



Structure of the Ultra-High-Affinity Colicin E2 DNase–Im2 Complex

Justyna Aleksandra Wojdyla¹, Sarel J. Fleishman^{2,3},
David Baker^{3,4} and Colin Kleanthous^{1*}

¹Department of Biology, University of York, Wentworth Way, Heslington, York YO10 5DD, UK

²Department of Biological Chemistry, Weizmann Institute of Science, Rehovot 76100, Israel

³Department of Biochemistry, University of Washington, Seattle, WA 98195, USA

⁴Howard Hughes Medical Institute, University of Washington, Seattle, WA 98195, USA

Received 2 November 2011;
received in revised form
10 January 2012;
accepted 13 January 2012
Available online
27 January 2012

Edited by R. Huber

Keywords:

protein–protein interaction;
colicin;
immunity protein;
high-affinity binding

How proteins achieve high-affinity binding to a specific protein partner while simultaneously excluding all others is a major biological problem that has important implications for protein design. We report the crystal structure of the ultra-high-affinity protein–protein complex between the endonuclease domain of colicin E2 and its cognate immunity (Im) protein, Im2 ($K_d \sim 10^{-15}$ M), which, by comparison to previous structural and biophysical data, provides unprecedented insight into how high affinity and selectivity are achieved in this model family of protein complexes. Our study pinpoints the role of structured water molecules in conjoining hotspot residues that govern stability with residues that control selectivity. A key finding is that a single residue, which in a noncognate context massively destabilizes the complex through frustration, does not participate in specificity directly but rather acts as an organizing center for a multitude of specificity interactions across the interface, many of which are water mediated.

© 2012 Elsevier Ltd. All rights reserved.

Introduction

How proteins distinguish specific protein binding partners from a variety of structural and functional homologues is a fundamental problem in molecular biology. Being able to tailor the specificity of any given protein–protein interaction (PPI) so that unwanted binding partnerships are avoided would have significant biotechnological and biomedical applications by, for example, reducing off-target effects in protein therapeutics and producing highly specific protein diagnostics. While the physicochem-

ical basis for complex formation and selectivity is understood for many model PPI systems,^{1,2} it is still a major challenge to integrate this information base as a starting point for rational and computational design.^{3,4} Nevertheless, major advances have been reported in the *de novo* design of PPIs, as well as in the redesign of naturally occurring complex specificity, in many instances incorporating negative design to enhance selectivity.^{4–10} The power of such approaches was shown by Fleishman *et al.* in their recent *de novo* design of proteins targeting the conserved stem region of influenza hemagglutinin, with the resulting binary complex structures closely matching those designed computationally.¹¹ The strategy adopted in this case involved computing important amino acid hotspot residues onto a guest scaffold and then optimizing shape complementarity and affinity from which high-affinity binders were isolated.¹¹ While the computational design of hot spots in PPIs is becoming increasingly common,

*Corresponding author. E-mail address: colin.kleanthous@york.ac.uk.

Abbreviations used: Im, immunity; PPI, protein–protein interaction; IPE, immunity protein exosite; PDB, Protein Data Bank.

such approaches invariably omit or ignore intervening waters, since these cannot be readily modeled or their contributions to stability and specificity readily predicted. In large part, this reflects the uncertainties regarding the role of buried water molecules in PPIs. Here, we describe the structure of an ultra-high-affinity PPI, which, when viewed in the context of previous structural and biophysical data, clearly delineates the importance of interfacial water molecules to both the stability and specificity of this model PPI.

How PPIs can have high affinities while maintaining high selectivity is one of the major questions in the field. This is emphasized by the protein design literature, where the best designed protein–protein complexes only achieve ~ 1000 -fold discriminations between cognate and noncognate complexes,^{12,13} although greater selectivities can be achieved by directed evolution.^{14,15} Among all macromolecular assemblies, PPIs are unique in exhibiting equilibrium dissociation constants (K_d values) that span the millimolar-to-femtomolar affinity range.¹⁶ Such varied PPI binding affinities underpin diverse biological functions such as electron transfer between redox proteins, antibody recognition of protein antigens, hormone recognition by receptors, intracellular signaling, and inhibition of hydrolytic enzymes. An important step in understanding the linkage between the structure and energetics of different protein–protein complexes has been taken by Kastriitis *et al.*, who collated and analyzed a nonredundant data set of 144 complexes for which K_d data spanning 10^{-5} – 10^{-14} M have been reported.¹⁷ Even so, our understanding of how protein complexes can achieve such varied affinities and exhibit high levels of discrimination remains rudimentary.

