doi:10.1016/j.jmb.2006.05.022 J. Mol. Biol. (2006) 361, 195–208

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Computational Design of a New Hydrogen Bond Network and at Least a 300-fold Specificity Switch at a Protein–Protein Interface

Lukasz A. Joachimiak[1], Tanja Kortemme[2] Barry L. Stoddard[3] and David Baker[1]*

[1]*Howard Hughes Medical Institute & Department of Biochemistry, Box 357350 University of Washington Seattle, WA 98195-7350, USA*

[2]*Department of Biopharmaceutical Sciences & California Institute for Quantitative Biomedical Research, Box 2540 University of California San Francisco, CA 94143, USA*

[3]*Basic Sciences Division Fred Hutchinson Cancer Research Center 1100 Fairview Avenue N. A3-023 Seattle, WA 98109, USA*

**Corresponding author*

The redesign of protein–protein interactions is a stringent test of our understanding of molecular recognition and specificity. Previously we engineered a modest specificity switch into the colicin E7 DNase–Im7 immunity protein complex by identifying mutations that are disruptive in the native complex, but can be compensated by mutations on the interacting partner. Here we extend the approach by systematically sampling alternate rigid body orientations to optimize the interactions in a binding mode specific manner. Using this protocol we designed a *de novo* hydrogen bond network at the DNase–immunity protein interface and confirmed the design with X-ray crystallographic analysis. Subsequent design of the second shell of interactions guided by insights from the crystal structure on tightly bound water molecules, conformational strain, and packing defects yielded new binding partners that exhibited specificities of at least 300-fold between the cognate and the non-cognate complexes. This multi-step approach should be applicable to the design of polar protein–protein interactions and contribute to the re-engineering of regulatory networks mediated by protein–protein interactions.

Keywords: computational design; hydrogen bond network; DNAse; immunity protein; protein–protein interactions

Introduction

Protein interactions mediate many cellular processes and form a substantial part of the regulatory networks responsible for biological function. To engineer protein–protein interactions requires an understanding of how proteins discriminate their natural binding partners from other competing proteins of similar structure and sequence. Simple rules to identify protein recognition sites and predict energetic hot spots in protein–protein interfaces often fail,[1] largely because of the extreme diversity in shape, chemical character and plasticity of protein–protein interfaces.[2] Computational tools, aided by an increase in sequence and structural information, have begun to make possible the rational engineering of protein interfaces,[3–7] and offer promise in building *de novo* protein interaction networks. To probe the contribution of interactions between a particular pair of proteins amongst many interactions made by either or both partners, methods for redesigning both partners to maintain their interaction but disrupt interactions with other proteins would be very useful.

A number of computational methods for the prediction and design of interaction specificity have been developed recently, based on experimental data on specific systems,[8,9] evolutionary information[10] and empirical energy functions.[11,12] Rotamer-search based methods have been applied to the design of new protein–small molecule,[13] coiled-coil[14,15] and protein–peptide interfaces, which have shown novel[15,16] or enhanced specificities.[17,18] Mayo and co-workers demonstrated that an automated computational design algorithm can be used to enhance calmodulin–peptide interface specificity.[17,18] Serrano and co-workers applied a computational design algorithm to alter the binding pocket in PDZ domains to recognize

Abbreviation used: SPR, surface plasmon resonance.

E-mail address of the corresponding author: dabaker@u.washington

novel peptide sequences,[16] while Havranek and Harbury developed a negative design strategy to achieve specificity in a coiled-coil system.[15] Bolon *et al*. showed that explicitly modeling competing states can increase specificity, but at a cost to stability.[19]

To engineer novel specificities at protein–protein interfaces we have chosen as a model system the complex between the colicin E7 DNase and the immunity protein Im7.[20] Previously, we engineered a new cognate pair that displayed a modest specificity switch by applying a ''second site suppressor'' strategy.[21] However the 30-fold difference in affinities between the new cognate pair (as compared to non-cognate complexes) was many orders of magnitude smaller than that observed in naturally occurring colicin/immunity protein pairs, which span five to eight orders of magnitude affinity differences between cognate and non-cognate pairs.[22] Here we have extended our previous work by (1) moving beyond the fixed peptide backbone approximation used in our design procedure by systematically sampling alternate rigid binding orientations, (2) designing a novel crystallographically confirmed hydrogen bond network and (3) achieving a specificity switch of at least 300-fold. This multi-step approach should be applicable to the engineering of *de novo* protein interaction networks mediated by orthogonal interacting protein pairs and the delineating of the critical interactions in complex signaling networks.

Figure 1. Rotation in the binding mode between the E7–Im7 and E9–Im9 complexes. The E7 DNase is colored colored in teal and the Im7 and Im9 proteins are colored in gray and pink, respectively. A 19° rigid-body rotation relates the orientation if the Im7 and Im9 proteins in their respective complexes superimposed onto the E7 DNase structure. A conserved YY motif (side-chains in yellow stick representation) on the immunity proteins serves as the anchor for the change in binding mode. Figures were made using PyMOL (Delano Scientific).

Results

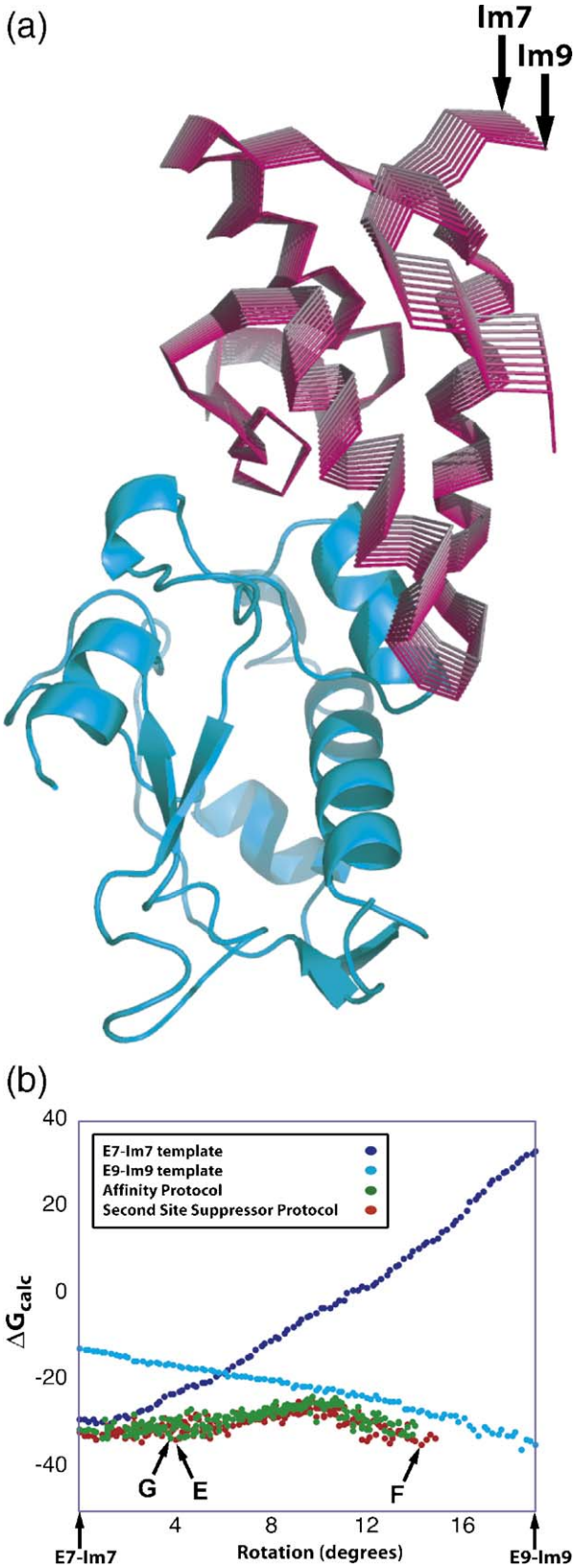
Design protocol

A limitation to our earlier second site suppressor approach[21] was that the relative orientation of the two protein partners in the complex was fixed to that of the wild-type E7–Im7 complex throughout the design process. Specificity in the native colicin DNase–immunity protein system comes about in part because of a 19° rigid body rotation that presents residues in a somewhat different orientation in the E7–Im7 and the homologous E9–Im9 complexes (Figure 1).[23] A conserved YY motif on the immunity protein shown by Kleanthous *et al.* to contribute significantly to binding serves as the pivot point for the rotation.

