. Mol. Biol.* **285**, 2177–2198 (1999).
3. Sharp, K.A. Calculation of HyHel10-lysozyme binding free energy changes: effect of ten point mutations. *Proteins* **33**, 39–48 (1998).
4. Massova, I. & Kollman, P.A. Computational alanine scanning to probe protein-protein interactions: a novel approach to evaluate binding free energies. *J. Am. Chem. Soc.* **121**, 8133–8143 (1999).
5. Huo, S., Massova, I. & Kollman, P.A. Computational alanine scanning of the 1:1 human growth hormone-receptor complex. *J. Comput. Chem.* **23**, 15–27 (2002).
6. Guerois, R., Nielsen, J.E. & Serrano, L. Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *J. Mol. Biol.* **320**, 369–387 (2002).
7. Kortemme, T. & Baker, D. A simple physical model for binding energy hot spots in protein-protein complexes. *Proc. Natl. Acad. Sci. USA* **99**, 14116–14121 (2002).
8. Brannetti, B., Via, A., Cestra, G., Cesareni, G. & Helmer-Citterich, M. SH3-SPOT: an algorithm to predict preferred ligands to different members of the SH3 gene family. *J. Mol. Biol.* **298**, 313–328 (2000).
9. Vaccaro, P. *et al.* Distinct binding specificity of the multiple PDZ domains of INADL, a human protein with homology to INAD from *Drosophila melanogaster*. *J. Biol. Chem.* **276**, 42122–42130 (2001).
10. Li, L., Shakhnovich, E.I. & Mirny, L.A. Amino acids determining enzyme-substrate specificity in prokaryotic and eukaryotic protein kinases. *Proc. Natl. Acad. Sci. USA* **100**, 4463–4468 (2003).
11. Aloy, P. & Russell, R.B. Interrogating protein interaction networks through structural biology. *Proc. Natl. Acad. Sci. USA* **99**, 5896–5901 (2002).
12. Wollacott, A.M. & Desjarlais, J.R. Virtual interaction profiles of proteins. *J. Mol. Biol.* **313**, 317–342 (2001).
13. Desjarlais, J.R. & Handel, T.M. *De novo* design of the hydrophobic cores of proteins. *Protein Sci.* **4**, 2006–2018 (1995).
14. Ventura, S. *et al.* Conformational strain in the hydrophobic core and its implications for protein folding and design. *Nat. Struct. Biol.* **9**, 485–493 (2002).
15. Bolon, D.N., Marcus, J.S., Ross, S.A. & Mayo, S.L. Prudent modeling of core polar residues in computational protein design. *J. Mol. Biol.* **329**, 611–622 (2003).
16. Pokala, N. & Handel, T.M. Review: protein design—where we were, where we are, where we're going. *J. Struct. Biol.* **134**, 269–281 (2001).
17. Hellinga, H.W., Caradonna, J.P. & Richards, F.M. Construction of new ligand binding sites in proteins of known structure. II. Grafting of a buried transition metal binding site into *Escherichia coli* thioredoxin. *J. Mol. Biol.* **222**, 787–803 (1991).
18. Dahiyat, B.I. & Mayo, S.L. *De novo* protein design: fully automated sequence selection. *Science* **278**, 82–87 (1997).
19. Nauli, S., Kuhlman, B. & Baker, D. Computer-based redesign of a protein folding pathway. *Nat. Struct. Biol.* **8**, 602–605 (2001).
20. Harbury, P.B., Plecs, J.J., Tidor, B., Alber, T. & Kim, P.S. High-resolution protein design with backbone freedom. *Science* **282**, 1462–1467 (1998).
21. Kuhlman, B. *et al.* Design of a novel globular protein fold with atomic level accuracy. *Science* **302**, 1364–1368 (2003).
22. Looger, L.L., Dwyer, M.A., Smith, J.J. & Hellinga, H.W. Computational design of receptor and sensor proteins with novel functions. *Nature* **423**, 185–190 (2003).
23. Harbury, P.B., Zhang, T., Kim, P.S. & Alber, T. A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants. *Science* **262**, 1401–1407 (1993).