The present work focuses on the interactions of colicin DNases with immunity (Im) proteins, one of the few PPI systems that span the millimolar-to-femtomolar affinity range. Colicins are a widespread group of plasmid-encoded three-domain protein antibiotics released by *Escherichia coli* following environmental stress as a means of killing neighboring closely related organisms during competition for resources. Cell killing is mediated by a C-terminal cytotoxic domain that is translocated to the cytoplasm of a susceptible bacterium following the binding of the colicin to receptor and translocator proteins in the outer membrane and contact with inner-membrane proteins.¹⁸ The cytotoxic domains of endonuclease E colicins (ColE2, ColE7, ColE8, and ColE9) are 15-kDa domains that belong to the H–N–H/ $\beta\beta\alpha$ -Me class of nucleases,^{19,20} eliciting cell death through random degradation of the bacterial genome.²¹ Colicin-producing *E. coli* avoid suicide through the action of a small 10-kDa Im protein that binds with high affinity to an immunity protein exosite (IPE) on the enzyme, inhibiting its activity

through steric and electrostatic occlusion of substrate DNA binding.²² Im proteins share $\sim 50\%$ sequence identity, while colicin DNases share $\sim 65\%$ sequence identity. An important consequence of exosite binding is that much of the sequence diversity in these proteins is found at the protein–protein interface,²³ a result of the positive selection for novel colicin DNase–Im variants between competing bacterial populations. These properties, along with the extensive characterization of colicin DNase–Im protein complexes reported in the literature, have led to their being adopted as a model system for investigating the coevolution of PPIs,²⁴ the development of NMR-based methods for structure determination of PPIs,²⁵ testing of the latest methodologies for computational docking of PPIs,²⁶ molecular dynamics simulations to follow PPI association,²⁷ and the directed evolution and design of PPI specificity.^{4,28–30}

Cognate colicin DNase–Im protein complexes are high-affinity PPIs, exhibiting K_d values of $\sim 10^{-14}$ – 10^{-16} M, while noncognate complexes, which also inhibit nuclease activity, display binding that is 6–10 orders of magnitude weaker.^{31–34} Mutational and biophysical analyses,^{32,35–39} along with crystal structures of various complexes,^{22,40–42} have provided a structural and thermodynamic framework for understanding how specificity is encoded in these high-affinity PPIs. Im proteins use conserved and variable helices to bind the hypervariable IPE on the DNase surface through a ‘dual-recognition’ mechanism. Binding affinity is dominated by a conserved hot spot, a common feature of PPIs,^{32,43,44} comprising two tyrosine residues (Tyr54 and Tyr55) and an aspartic acid (Asp51) in helix III of the Im protein, with neighboring variable specificity residues in helix II (centering on position 33) making positive, neutral, or negative contributions to specificity. Similar dual-recognition mechanisms have since been reported to underpin specificity in the associations of many other PPIs, including a bacterial chemoreceptor binding CheR methyltransferase,⁴⁵ animal toxins binding voltage-gated potassium channels,⁴⁶ IL-13 binding the IL-13 receptor,⁴⁷ and regulators of G-protein signaling binding G proteins,⁴⁸ although the molecular details in each case differ.

We have reported previously the crystal structures of the colicin E9 DNase in complex with its cognate Im protein Im9 and with the noncognate partner Im2, which differ in binding affinity by 7 orders of magnitude.^{22,42} Structural comparisons and computational analysis highlighted the importance of chemical ‘frustration’ at the center of the DNase–Im protein complex, whereby a destabilizing specificity contact (Im2 Asp33...E9 DNase Phe86) is tolerated due to the highly stabilizing interactions of the conserved Im protein hot spot. The frustrated interface is thus primed for high-affinity binding, which can be relieved through the mutation of a

limited number of interface residues, in particular residue 33. While these studies provided molecular insight into how the colicin E9 DNase discriminates between Im2 and Im9, they did not explain how the colicin E2 DNase binds Im2 specifically. To address this question, we determined the crystal structure of the E2 DNase–Im2 complex, which has a K_d of 10^{-15} M at pH 7, 25 °C, and 200 mM NaCl.³² Viewed in the context of previous biochemical and biophysical data on colicin DNase–Im protein complexes, along with structural comparisons to free Im2 and other colicin DNase–Im protein complexes, the study provides a mechanistic basis for the 8 -orders-of-magnitude Im2/Im9 discrimination exhibited by the colicin E2 DNase and gives one of the most complete pictures yet of how specificity is encoded in high-affinity PPIs.