To go beyond the fixed backbone limitation of our previous work, we followed nature's lead and created ensembles of conformations by sampling along the rigid body transformation relating the E7–Im7 and E9–Im9 complexes in 0.2° increments centered around the conserved YY motif (Figure 2(a)). In all of our design calculations we chose to keep the side chain conformations of this YY motif fixed. We used two different protocols to create specific and high affinity complexes using these ensembles. The first protocol is a generalization of our second site suppressor strategy, which has an explicit negative design element. Computational screening was used to find mutations that are most destabilizing to the native binding mode. Binding orientation diversity was introduced as described above, and the interface was redesigned to optimize the interacting residues to compensate for the introduced mutations in the context of each of the binding modes in the ensemble. In the second affinity protocol, we optimized all interface positions for each binding mode in the ensemble without explicitly including mutations destabilizing the wild-type interface. The predicted binding energies of the optimized interfaces produced using the second site suppressor and affinity protocols were computed as described by Kortemme & Baker[7] and the sequences with the highest predicted affinities were selected for analysis (Tables 1 and 2A).

The importance of including explicit negative design in specificity calculations has been extensively discussed.[15,18,19] We explored the affinity protocol because it appeared likely that if the variation in the binding mode is substantial, such a positive design protocol alone could generate specificity. To test this idea, we computed the binding energies of the E7–Im7 and E9–Im9 complexes without sequence redesign for each rigid body orientation in the ensemble. For each orientation, we optimized the side-chain conformations of all interface residues using a Monte-Carlo rotamer sampling protocol (see Methods). As is evident in Figure 2(b), the predicted

binding energy of both native complexes rises sharply as the orientation deviates from that in the wild-type structure. Thus, an interface sequence optimized in one binding mode is predicted to be far from optimal in another.



Designed interfaces

The designs with the lowest binding energy from the affinity and second site suppressor calculations that lacked unsatisfied hydrogen bond donors and acceptors were selected for experimental characterization. Of the 11 designs chosen for characterization (six from the affinity protocol and five from the second site suppressor protocol), in four cases the immunity protein was insoluble (the native immunity protein is already marginally stable) and in four of the remaining cases no binding was observed. The three remaining designs, indicated by arrows in Figure 2(b), were well expressed, soluble, and demonstrated binding in our initial assay (described in the next section); the sequences of these designs are shown in Table 1. For continuity with our previous work, our naming convention begins with ''E'', and the non-cognate pairs are denoted with the DNase first and the immunity protein second. Designs E and F are from the second site suppressor protocol, and design G is from the affinity protocol.

The E design (E7_E/Im7_E), derived from a modest binding mode perturbation (3.4°), contains two second site suppressor mutations, D23Y and N26I (Figure 2(c)). The N26I mutation is predicted to destabilize the native interface by disrupting a

Figure 2. Design using a rigid body ensemble. (a) Illustration of the rigid body ensemble used in design calculations. The DNase is colored in teal and is shown in cartoon representation. The immunity protein structures are colored from gray (Im7 orientation) to magenta (Im9 orientation) and are shown in wireframe representation. (b) The binding energy was computed for rigid body orientations on the E7–Im7 to E9–Im9 trajectory for the wild-type E7–Im7 sequence (blue), the E9–Im9 sequence (cyan), the lowest binding energy structures from the affinity protocol (green), and the second site suppressor protocol (red). As the rigid body orientation moves from the E7–Im7 binding mode towards the E9–Im9 binding mode, the interactions from the E7–Im7 sequence are no longer predicted to be favorable (blue). A similar result is obtained for the E9–Im9 sequence with the native structure perturbed towards the E7–Im7 binding mode (cyan). (c)–(e) Comparison of wild-type and designed interfaces. The DNase is colored in teal and the immunity protein is gray. (c) Model of the E design. The design was selected from the second site suppressor protocol and derived from a modest binding mode perturbation (3.4°); it contains two second site suppressor mutations, D23Y and N26I. (d) Model of the F design. The design was selected from the second site suppressor protocol, contains two destabilizing mutations (D23Y and N26V) and has a significant (15°) rotation in the binding mode. 15 additional mutations were selected, incorporating a combination of steric and polarity switches, yielding an interface predicted to be optimized for this substantially altered binding mode. (e) Model of the G design. This design is selected from the affinity protocol, is derived from a modest change in binding mode (3.2°) and contains seven mutations. The corresponding residues in the wild-type crystal structure are shown on the left in each panel.

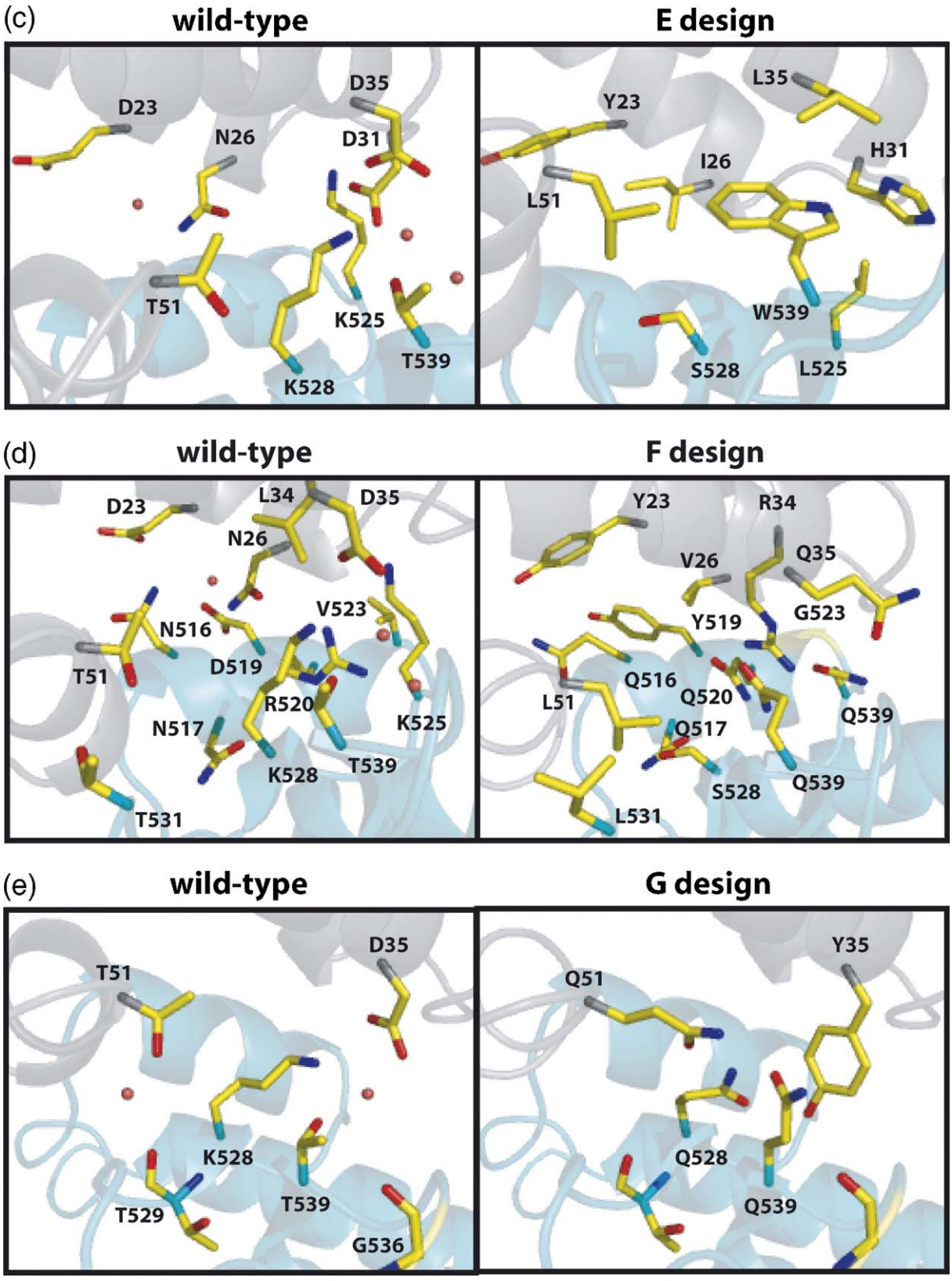


Figure 2 (*legend on previous page*)

hydrogen bond network. In the design it serves as the anchor point for a polarity switch, which replaces a polar interaction network with predicted interactions dominated by van der Waals packing. The second position, D23Y, is predicted to disrupt a salt-bridge in the native interface while in the design this mutation caps the hydrophobic interface.