24. Havranek, J.J. & Harbury, P.B. Automated design of specificity in molecular recognition. *Nat. Struct. Biol.* **10**, 45–52 (2003).
25. Shifman, J.M. & Mayo, S.L. Modulating calmodulin binding specificity through computational protein design. *J. Mol. Biol.* **323**, 417–423 (2002).
26. Reina, J. *et al.* Computer-aided design of a PDZ domain to recognize new target sequences. *Nat. Struct. Biol.* **9**, 621–627 (2002).
27. Shifman, J.M. & Mayo, S.L. Exploring the origins of binding specificity through the computational redesign of calmodulin. *Proc. Natl. Acad. Sci. USA* **100**, 13274–13279 (2003).
28. Ko, T.P., Liao, C.C., Ku, W.Y., Chak, K.F. & Yuan, H.S. The crystal structure of the DNase domain of colicin E7 in complex with its inhibitor Im7 protein. *Structure Fold. Des.* **7**, 91–102 (1999).
29. Schamberger, G.P. & Diez-Gonzalez, F. Selection of recently isolated colicinogenic *Escherichia coli* strains inhibitory to *Escherichia coli* O157:H7. *J. Food. Prot.* **65**, 1381–1387 (2002).
30. Murinda, S.E., Rashid, K.A. & Roberts, R.F. In vitro assessment of the cytotoxicity of nisin, pediocin, and selected colicins on simian virus 40-transfected human colon and Vero monkey kidney cells with trypan blue staining viability assays. *J. Food. Prot.* **66**, 847–853 (2003).
31. Kuhlman, B. & Baker, D. Native protein sequences are close to optimal for their structures. *Proc. Natl. Acad. Sci. USA* **97**, 10383–10388 (2000).
32. Kortemme, T., Morozov, A.V. & Baker, D. An orientation-dependent hydrogen bonding potential improves prediction of specificity and structure for proteins and protein-protein complexes. *J. Mol. Biol.* **326**, 1239–1259 (2003).
33. Morozov, A.V., Kortemme, T. & Baker, D. Evaluation of models of electrostatic interactions in proteins. *J. Phys. Chem. B* **107**, 2075–2090 (2003).
34. Wallis, R. *et al.* Protein-protein interactions in colicin E9 DNase-immunity protein complexes. 2. Cognate and noncognate interactions that span the millimolar to femtomolar affinity range. *Biochemistry* **34**, 13751–13759 (1995).
35. Kleanthous, C., Hemmings, A.M., Moore, G.R. & James, R. Immunity proteins and their specificity for endonuclease colicins: telling right from wrong in protein-protein recognition. *Mol. Microbiol.* **28**, 227–233 (1998).
36. Kuhlmann, U.C., Pommer, A.J., Moore, G.R., James, R. & Kleanthous, C. Specificity in protein-protein interactions: the structural basis for dual recognition in endonuclease colicin–immunity protein complexes. *J. Mol. Biol.* **301**, 1163–1178 (2000).
37. Covell, D.G. & Wallqvist, A. Analysis of protein-protein interactions and the effects of amino acid mutations on their energetics. The importance of water molecules in the binding epitope. *J. Mol. Biol.* **269**, 281–297 (1997).
38. Dunbrack, R.L., Jr. & Cohen, F.E. Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Sci.* **6**, 1661–1681 (1997).
39. Lazaridis, T. & Karplus, M. Effective energy function for proteins in solution. *Proteins* **35**, 133–152 (1999).
40. Otwinowski, J. & Minor, W. Processing of X-ray data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
41. Kissinger, C.R., Gehlhaar, D.K. & Fogel, D.B. Rapid automated molecular replacement by evolutionary search. *Acta Crystallogr. D* **55**, 484–491 (1999).
42. McRee, D.E. XtalView/Xfit—a versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol.* **125**, 156–165 (1999).
43. Brunger, A.T. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).

© 2004 Nature Publishing Group http://www.nature.com/natstructmolbiol