Results and Discussion

Structure of the E2 DNase–Im2 complex

The crystal structure of the E2 DNase–Im2 complex was solved to a resolution of 1.7 Å and refined to an R -factor of 16.2% ($R_{\text{free}} = 20.2\%$) (Fig. 1a

and Table 1), following coexpression and purification of the complex (see Materials and Methods). Electron density maps allowed the fitting of all E2 DNase and Im2 residues, with the exception of the N-terminal residues of both proteins. The structure contains 319 solvent molecules, as well as 1 calcium ion and 1 metal ion. The H–N–H/ $\beta\beta\alpha$ -Me motif of colicin DNases is known to bind a variety of divalent metal ions.^{50–52} Hence, we performed a wide-range absorption energy scan to identify the metal ion bound to the E2 DNase domain, which revealed that the majority of sites are occupied by zinc, presumably acquired during the expression of the E2 DNase–Im2 complex in bacterial cells. A small amount of nickel was also detected, most likely coming from the nickel affinity purification step. We have shown previously that the active-site motif in the colicin E9 DNase binds zinc and nickel ions with nanomolar and micromolar affinities, respectively.⁵¹ Based on the energy scan and previous biophysical data on metal binding to the H–N–H/ $\beta\beta\alpha$ -Me motif of these enzymes, a Zn^{2+} with an occupancy of 1 was modeled into the electron density map of the E2 DNase–Im2 complex where, as in E7 and E9 DNases,^{53,54} it is coordinated by three of the four histidine residues of the motif.

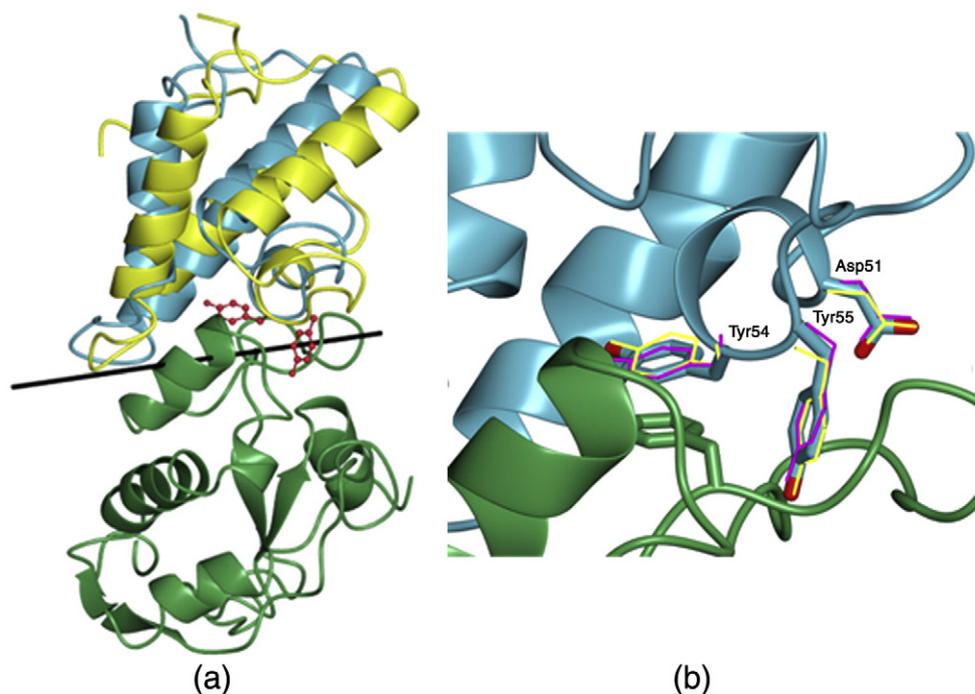


Fig. 1. Structure of the E2 DNase–Im2 complex and its comparison to other cognate colicin–Im protein complexes. (a) Structure of the E2 DNase (green) bound to the Im2 (cyan) shown in ribbons and highlighting the two hotspot tyrosine residues of Im2: Tyr54 and Tyr55. Rigid-body rotations of Im proteins with respect to the superposed DNase domain in the E2 DNase–Im2 and E7 DNase–Im7 complexes were analyzed with the program DynDom3D.⁴⁹ Im7 is shown in yellow, while the rotation axis is shown with a black line. For clarity, only the E2 DNase (green) is shown. (b) Hotspot residues at the interface of the E2 DNase–Im2 complex are identically positioned in cognate colicin DNase–Im complexes. Im2 Asp51, Tyr54, and Tyr55 are shown as blue cylinders. Equivalent residues of E9 DNase–Im9 (magenta; PDB ID: 1emv) and E7 DNase–Im7 (yellow; PDB ID: 1mz8) complexes are shown as thin lines (DNase domains were superposed).