Table 1. Positions designed in the colicin DNase–immunity protein interface

Sequence position	Immunity protein										DNAse										
	19	20	23	26	31	34	35	51	52	68	514	516	517	519	520	523	525	528	530	531	539
Wild-type	L	K	E	N	D	L	D	T	D	I	S	N	N	D	R	V	K	K	R	T	T
Second site suppressor protocol																					
E design		M	Y	I	H		L	L	N			Q			K		L	S			W
F design		L	Y	V		R	Q	L	N		T	Q	Q	Y	Q	G	N	S		L	Q
Affinity protocol																					
G design							Y	Q				T	Q				R	Q			Q
Structure based optimization																					
G2 design							Y	Q				T					R	Q			Q
G2_68F design	V						Y	Q		F		T					R	Q			Q
G2_68W design							Y	Q		W		T					R	Q			Q

Empty fields indicate wild-type residues. Initial destabilizing mutations are in gray boxes.

The F design (E7_F/Im7_F) contains two second site suppressor mutations, D23Y and N26V, and is derived from a 15° rotation in the binding mode (Figure 2(d)). To accommodate this large shift in binding mode, 15 additional mutations incorporate a combination of steric and polarity switches to yield an interface optimized for a substantially altered binding mode. A key feature of this interface is that position V523 of the DNase was changed to a glycine residue to accommodate the altered position of a loop on the immunity protein.

The G design (E7_G/Im7_G) was generated using the affinity protocol, and is derived from a binding mode with a modest 3.2° rotation. It contains seven mutations, which cluster into two sites. The first site contains four mutations, which form a single hydrogen bond network coordinated by K528Q and T539Q on the DNase and D35Y and T51Q on the immunity protein (Figure 2(e)). In the native interface this region contains a salt-bridge between D35 and K528 and a water-mediated hydrogen bond from the hydroxyl of T539 to the delta oxygen of D35. At the fourth position, T51, the hydroxyl forms a water-mediated hydrogen bond to the backbone carbonyl of T529. At the second site on the DNase, the mutations N516T and N517Q replace water-mediated contacts across the interface with direct contacts. The arginine introduced, the last mutation (K525R) in the G design, forms a salt bridge similar to that made by the native lysine residue.

Table 2. Computed binding energies are consistent with experimentally observed specificity between cognate and non-cognate interactions

Complex	ΔG_{calc}
A. *Comparison of the predicted binding energies for the three models chosen for experimental characterization and the two crystal structures of native cognate complexes*	
E design	−31.8
F design	−34.0
G design	−31.4
E7-Im7 (7CEI.pdb)	−29.0
E9-Im9 (1EMV.pdb)	−34.4
B. *Comparison of the predicted binding energies for the G design variants using the G design crystal structure as the template*	
G design (2ERH.pdb)	−32.0
G2_68W design	−31.5
G2_68F design	−33.7
E7_WT/Im7_G_68W	−22.0
E7_WT/Im7_G_68F	−23.8

The non-cognate interactions are predicted to be substantially weaker, as experimentally observed. Binding energy calculations are performed as described by Kortemme *et al.*[7]

Characterization of the designed complexes

The designed endonuclease and immunity protein pairs were co-expressed from a bi-cistronic construct containing a poly-histidine tag on the DNase. Co-elution of the complex from a Ni-NTA affinity column is an indicator of binding of the untagged immunity protein to the DNase. The elution profile was inspected on a Coomassie stained SDS-PAGE gel to determine the amount of immunity protein retained during the elution. The E and F designed complexes did coelute, however, the stoichiometry did not appear to be 1:1, suggesting weak binding, whereas the G design eluted as a complex. The eluant from the G design purification was analyzed on a gel filtration column and migrated as a single monodisperse peak corresponding to a M_r of ~26 kDa (expected M_r = 25.7 kDa), while the E and F designs migrated as a population of monomers and heterodimers (data not shown). Based on these results, further characterization focused on the G design.

We used surface plasmon resonance (SPR) to compare the *in vitro* binding of the cognate G design and non-cognate E7_G/Im7_WT hybrid complexes. Consistent with previous SPR analysis of designed colicin interactions, and the work of Kleanthous and co-workers,[22] the association rates of the cognate and the non-cognate complexes are quite similar (between 1×10^5 and 5×10^5 M^{-1} s^{-1}). In contrast, the dissociation rates vary substantially

and are the major determinant of differences in affinity. The rate of dissociation for the cognate G design is $k_{off,app} = 0.02\ s^{-1}$ (Suppl. Mat., Figure 1(a)); this corresponds to an estimated dissociation constant of 65 nM (assuming $k_{on} = 3.2 \times 10^6\ M^{-1}\ s^{-1}$). Comparison of the cognate G design dissociation rate to the similar non-cognate E7_G/Im7_WT dissociation rate suggests a small specificity switch (<threefold, Suppl. Mat., Figure 1(b)). The non-cognate E7_G/Im7_WT complex shows threefold weaker binding, while the second non-cognate E7_WT/Im7_G complex, which differs from the high affinity WT complex by only two mutations (on the immunity protein) dissociates very slowly.

Crystal structure of the G design

To better characterize the G design we grew crystals and were able to determine the crystal structure to 2.0 Å resolution. The electron density surrounding the designed residues is well resolved with *B*-factors ranging between 20 $Å^2$ to 30 $Å^2$ for buried residues and slightly higher *B*-factors for residues exposed to solvent (Figure 3(a)). An overlay of the designed interface region in the G design crystal structure and in the computational model reveals that the interactions in the designed hydrogen bond network are predicted correctly (Figure 3(b)–(d)) with all-atom and Cα r.m.s.

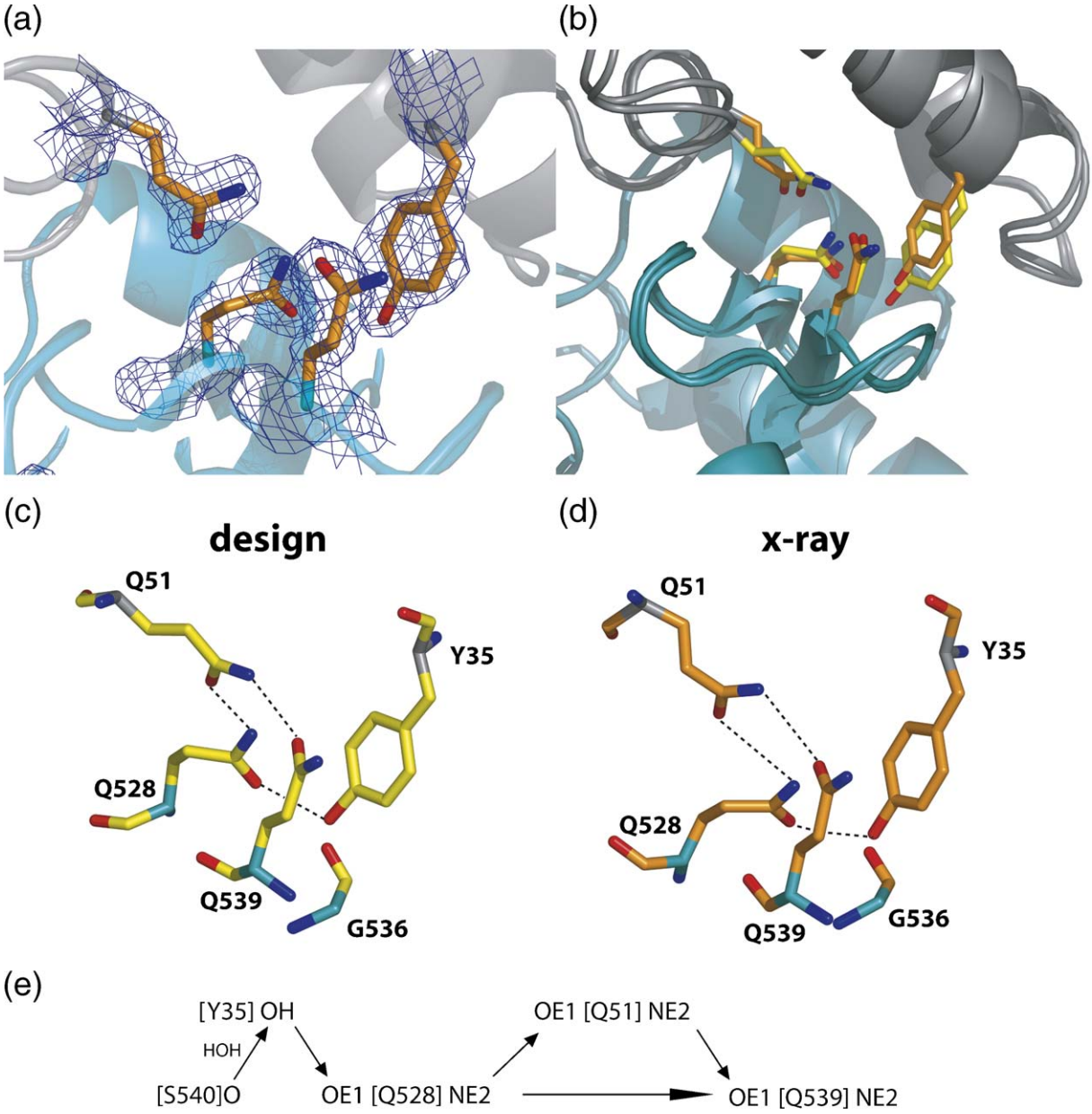


Figure 3. The crystal structure of the G design. The DNase backbone is in teal and the immunity protein backbone is in gray. (a) $2F_o - F_c$ density is contoured at 1.3σ (blue). (b) Overlay of the design model (yellow side-chains) with the experimentally determined structure (orange side-chains). (c) and (d) Comparison between the engineered hydrogen bond network in the design model (c) and the crystal structure (d). Connectivity in the designed network and how it is integrated into the native surrounding interface region (e).

deviation of 0.63 Å and 0.3 Å, respectively, for all interface residues (defined as residues having at least one atom within 4 Å of an atom in the partner). The side-chain conformations of four of the designed interface residues, Y35, Q51, Q528 and Q539, in the crystal structure match well with the rotamers used in the computational design model with an all-atom r.m.s. deviation of 0.72 Å (Figure 3(c) and (d)). However, subtle changes in the crystal structure backbone lead to a different rotamer for the central Q51 of the hydrogen bond network, which strains the hydrogen bond geometries (Table 3). While the rotamer is incorrectly predicted, the $\chi 2$ angle compensates to preserve the two hydrogen bonds to the epsilon nitrogen of Q528 and the epsilon oxygen of Q539. The Y35 side-chain hydroxyl group makes an additional water-mediated hydrogen bond to the carbonyl backbone of S540 (Figure 3(e)) which was not predicted because water molecules were not included in the design calculations. As none of the hydrogen bonds are present in the WT complex, we have successfully engineered and structurally validated a new hydrogen bond network coordinated by four designed side-chains.

There are two other mutations in this designed interface not involved in the hydrogen bond network: N516T and N517Q. The correctly modeled threonine at position 516 forms a hydrogen bond across the interface to the backbone carbonyl of I54. The side-chain hydroxyl group replaces a water molecule, showing that water-mediated interactions in some cases can be substituted with side-chain atoms. The other DNase mutation, N517Q, is not predicted correctly (Figure 4). In the native interface, an asparagine makes a side-chain hydrogen bond to the backbone amide of K528 and a water-mediated interaction to S514 (Oγ). In the G design model N517 is replaced by a Q, which is predicted to displace the bound water. However, the crystal structure shows that the water is not in fact displaced and maintains a tetrahedral coordination with two donors (S514 N, Q517 Nε^2H) and two acceptors (Y55 O and S514 Oγ); to compensate the Q bends back by adopting unusual $\chi 1$ and $\chi 2$ angles to maintain asparagine-like hydrogen bonding interactions and the two proteins are forced slightly apart (Figure 4). The water molecule is present in the same position in both the unbound E7 DNase and the unbound Im7 structures, suggesting that it is quite tightly bound. By underestimating the binding affinity of the water and replacing the N with a Q we likely introduced strain into the interface and perturbed the binding mode. The comparison of the binding modes between the experimental structure and design model reveals that in the crystal structure the design G immunity protein is shifted towards the native E7–Im7 binding orientation. This difference between the predicted and observed binding modes is likely due to the strain introduced by the Q517 residue, and may also strain the hydrogen bond geometries within the hydrogen bond network.

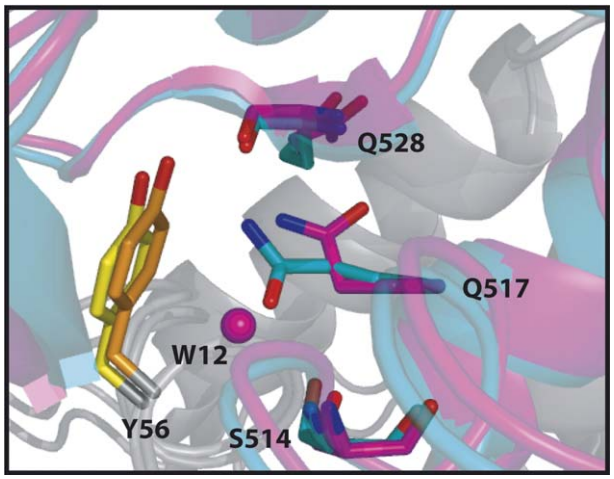


Figure 4. The N517Q mutation in the G design induces a backbone shift in the DNase. Overlay of the design model (teal and yellow side-chains) with the experimentally determined structure (magenta and orange side-chains). The Q517 side-chain in the G design crystal structure does not displace a tightly bound water molecule (magenta, W12) resulting in a backbone shift to accommodate a different Q side-chain rotamer. The immunity protein backbones are colored in gray.

Structure-based design optimization

We used the G design crystal structure as a scaffold for redesign to attempt to improve affinity of the cognate interaction and gain specificity against the wild-type partners. We reverted Q517 back to the wild-type N based on the knowledge that the water is tightly bound and unlikely to be displaced; this variant will be referred to subsequently as the G2 design. We observed further in the crystal structure of the G design that the packing against the methyl group of T516 is suboptimal (Figure 5(a)), while the

Table 3. Comparison of hydrogen-bond geometries between the crystal structure of the G design and the design model

Donor atom	Acceptor atom	δ_{HA} (Å)		θ (°)		ψ (°)	
		X-ray	Design	X-ray	Design	X-ray	Design
Y35 OH	$O^{\varepsilon 1}$ Q528	2.2	1.7	145.2	170.8	118.7	108.0
Q528 $N^{\varepsilon 2}$H	$O^{\varepsilon 1}$ Q51	2.2	1.7	145.4	168.2	97.0	127.4
Q51 $N^{\varepsilon 2}$H	$O^{\varepsilon 1}$ Q539	2.3	1.8	117.4	173.6	122.1	116.1
Q528 $N^{\varepsilon 2}$H	$O^{\varepsilon 1}$ Q539	3.0	2.1	108.5	121.7	124.8	141.0

$\delta_{HA,}$ distance between the hydrogen and acceptor atoms; θ, angle at the hydrogen atom; ψ, angle at the acceptor atom. The parameters used to describe the hydrogen bonds are as defined by Kortemme *et al.*[33]