Table 1. Data collection and refinement statistics for the E2 DNase–Im2 complex

<i>Data collection</i>	
Space group	<i>P</i> 2 ₁ 2 ₁ 2
Unit cell parameters <i>a</i> , <i>b</i> , <i>c</i> (Å)	121.81, 53.28, 32.78
Wavelength (Å)	0.87260
Resolution range (Å)	50.0–1.71 (1.82–1.71) ^a
Mean <i>I</i> / σ (<i>I</i>)	16.4 (3.23) ^a
<i>R</i> _{sym} ^b (linear) (%)	8.0 (42.2) ^a
Redundancy	6.8 (5.4) ^a
Number of observations	160,403
Number of unique reflections	23,541
Completeness (%)	99.0 (94.0) ^a
<i>Refinement</i>	
Resolution range (Å)	48.8–1.72
Number of working/free reflections	22,213/1206
Number of protein residues	92 (A), 132 (B)
Number of zinc ions	1
Number of calcium ions	1
Number of water molecules	319
<i>R</i> _{work} ^c / <i>R</i> _{free} ^d (%)	16.2/20.2
<i>B</i> average (Å ²)	18.2
RMSD from ideal values	0.025
RMSD bond lengths (Å)	1.0
RMSD bond angles (°)	

^a Numbers given in parentheses are from the last-resolution shell.

^b $R_{\text{sym}} = (S_{hkl} S_i | I_i(hkl) - \langle I(hkl) \rangle) / S_{hkl} S_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of the *i*th measurement of reflection (*hkl*), and $\langle I(hkl) \rangle$ is the average intensity.

^c $R_{\text{work}} = (S_{hkl} | F_o - F_c |) / S_{hkl} F_o$, where F_o and F_c are the observed and calculated structure factors, respectively.

^d R_{free} is calculated as for R_{work} but from a randomly selected subset of the data (5%) that were excluded from refinement.

ColE2 was the first colicin to be identified as a DNase, and Im2 was the first reported Im protein to inhibit DNase activity⁵⁵; however, up to now, no structure for the enzyme or its complex with Im2 has been reported. Our structure of the E2 DNase–Im2 complex is the third cognate colicin DNase–Im protein complex to be solved, with the structure showing many similarities but also many unique features that are detailed in the following sections. The complexed E2 DNase has a mixed α/β fold similar to those of the unbound DNase domains of ColE7 [Protein Data Bank (PDB) ID: 1m08] and ColE9 (PDB ID: 1fsj), with RMSDs of 0.66 Å (for 129 C α atoms) and 0.7 Å (for 130 C α atoms), respectively. In both of these latter cases, the enzyme undergoes only minor structural changes upon binding the Im protein, from which we surmise that Im2 binding most likely also causes little or no change in the structure of the enzyme. Bound Im2 contains four α -helices (I–IV), with a characteristic short helix III, and is very similar to the solution structure of free Im2 (PDB ID: 2no8). However, some notable differences in C α trace were identified between residues 22 and 57, including helices I–III and loops I and II. Many of the residues in this region are involved in the formation of the interface

with the E2 DNase (a point we return to later). In the present study, Im2 was purified via a C-terminal noncleavable His₆ tag, which is clearly visible in the electron density map. Since attempts to crystallize the tag-less E2 DNase–Im2 complex failed, it implies that the presence of His₆ tag facilitates crystallization and stabilizes crystal contacts. The His₆ tag (residues 89–94) interacts through four direct hydrogen bonds with symmetry-related E2 DNase (His89–Arg96) and Im2 (His94–Asp51, His94–Gln30, and His94–Gln31) molecules. Importantly, the His₆ tag residues make no contribution to the protein–protein interface.