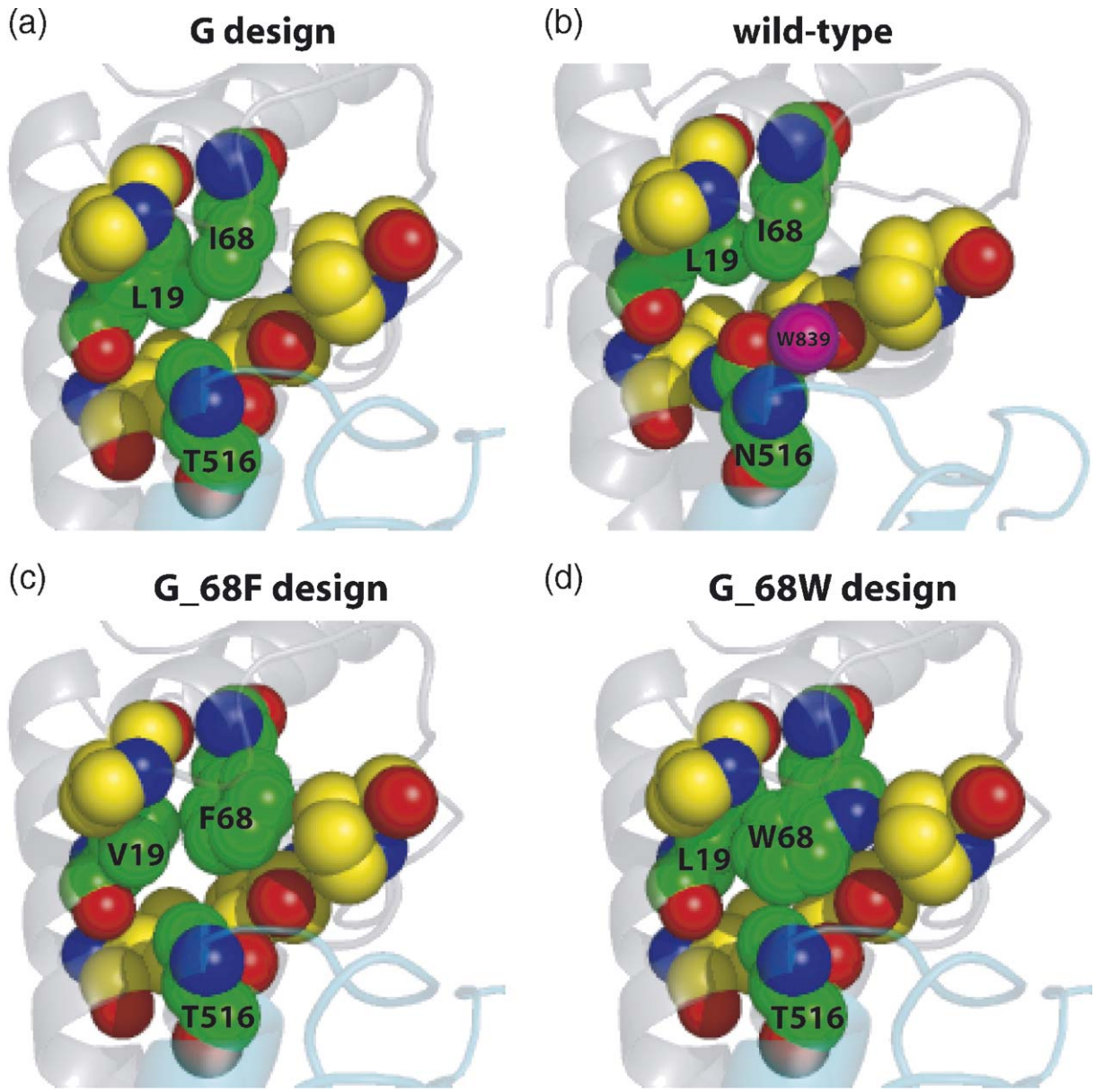


Figure 5. Structure-based optimization of the G design. The DNase is colored in teal and the immunity protein in gray. Residues participating in the interaction that have been changed or were allowed to vary are shown in space-fill representation, in green and yellow, respectively. (a) In the G design crystal structure the T516 hydroxyl group makes a hydrogen bond to the backbone carbonyl of I54, but the methyl group of the threonine is sub-optimally packed. (b) In the wild-type interface N516 forms a water-mediated (magenta) hydrogen bond to the backbone carbonyl of I54. Following sequence optimization surrounding T516 using the G design structure, the two sequences with the lowest predicted binding energies contained the L19V/I68F(c) and I68W mutations (d) (named G_68F and G_68W, respectively).

designed hydrogen bond from the T516 hydroxyl to the I54 carbonyl is made as predicted, replacing a water-mediated hydrogen bond between the delta oxygen of N516 and the backbone carbonyl of I54 in the native structure (Figure 5(b)). To improve specificity we redesign the positions around T516 (I22, I68 and L19); these residues were not selected in the first round of design because they are further away from the interface, forming part of the second shell of interactions. The sequence with the lowest predicted binding energy contained two mutations: L19V and I68F (cognate G_68F complex, Table 1, Figure 5(c)). The second lowest predicted binding energy sequence contained a single mutation: I68W (cognate G_68W complex, Table 1, Figure 6(d)). These mutations are predicted to add specificity because of clashes with the delta oxygen of N516 when in complex with the wild-type immunity protein (Table 2B).

Biophysical characterization of optimized variants

We again used SPR as a tool to assess the extent of specificity *in vitro* between optimized cognate and non-cognate interactions. Analysis of the wild-type

E7–Im7 binding by SPR shows no significant dissociation over the time scale of the experiment (1800 s, Suppl. Mat. Figure 1(c)). The dissociation for the G2 design (Figure 6(a)) is similar to the wild-type dissociation (Suppl. Mat. Figure 1(c)) and much slower than for the G design (Suppl. Mat. Figure 1(a)). Hence, the removal of the strain at the interface resulting from the failure to displace the tightly bound water produces a significant decrease in the dissociation rate. The non-cognate E7_G2/Im7_WT dissociation is faster (Figure 6(b)) than the cognate G2 design dissociation, suggesting modest specificity.

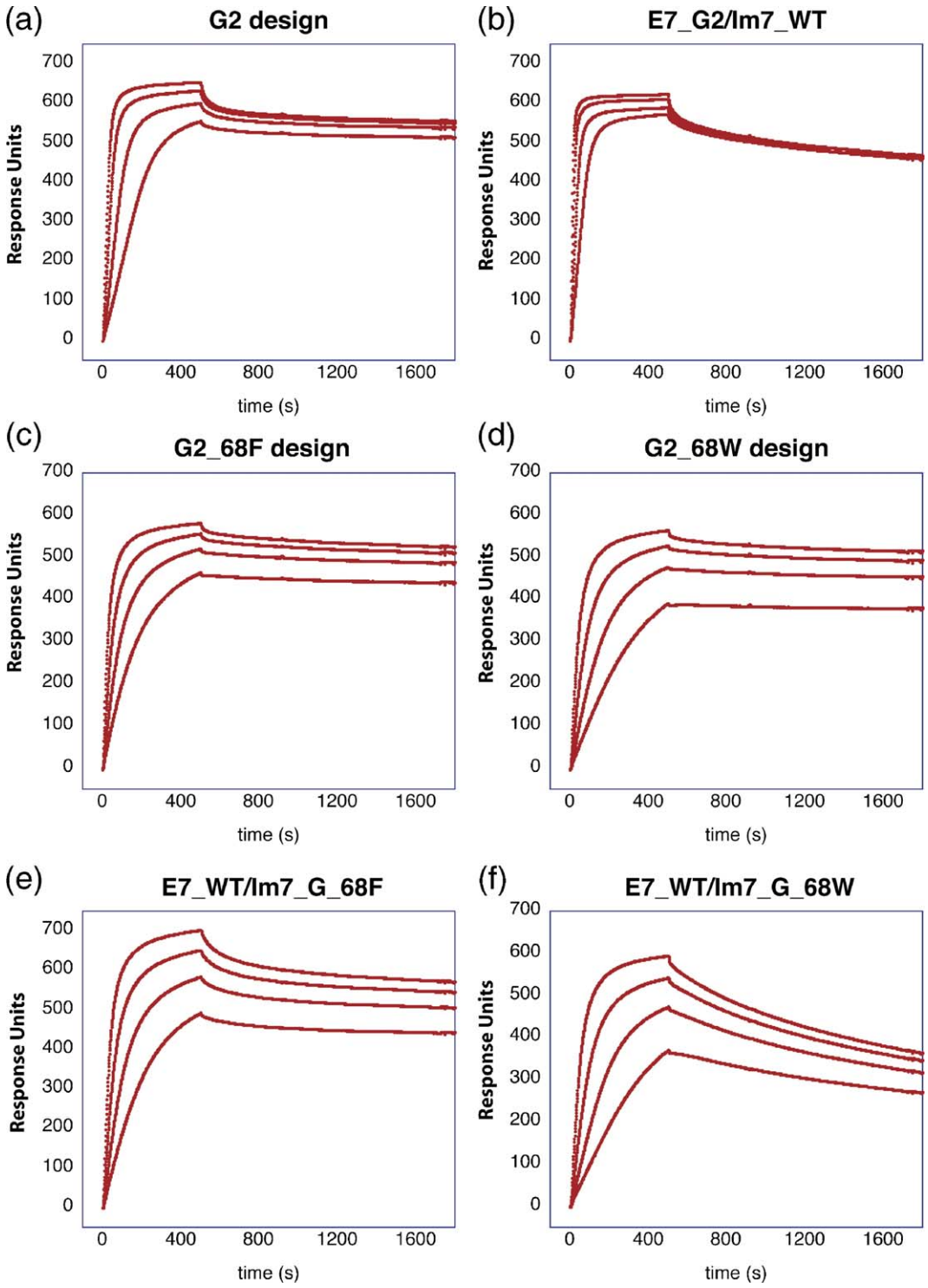


Figure 6. SPR sensograms for the G2 design and structure-based optimized variants. A comparison of SPR traces for the cognate G2 design (a), non-cognate E7_G2/Im7_WT (b), cognate G2_68F design (c), cognate G2_68W design (d), non-cognate E7_WT/Im7_G_68F (e) and non-cognate E7_WT/Im7_G_68W (f) at concentrations of 125 nM, 62.5 nM, 31.25 nM and 15.625 nM immunity protein.

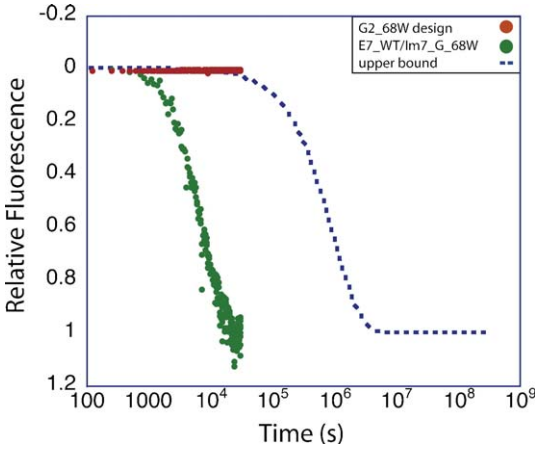


Figure 7. The G2_68W cognate complex dissociates much more slowly than the non-cognate complex. Dissociation rates monitored by the change of tryptophan fluorescence upon binding. There is a significant difference between the rate of dissociation for the cognate G2_68W interaction (red) and the non-cognate E7_WT/Im7_G_68W interaction (green). The estimated upper bound for the cognate dissociation rate, 1×10^{-6} s^{-1}, is represented by the dotted curve.

Both the cognate G2_68F (Figure 6(c)) and G2_68W (Figure 6(d)) designed complexes dissociate with rates similar to the cognate G2 complex (Figure 6(a)). A comparison of the dissociation rates between the non-cognate E7_WT/Im7_G_68W (Figure 6(f)) and the cognate G2_68W complex (Figure 6(d)) suggests a significant specificity switch.

Since the dissociation curves from the SPR experiments for the cognate interactions do not drop to baseline it is difficult to fit them accurately. To obtain a more quantitative measure of the dissociation rate we performed intrinsic fluorescence competition experiments following the substantial fluorescence change upon binding observed in the W-containing complexes. The comparison of the dissociation of the cognate G2_68W and non-cognate E7_WT/Im7_G_68W is shown in Figure 7. The dissociation constant for the non-cognate complex is $1.78\times10^{-4}\pm5.5\times10^{-5}$ s^{-1}. The signal change for the dissociation of the cognate complex is not measurable during the 8 h experiment; we can estimate an upper bound for k_{off} of 1×10^{-6} s^{-1} (Figure 7). Combining the fluorescence dissociation data and the SPR association rates we estimate the specificity switch between the cognate G2_68W and non-cognate E7_WT/Im7_G_68W complexes to be greater than 300-fold (Table 4).

To determine which mutations on the immunity protein of the G2_68W design contribute to the observed specificity switch, we introduced the I68W or D35Y/T51Q mutations into the wild-type E7−Im7 interface. Competition experiments showed that the rate of dissociation for the I68W complex is 5.6×10^{-5} s^{-1} (data not shown), which is over an order of magnitude slower than the non-cognate E7_WT/Im7_G_68W interaction. For the second set of mutations, D35Y/T51Q, there is no W in the interface but we used a previously characterized variant of the wild-type immunity protein containing a Y56W mutation[21] to monitor the fluorescence signal change. The fluorescence signal change for the dissociation of the wild-type D35Y/T51Q complex is not measureable during the 8 h time-course of the experiment and very little dissociation is observed in SPR experiments (data not shown). These results suggest that the I68W mutation is responsible for the majority, but not all of the specificity in the G2_68W design. The experimental binding data (Table 4) are consistent with the predicted specificity of the designs (Table 2B); rotamer changes can relieve clashes with the delta oxygen of N516 in the non-cognate E7_WT/Im7_G_68F complex but not with the bulkier W in the non-cognate E7_WT/Im7_G_68W complex.

Discussion

We used computational design followed by structure determination and a subsequent round of design to successfully engineer a new variant of the E7−Im7 colicin DNase−immunity protein complex that contains seven mutations (four on the E7 side and three on the immunity protein side) relative to wild-type, binds with near native affinity and exhibits specificity for cognate over non-cognate complex on the order of at least 300-fold (Table 4). This is an improvement over our

Table 4. Kinetic data for cognate G2_68W and non-cognate E7_WT/Im7_G_68W designed complexes

		$k_{app,on}$ (M^{-1} s^{-1}) ($\times10^5$)	$k_{app,off}$ (s^{-1})	$K_{app,d}$ (nM)
G2_68W design	SPR	1.80	–	–
	Fluorescence	–	$<1\times10^{-6}$	<0.0055
E7_WT/Im7_G_68W	SPR	1.00	–	–
	Fluorescence	–	0.00018	1.8
E7_WT/Im7_WT_68W	SPR	–	–	–
	Fluorescence	–	5.6×10^{-5}	–

SPR and fluorescence kinetic data for the cognate G2_68W and E7_WT/Im7_G68W non-cognate complexes. The apparent affinity $K_{app,d}$ is the ratio of the apparent dissociation and association rate constants $k_{app,off}$ and $k_{app,on}$. Both SPR and fluorescence experiments were carried out in 50 mM Mops (pH 7.5) and 150 mMNaCl. The affinity of the interaction between between the E7 DNase and its cognate Im7 immunity protein has not been reported in the literature, but the affinity of the similar E9−Im9 complex has been measured at 10^{-14} M to 10^{-16} M. [22]

previous engineering, which yielded variants with similar numbers of mutations but weaker affinity ($K_{d,app} = 3.4 \times 10^{-10}$ M), and which exhibited a much smaller specificity switch (30-fold). The crystal structure of the G design validated a *de novo* engineered hydrogen bond network, which is coordinated by four designed side-chains and extends into the native polar network. To our knowledge this is the first crystal structure of a successfully designed hydrogen bond network.

Inspired in the past by the solution to the specific molecular recognition problem in the naturally occurring colicin toxins and immunity proteins we developed a protocol that couples sequence design with sampling of alternate rigid body orientations. The sampling diversity achieved with the use of different rigid binding modes, coupled with our second site suppressor or affinity protocols, resulted in low energy sequence solutions that are divergent from naturally occurring colicin–immunity protein pairs. Experimental characterization of the initial designs revealed that modest binding mode changes with relatively few sequence changes were the most successful at retaining high affinity. While the specificity achieved here is significant, it is still lower than the 10^7–10^8-fold binding affinity differences observed between naturally occurring colicin–immunity protein pairs. Loop variability in the naturally occuring colicins, in addition to the rigid body variation, is likely to contribute to the specificity. Combining the sampling of binding modes with small backbone changes (i.e. relaxing the rigid-body approximation) could yield higher affinity and more specific designed complexes and is an area for future work.

In addition to verifying the accuracy of the designs, structural analysis is even more critical in revealing the shortcomings of current design techniques. In the case of the initial G design we attempted to replace a water-mediated contact with a direct contact, but the water is sufficiently tightly bound that instead the water was retained and the introduced side-chain adopted a strained conformation and the proteins were pushed slightly apart. The importance of interfacial water molecules that mediate recognition and stabilize the interface has previously been noted in protein–DNA[24,25] and protein–protein interfaces.[2,23,26–29] One way of incorporating these fixed water molecules into design is to supplement the design calculations with ''solvated rotamers'' with one or more water molecules in optimal hydrogen bonding geometries attached to standard side-chain rotamers.[30] This strategy has been successful in the prediction of water binding sites[30] but is challenging for design due to the large number of rotamers being modeled. A simpler approach is suggested by the well resolved water molecule that displaced Q517 in the G design: in the G design crystal structure the water that displaced the Q517 has the fifth lowest *B*-factor (17 Å^2), in the native E7–Im7 (7CEI.pdb) structure the equivalent water bridges two donors (S514 N and N517 Nδ^2H) and two acceptors (Y55 O and S514 Oγ) and has the second lowest *B*-factor (18.9 Å^2) out of 161 water molecules, in the unbound E7 DNase structure (1MO8.pdb) the water is very well resolved and has the lowest *B*-factor of all water molecules (6.0 Å^2), and in the related E9–Im9 structure (1EMV.pdb) the water is well defined and is coordinated by similar donor and acceptor groups (instead of S514 Oγ there is an asparagine Oδ^1) and has the fourth lowest *B*-factor of 154 water molecules (20.2 Å^2). Retaining such highly conserved bound water molecules during the design process would have improved our initial designs and may generally be useful in protein design calculations.