Comparison with other structures of colicin DNase–Im protein complexes and the role of rotation in defining specificity

Twenty-three residues from the E2 DNase and 27 residues from Im2 are involved in the formation of the complex, resulting in the burial of 1697 Å² of accessible surface area at the interface, the largest of all the colicin DNase–Im protein complexes solved to date (the second largest being that of the noncognate E9 DNase–Im2 complex; 1566 Å²). Overall, the E2 DNase–Im2 complex is very similar in structure to the cognate E9 DNase–Im9 (PDB ID: 1emv) and E7 DNase–Im7 (PDB ID: 1mz8) complexes, as well as to the noncognate E9 DNase–Im2 (PDB ID: 2wpt) complex, with RMSDs of 0.96 Å (for 212 C α atoms), 1.53 Å (for 212 C α atoms), and 1.00 Å (for 195 C α atoms), respectively. The DNase domains of ColE2, ColE7, and ColE9 in all four complexes superpose very well, with minor differences in the loops located far from the complex interface. All four Im proteins also superpose well (RMSD <1 Å), with minor structural differences observed in the C-terminal region of helix I, loop I, and loop III. The structural similarity of the complexes can be readily appreciated by the close superposition of the helix III hotspot residues of Im2, Im7, and Im9 (Asp51, Tyr54, and Tyr55) when bound to their cognate enzymes (Fig. 1b).

The interface formed between the basic colicin E2 DNase and acidic Im2 shows a high degree of shape ($S_c = 0.72$; adapted from Lawrence and Colman⁵⁶) and charge complementarity equivalent to those of other colicin DNase–Im protein complexes (Table 2). The interfaces of colicins E2 and E7 DNases with their cognate Im proteins are the most polar of the four complex structures reported, involving a similar high number of interfacial direct hydrogen bonds and buried water molecules (Tables 2–4). Of the three cognate complexes, the colicin E9 DNase–Im9 complex is the least polar and involves the least number of interfacial hydrogen bonds and buried water molecules. These global characteristics are consistent with the thermodynamic signatures of cognate Im protein binding; E2 and E7 DNases are

Table 2. Comparison of colicin DNase–Im complex interfaces

Complex	Buried surface area (Å ²)	Direct hydrogen bonds	One hydrogen bond per buried area (Å ²)	S _c	Buried waters
E2–Im2	1697	18	94	0.72	7
E9–Im2	1566	15	104	0.75	8
E9–Im9	1500	13	115	0.71	5
E7–Im7	1370	19	72	0.71	7

All quoted values, with the exception of E2 DNase–Im2 from this work, were adapted from Meenan *et al.*⁴²

Buried surface area was calculated with the PISA server.⁵⁷ Direct hydrogen bonds and buried waters were calculated with the programs CONTACT and AREAIMOL, respectively, from the CCP4 program suite. Protein surface complementarity (S_c) was analyzed with the program Sc.⁵⁶

strongly enthalpically driven associations that are entropically disfavored, while the E9 DNase–Im9 complex is weakly enthalpically driven but entropically favored.³¹ However, these structure-based interpretations of thermodynamic data breakdown for the noncognate colicin E9 DNase–Im2 complex which, despite having structural properties similar to a cognate complex (similar buried surface area, number of direct hydrogen bonds and buried water

molecules, and high degree of complementarity), has a thermodynamic signature characteristic of most of the noncognate complexes, which tend to be weakly enthalpically driven and entropically favored.³¹

Kuhlmann *et al.* reported previously that Im7 and Im9 are related by a 19° rotation axis when bound to their cognate enzymes due to rigid-body rotations centered on helix III.⁴¹ The rotation enables different regions of the specificity helix II to contact the enzyme while maintaining all the conserved interactions of the hotspot residues.⁴¹ The importance of such ‘rotamer’ states to the evolution of Im protein specificity has been demonstrated by directed evolution experiments in which Im9 was evolved toward ColE7 specificity.³⁰ A newly evolved nevIm7 protein exhibited a ColE7-bound conformation intermediate between that of Im9 and that of Im7 bound to their cognate enzymes. Using the program DynDom3D,⁴⁹ we compared the rotamer status of Im2 with those of other Im proteins bound to cognate enzymes. This analysis revealed that Im2 and Im7 are displaced relative to each other by a 13.4° rotation axis that passes through the helix III hotspot residues on the Im protein and E2 DNase Phe86, the key specificity site on the enzyme (Fig.

Table 3. Direct hydrogen bonds at the interfaces of the E2 DNase–Im2, E9 DNase–Im9, and E9 DNase–Im2 complexes