It is likely that by producing and testing a large number of designs more complete specificity changes and higher affinity binding could have been obtained in the initial search through the binding modes. Here, we only characterized the 11 designs from the initial rigid body searches and the two variants of the G design, which were all chosen based on their very low energies; no other proteins were tested. This focus on a small number of designs, while perhaps not yielding the greatest specificity switch, allowed us to do detailed structural and biophysical characterization necessary to learn from both the failures and successes and gain insights into the determinants of interaction specificity and affinity.

Conclusions

We have made advances in the engineering of polar interaction networks in protein–protein interfaces. The crystal structure of the G design demonstrated that novel hydrogen bond networks can be engineered and highlighted a shortcoming in our approach, the neglect of tightly bound water molecules. In a second round of design based on the crystal structure we generated a design with very high affinity and specificity. Taken together, our results highlight both the power and limitations of emerging protein design methodology; our new approaches for sampling alternate rigid body orientations to achieve specificity had to be complemented with an iterative structure-based design process to be successful. More generally, this work illustrates how experimental structural and biophysical characterization of computationally designed protein interfaces provides a stringent test of our understanding of macromolecular interactions and highlights the areas needing improvement.

Methods

Generation of template structures

Superposition of the E7–Im7 and the E9–Im9 complex structures along the coordinates of the DNase reveals a 19° rigid body rotation between the Im7 and Im9 proteins binding to their respective DNase partner. The hydroxyl group on the conserved Y56 of Im7 (equivalent

to Y55 of Im9) forms a hydrogen bond to a backbone carbonyl on the respective DNase. To create an ensemble of rigid body orientations as starting points for specificity redesigns we rotated the immunity protein around the axis defined by the Cα of Y56 and the center of mass of Im7 to maintain the conserved hydrogen bond interaction. A total of 100 structures were generated along a trajectory starting at the E7–Im7 binding mode and ending at the E9–Im9 binding mode that sampled the 19° rotation in 0.2° increments.

Computational protein design

The general design strategy is described in the Results. Positions identified for the initial sequence perturbations used in the second site suppressor protocol were E23 and N26 on the immunity protein. We computationally modeled single mutations at each position separately, and compared the predicted binding energies of each mutated immunity protein in complexes with (a) the wild-type DNase to estimate the destabilizing effect of the mutation on the wild-type interface (perturbed interface) and (b) an altered DNase where all interface residues were simultaneously redesigned (redesigned interface). E23 and N26 interact with a number of polar residues on the immunity protein, including several water molecules with low temperature factors (not shown). We sought to perturb this polar interaction network by the insertion of a hydrophobic or aromatic residue (H, F, Y, W, I, L or V) at position 23 and 26. The computational screen identified valine or isoleucine at position 26 as the amino acid with a sizeable difference in predicted binding energies of the perturbed and redesigned interfaces. The neighboring interface residues around the 23/26 sites were optimized for each binding mode (see Table 1). In a separate set we allowed all interface positions to be varied, termed the ''affinity'' protocol. In both protocols the conserved YY motif (positions 55 and 56) was fixed and only allowed to change rotamers.

For each sequence combination, we optimized the total energy of the complex using a Monte-Carlo simulated annealing protocol similar to that described by Kuhlman & Baker.[31] In brief, a move consisted of a replacement of an amino acid side-chain in the interface at a randomly picked position by a rotamer from the library, followed by a comparison of the previous and current total energies of the complex. The energy function consists of a linear combination of the attractive part of a 6–12 Lennard-Jones potential, a linear repulsive part, an explicit hydrogen bonding potential,[7] an implicit solvation model,[32] statistical terms representing the backbone-dependent internal free energies of amino acid rotamers and amino acid type-specific reference energies, as described.[33] In a second step, we selected the best sequences based on their calculated binding energy computed as by Kortemme & Baker[7] (the total energies of these sequences were also among the lowest sampled).

Construction and cloning of DNase/immunity protein designs

A plasmid for the wild-type E7–Im7 DNase/immunity construct (pHBH, a derivative of pQE30, Qiagen) was a kind gift from Dr Kin-Fu Chak (National Yang Ming University, Taiwan). Designed constructs were cloned by standard methods. An enzymatically inactive variant of each design was created by site-directed mutagenesis to introduce the DNase mutation H569A (distant from the DNase–immunity protein interface). These inactive constructs could be expressed with higher yields[21] and were used for the SPR and fluorescence binding assays and the crystallography.

Purification and separation of complexes

All computationally selected variants were transformed into SG13009 [pREP4] *Escherichia coli* (Qiagen), expressed in 2XTY and purified as described.[21] For SPR and fluorescence binding experiments, the complex was separated on a Ni-NTA (Qiagen) column by eluting the immunity protein with 7 M Gu-HCl. The DNase was then eluted with an imidazole step gradient.

Gel filtration analysis

Gel filtration was carried out with 50 μM protein on a Superdex-75 column (Amersham Biosciences) in 50 mM Tris (pH 7.5), 150 mM NaCl, 1 mM EDTA at 25 °C.

SPR binding analysis

SPR measurements were performed with a BIAcore 2000 (BIAcore AB) in buffer containing 50 mM Mops (pH 7.5), 150 mM NaCl and 0.005% surfactant P20. Protein concentrations were determined from the absorbance at 280 nm, using a calculated molar extinction coefficient.[34] 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 in a range of concentrations from 0.1 nM to 500 nM at 25 °C. The SPR titrations with high response units were used as a qualitative measure to determine the extent of specificity. Extensive discussion of the analysis of the SPR sensograms and resolution of associated problems is in Kortemme *et al.*[21]

Fluorescence binding analysis

Binding can be measured by fluorescence if there is a W residue at the interface, whose intrinsic fluorescence changes between the bound and the unbound states. The cognate G2_68W complex contains a W at position 68, whose intrinsic fluorescence changes upon binding by 30%. The dissociation competition experiments were carried out using manual mixing on a Spex Fluorolog 1681 0.22 m spectrometer with a 292 nm excitation and 345 nm emission wavelength. A preformed 10 μM DNase/immunity protein complex was chased with a threefold excess of WT immunity protein, in 50 mM Mops (pH 7.5) and 150 mM NaCl at 25 °C. The subsequent traces fit to single exponential curves using the Kaleidograph software package (Synergy inc., Reading PA). The rate constant was independent of the concentration of the excess wild-type immunity protein competitor as expected for the dissociation rate constant; the faster association with the competitor observed in stopped-flow experiments[21] was not resolvable in these manual mixing experiments. This same experiment was carried out for the non-cognate E7_WT/Im7_G_68W complex where there is a 20% change in the intrinsic tryptophan fluorescence signal upon binding. A more detailed discussion of the methods used to determine and analyze the fluorescence binding traces can be found in Kortemme *et al.*[21]

Crystallization and data collection

Crystals of the G design complex were grown at room temperature in hanging drops with 1 μl reservoir containing 30% (w/v) PEG 4000, 600 mM $(NH_4)_2$ acetate, 50 mM sodium acetate (pH 4.6), 25% (v/v) glycerol, 5% (v/v) dimethylsulfoxide, mixed with 1 μl of 25–30 mg/ml protein. Crystals (700 μm × 300 μm × 300 μm; space group $I222$; $a = 61.038$, $b = 72.969$, $c = 121.167$) appeared in several days. They were flash frozen in liquid nitrogen. Diffraction data were recorded to 2.0 Å resolution at the Advanced Light Source beamline 5.0.3. For the resolution range (50–2.0 Å) the data were 99% complete with an R_{merge} of 4.6%. There were 38,937 total reflections recorded of which 20,886 were unique. The intensities were integrated using DENZO and SCALEPACK[35] (see Supplementary Table 1 for complete data collection and refinement statistics).