	E2 DNase residue	Im2 residue	Distance (Å)	E9 DNase–Im9	E9 DNase–Im2
1	NH1 Arg54	OE2 Glu30	2.86	Yes (2.84) ^a	Yes, NH1-OE1 (2.91)
2	NH2 Arg54	OE1 Glu30	2.85	Yes (3.02) ^a	Yes, NH2-OE2 (2.80)
3^b	NZ Lys72	O Pro56	2.78	No, Lys72 → <i>Asn72</i>	No, Lys72 → <i>Asn72</i>
4^b	NZ Lys72	OD1 Asp58	2.74	No, Lys72 → <i>Asn72</i>	No, Lys72 → <i>Asn72</i>
5^b	N Gly73	OD2 Asp62	2.85	No, Gly73 → <i>Pro73</i>	No, Gly73 → <i>Pro73</i> ND2 Asn72-OD1 and OD2 Asp62 (3.13 and 3.17)
6	NZ Lys83	O Ala25	3.08	No, different conformation (Lys83 → <i>Tyr83</i>)	No, OH <i>Tyr83</i> -A2017 (2.72)-O Ala25 (2.76)
7	NZ Lys83	O Gly27	2.94	Yes, OH <i>Tyr83</i> -O <i>Thr27</i> (2.98)	No, OH <i>Tyr83</i> -A2017 (2.72)-O Gly27 (2.56)
8	N Ala84	OE2 Glu30	3.16	Yes, N <i>Ser84</i> -OE2 Glu30 (2.92)	Yes, N <i>Ser84</i> -OE2 Glu30 (2.94)
9	O Phe86	OH Tyr55	2.64	Yes (2.68)	Yes (2.71)
10	NZ Lys89	OE1 Glu41	2.83	No, Lys89 facing away; OE2 Glu41-NZ Lys97 (3.22)	Yes, but also OE1 Glu41-NZ Lys97 (2.87)
11	N Lys89	OD1 Asp51	2.77	Yes (2.83)	Yes (2.79)
12	NE2 Gln92	OG Ser50	3.00	Yes (2.90)	Yes (2.92)
13^b	O Gly95	NH2 Arg38	3.52	No, Arg38 → <i>Thr38</i>	No, Arg38 facing away O Cys95-ND2 Asn34 (2.89)
14^b	OE1 Glu97	NH2 Arg38	2.95	No, Arg38 → <i>Thr38</i> Lys97 interacts with Glu41 (specificity interaction) ^a	No, Arg38 facing away (because Glu31 → <i>Cys31</i>); Lys97 interacts with Glu41
15^b	OE2 Glu97	NE Arg38	2.74	No, Arg38 → <i>Thr38</i> ; Lys97 interacts with Glu41 (specificity interaction) ^a	No, Arg38 facing away (because Glu31 → <i>Cys31</i>)
16^b	NH2 Arg98	O Glu30	3.45	No, Arg98 → <i>Val98</i>	No, Arg98 → <i>Val98</i>
17^b	NH2 Arg98	OD1 Asn34	2.98	No, Asn34 → <i>Val34</i>	No, Arg98 → <i>Val98</i> and Asn34 facing away (ND2 Asn34-O Cys95)
18^b	NE Arg98	OD1 Asn34	2.87	No, Asn34 → <i>Val34</i>	No, Arg98 → <i>Val98</i> and Asn34 facing away (ND2 Asn34-O Cys95)

^a Salt bridges in the E9 DNase–Im9 complex (adapted from Kuhlmann *et al.*⁴¹); hydrogen bonds conserved in the E2 DNase–Im2, E9 DNase–Im9, and E9 DNase–Im2 complexes are labeled in boldface; residues that are not conserved between DNase–Im complexes are italicized.

^b Specificity interaction in E2 DNase–Im2.

Table 4. Water-mediated hydrogen bonds in structurally characterized DNase–Im complexes

Number	Water	E9 DNase–Im9	E9 DNase–Im2	E7 DNase–Im7	E2 DNase and Im2 residues	Distance (Å)	B-value (Å ²)	Accessible surface area (Å ²)
1	32 N Lys72B O Tyr54A	Yes, 87A (0.53)	Yes, 2039A (0.36)	Yes, 606A (0.77)	ND2 Asn75B 3.01 2.81	2.98	9.11	<5
2	33 O Tyr54A	No	No	No	N Asn75B 2.81	2.76	10.97	0
3	34 O Ile53A OD1 Asp62A	No	No	Yes, 605A (1.18)	N Ser74B 2.73 2.68	2.97	11.44	0
4	36 O Cys23A	No	No	No	OG1 Thr77B 2.73	2.82	15.42	>10
5	42 OH Tyr54A O Ile22A	No	No	No	ND2 Asn78B 2.72 2.77	3.05	11.54	0
6	43 NH2 Arg98B OD2 Asp33A	No	No	No	OD1 Asn78B 2.84 2.78	2.71	9.74	0
7	55 OE2 Glu30A	No	No	No	O Gly82B 2.67	2.77	14.30	>10
8	87 OD1 Asp51A OG Ser50A	Yes, 88A (0.59)	Yes, 2037A (0.39)	Yes, 611A (0.86)	O Ala87B 2.74 2.75	2.94	11.24	0
9	88 OD2 Asp51A	Yes, 147A (0.5)	Yes, 2038A (0.24)	Yes, 656A (0.46)	N Lys90B 2.69	2.96	13.72	>10
10	89 OD2 Asp51A	No	No	No	NH1 Arg88B 3.07	2.71	21.67	>10
11	135 OE2 Glu41	No	No	No	OE1 Gln92 2.67	2.79	20.13	<5
12	224 OE1 Glu97 ND2 Asn34	Yes, 418B (0.49) interacts only with DNase	Yes, 2029A (0.31)	Yes, 616B (0.55) interacts only with DNase	N Glu97B 2.94 3.10	2.79	22.01	>10
13	225 OD1 Asp62	No	Yes, 2049A (1.76)	No	OG Ser74B 2.98	2.81	27.31	<10
14	251 OE2 Glu26A	No	Yes, 2015A (0.89)	No	NZ Lys81B 2.63	3.04	34.90	>10

Waters highlighted in boldface occupy conserved positions and are shown in Fig. 4. Numbers in parentheses show displacement (Å) from the water position in the E2 DNase–Im2 complex.

1a).³⁵ Although not as displaced as Im9 relative to Im7, the effect is similar, allowing more of helix II in Im2 to be exposed to the DNase surface for specificity interactions with the enzyme. Interestingly, the E2 DNase–Im2 and E9 DNase–Im9 complexes are related by a modest 5.6° rotation axis that, unusually, does not pass through the hotspot residues but instead connects Leu16 in helix I with Ser50 in helix III of the Im protein (data not shown). This rotation causes a shift in helix II residues by 1.0–1.5 Å. Hence, in this instance, rotation serves to fine-tune the position of common specificity sites rather than exposing or sequestering distinct regions of the specificity helix II.

We have argued previously that the biphasic association kinetics observed for all colicin DNase–Im protein complexes in pre-steady-state experiments are the result of rigid-body rotations of the Im protein on the DNase surface, following rapid formation of an electrostatically steered intermediate centered on helix III.³⁶ Such a kinetic mechanism underpins the dual recognition of colicin DNases by Im proteins, where they can be broadly cross-

reactive (yielding noncognate complexes) yet highly specific for a particular colicin. A comparison of the orientations of all Im proteins for which structures have been solved while bound to colicin DNases shows a distribution of rotamer states (Fig. 2a). We speculate that this distribution in crystal structures reflects rigid-body rotations on the DNase surface in solution, with specific rotamers becoming ‘frozen out’ due to specificity contacts unique to particular complexes. This in turn suggests that rotations should persist in noncognate colicin DNase–Im protein complexes in solution, consistent with the complex dissociation kinetics observed for such complexes.^{34,36}

We note that the ability of Im proteins to adopt distinct rotamer states on the DNase surface is facilitated by the architecture of the IPE itself. If the IPE were flat and undulating, as for many PPIs,⁵⁸ then surface rotation would be limited by steric clashes. The IPE, in contrast, is shaped like a bow tie, composed of a central convex ‘collar’ flanked by two wider ‘bows’ of varying dimensions (Fig. 2b). Helix III of the Im protein docks in the cleft at the base