Data refinement and model building

The structure was solved *via* molecular replacement using CNS[36] with the E7–Im7 (pdb:7CEI) complex as the initial search model; the correlation coefficient for the solution was 55%. The structure was modeled in XtalVIew[37] and refined using CNS[36] with a 9.5% data set for cross-validation. The final R_{work} and R_{free} for the E7_G/Im7_G complex were 22.5% and 27%, respectively.

Data bank accession codes

The atomic coordinates have been deposited in the RCSB Protein Data Bank†, with accession code 2ERH.

Acknowledgements

We thank Jim Havranek, John Karanicolas and Vanita Sood for critical reading of the manuscript and the whole Baker laboratory for helpful comments. We thank Tomoko Hamma for collecting the G design data set at the Advanced Light Source 5.0.3 Beam Line and Bin Qian for generating the set of structures with different binding modes along the E7–Im7 to E9–Im9 trajectory. This work was supported by a US National Institutes of Health (NIH) training grant (L.A.J.) and a grant from the NIH (D.B.).

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2006.05.022

References

1. Bogan, A. A. & Thorn, K. S. (1998). Anatomy of hot spots in protein interfaces. *J. Mol. Biol.* **280**, 1–9.
2. Conte, L. L., Chothia, C. & Janin, J. (1999). The atomic structure of protein-protein recognition sites. *J. Mol. Biol.* **285**, 2177–2198.
3. Sharp, K. A. (1998). Calculation of HyHel10-lysozyme binding free energy changes: effect of ten point mutations. *Proteins: Struct. Funct. Genet.* **33**, 39–48.
4. Massova, I. & Kollman, P. A. (1999). Computional alanine scanning to probe protein-protein interactions: a novel approach to evaluate binding free energies. *J. Am. Chem. Soc.* **121**, 8133–8143.
5. Huo, S., Massova, I. & Kollman, P. A. (2002). Computational alanine scanning of the 1:1 human growth hormone-receptor complex. *J. Comput. Chem.* **23**, 15–27.
6. Guerois, R., Nielsen, J. E. & Serrano, L. (2002). Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *J. Mol. Biol.* **320**, 369–387.
7. Kortemme, T. & Baker, D. (2002). A simple physical model for binding energy hot spots in protein-protein complexes. *Proc. Natl Acad. Sci. USA* **99**, 14116–14121.
8. Brannetti, B., Via, A., Cestra, G., Cesareni, G. & Helmer-Citterich, M. (2000). SH3-SPOT: an algorithm to predict preferred ligands to different members of the SH3 gene family. *J. Mol. Biol.* **298**, 313–328.
9. Vaccaro, P., Brannetti, B., Montecchi-Palazzi, L., Philipp, S., Helmer Citterich, M., Cesareni, G. & Dente, L. (2001). 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.
10. Li, W., Hamill, S. J., Hemmings, A. M., Moore, G. R., James, R. & Kleanthous, C. (1998). Dual recognition and the role of specificity-determining residues in colicin E9 DNase-immunity protein interactions. *Biochemistry*, **37**, 11771–11779.
11. Aloy, P. & Russell, R. B. (2002). Interrogating protein interaction networks through structural biology. *Proc. Natl Acad. Sci. USA* **99**, 5896–5901.
12. Wollacott, A. M. & Desjarlais, J. R. (2001). Virtual interaction profiles of proteins. *J. Mol. Biol.* **313**, 317–342.
13. Looger, L. L., Dwyer, M. A., Smith, J. J. & Hellinga, H. W. (2003). Computational design of receptor and sensor proteins with novel functions. *Nature*, **423**, 185–190.
14. Harbury, P. B., Zhang, T., Kim, P. S. & Alber, T. (1993). A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants. *Science*, **262**, 1401–1407.
15. Havranek, J. J. & Harbury, P. B. (2003). Automated design of specificity in molecular recognition. *Nature Struct. Biol.* **10**, 45–52.
16. Reina, J., Lacroix, E., Hobson, S. D., Fernandez-Ballester, G., Rybin, V., Schwab, M. S. *et al.* (2002). Computer-aided design of a PDZ domain to recognize new target sequences. *Nature Struct. Biol.* **9**, 621–627.
17. Shifman, J. M. & Mayo, S. L. (2002). Modulating calmodulin binding specificity through computational protein design. *J. Mol. Biol.* **323**, 417–423.
18. Shifman, J. M. & Mayo, S. L. (2003). Exploring the origins of binding specificity through the computational redesign of calmodulin. *Proc. Natl Acad. Sci. USA* **100**, 13274–13279.
19. Bolon, D. N., Grant, R. A., Baker, T. A. & Sauer, R. T. (2005). Specificity versus stability in computational

† www.rcsb.org

protein design. *Proc. Natl Acad. Sci. USA* **102**, 12724–12729.
20. Ko, T. P., Liao, C. C., Ku, W. Y., Chak, K. F. & Yuan, H. S. (1999). The crystal structure of the DNase domain of colicin E7 in complex with its inhibitor Im7 protein. *Struct. Fold. Des.* **7**, 91–102.
21. Kortemme, T., Joachimiak, L. A., Bullock, A. N., Schuler, A. D., Stoddard, B. L. & Baker, D. (2004). Computational redesign of protein-protein interaction specificity. *Nature Struct. Mol. Biol.* **11**, 371–379.
22. Wallis, R., Leung, K. Y., Pommer, A. J., Videler, H., Moore, G. R., James, R. & Kleanthous, C. (1995). 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.
23. Kuhlmann, U. C., Pommer, A. J., Moore, G. R., James, R. & Kleanthous, C. (2000). Specificity in protein-protein interactions: the structural basis for dual recognition in endonuclease colicin-immunity protein complexes. *J. Mol. Biol.* **301**, 1163–1178.
24. Shakked, Z., Guzikevichguerstein, G., Frolow, F., Rabinovich, D., Joachimiak, A. & Sigler, P. B. (1994). Determinants of repressor-operator recognition from the structure of the Trp operator binding-site. *Nature*, **368**, 469–473.
25. Joachimiak, A., Haran, T. E. & Sigler, P. B. (1994). Mutagenesis supports water mediated recognition in the Trp repressor-operator system. *EMBO J.* **13**, 367–372.
26. Bhat, T. N., Bentley, G. A., Boulot, G., Greene, M. I., Tello, D., Dallacqua, W. *et al.* (1994). Bound water-molecules and conformational stabilization help mediate an antigen-antibody association. *Proc. Natl Acad. Sci. USA* **91**, 1089–1093.
27. Covell, D. G. & Wallqvist, A. (1997). 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.
28. Buckle, A. M., Schreiber, G. & Fersht, A. R. (1994). Protein-protein recognition - crystal structural-analysis of a barnase barstar complex at 2.0-angstrom resolution. *Biochemistry*, **33**, 8878–8889.
29. Braden, B. C., Fields, B. A. & Poljak, R. J. (1995). Conservation of water molecules in an antibody-antigen interaction. *J. Mol. Recogn.* **8**, 317–325.
30. Jiang, L., Kuhlman, B., Kortemme, T. A. & Baker, D. (2005). A ''solvated rotainer'' approach to modeling water-mediated hydrogen bonds at protein-protein interfaces. *Proteins: Struct. Funct. Bioinformat.* **58**, 893–904.
31. Kuhlman, B. & Baker, D. (2000). Native protein sequences are close to optimal for their structures. *Proc. Natl Acad. Sci. USA,* **97**, 10383–10388.
32. Lazaridis, T. & Karplus, M. (1999). Effective energy function for proteins in solution. *Proteins: Struct. Funct. Genet.* **35**, 133–152.
33. Kortemme, T., Morozov, A. V. & Baker, D. (2003). An orientation-dependent hydrogen bonding potential improves prediction of specificity and structure for proteins and protein-protein complexes. *J. Mol. Biol.* **326**, 1239–1259.
34. Gill, S. & von Hippel, P. H. (1989). Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* **182**, 319–326.
35. Otwinowski, Z. A. M. (1997). Processing of X-ray data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326.
36. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al.* (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallog. sect. D,* **54**, 905–921.
37. McRee, D. E. (1999). XtalView/Xfit–a versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol.* **125**, 156–165.

Edited by C. R. Matthews

(Received 9 February 2006; received in revised form 6 May 2006; accepted 10 May 2006)
Available online 24 May 2006