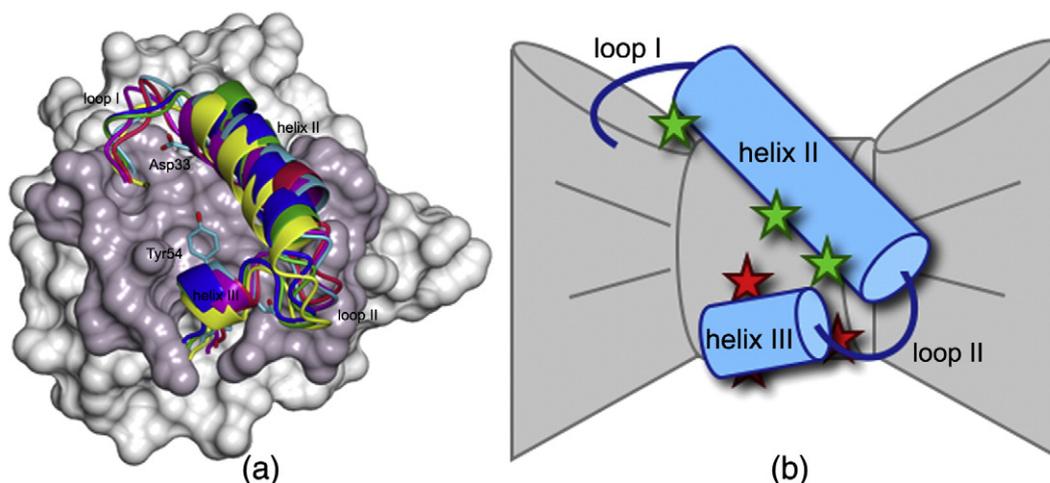


Fig. 2. Binding of Im proteins to the colicin DNase IPE. (a) Rotamer distribution of Im proteins bound to colicin DNases. The figure shows a molecular surface representation of the E2 DNase identifying that buried by Im2—with overlays of helices II and III, including adjoining loops I and II—for six Im proteins for which the structures of bound complexes have been determined. Cognate complex Im9 is shown in magenta, Im7 is shown in yellow, Im2 is shown in light blue, noncognate Im2 (E9 DNase–Im2 complex) is shown in red, nevIm7 R12-2 (PDB ID: 3gkl) is shown in blue, and nevIm7 R12-13 (PDB ID: 3gjn) is shown in green. Also shown are key residues in helix III of Im2 that clamp the Im protein to the base of the binding site (Asp51, Tyr54, and Tyr55) and the specificity residue Asp33, which lies close to loop I (see the text for details). The figure emphasizes the different rotameric states that Im proteins can adopt when bound to colicin DNases. (b) Cartoon depicting the ‘bow tie’ of the colicin DNase IPE and the main contact points of the Im protein. Helix III (red stars) forms water-mediated hotspot interactions with the collar of the bow tie, while helix II (green stars) forms a range of specificity contacts with the ribbons of the bow tie, some of which are water mediated (data not shown), as in the case of the E2 DNase–Im2 complex reported here.

between the two bows, clamped to the collar through the conserved hotspot interactions of Asp51, Tyr54, and Tyr55 (Fig. 2a). Helix II and its adjoining loops, which are also involved in specificity interactions, lie diagonally across the bow tie. Importantly, rotation of the Im protein around the central collar juxtaposes different specificity residues of the Im protein with distinct parts of the two bows while simultaneously maintaining all the conserved hotspot interactions and avoiding steric and electrostatic clashes with the enzyme.

These aspects of the DNase IPE are most readily appreciated by comparing the buried surfaces of the individual IPEs when bound to their cognate Im proteins (Fig. 3). The DNase IPE of colicins E2, E7, and E9 spans a near-contiguous 30-residue sequence, only five of which are invariant among the colicin DNases; Asn75, Gly82, Pro85, Gly94, and Arg96. The two glycines and proline have structural roles, and Arg96, which points towards the active site away from the Im protein, is involved in DNA binding.⁵⁴ Asn75 is the only conserved residue involved in stabilizing the complex with the Im protein, but this is indirect via a hydrogen bond with an intervening water molecule (W32) (Fig. 4b). While this interaction is conserved in all colicin DNase–Im complexes, its contribution to stabilization is nevertheless context dependent; mutation of E9 DNase Asn75 to Ala

yields $\Delta\Delta G_{\text{binding}}$ values of 2.3 and 1.2 kcal/mol for Im9 and Im2 binding, respectively.³⁵ Hence, even conserved Asn75 contributes to the specificity of the colicin DNase–Im protein complex although the thermodynamic basis for this is unclear at present. Asn75 is part of the collar of the IPE bow tie (Fig. 3), which, interestingly, given its sequence variation, has similar dimensions in all colicin DNase–Im protein complexes (~ 8 Å wide at its narrowest point). Another intriguing aspect of the IPE bow-tie architecture is how it is able to accommodate the conserved residues of helix III of the Im protein. Hydrogen bonds either to the main chain (Tyr55 to the carbonyl of the DNase specificity residue 86) or are mediated by water molecules. The bulky side chains of the Im protein tyrosine hotspot residues Tyr54 and Tyr55 are accommodated by van der Waals interactions. Tyr54 of the Im protein contacts the DNase specificity residue at position 86,^{35,41} which is centrally located in the collar of the bow tie (Fig. 3). In the case of the E2 DNase–Im2 complex, this contact directs E2 DNase Phe86 towards Im2 Val37 in helix II but also Im2 Asp33 (Fig. 5), a potentially ‘frustrating’ interaction in this high-affinity complex (a point we return to below). Tyr55 of the Im protein slots into the cleft at the base of the bow tie, the dimensions of which vary in the different cognate complexes (Figs. 2 and 3). In the E2

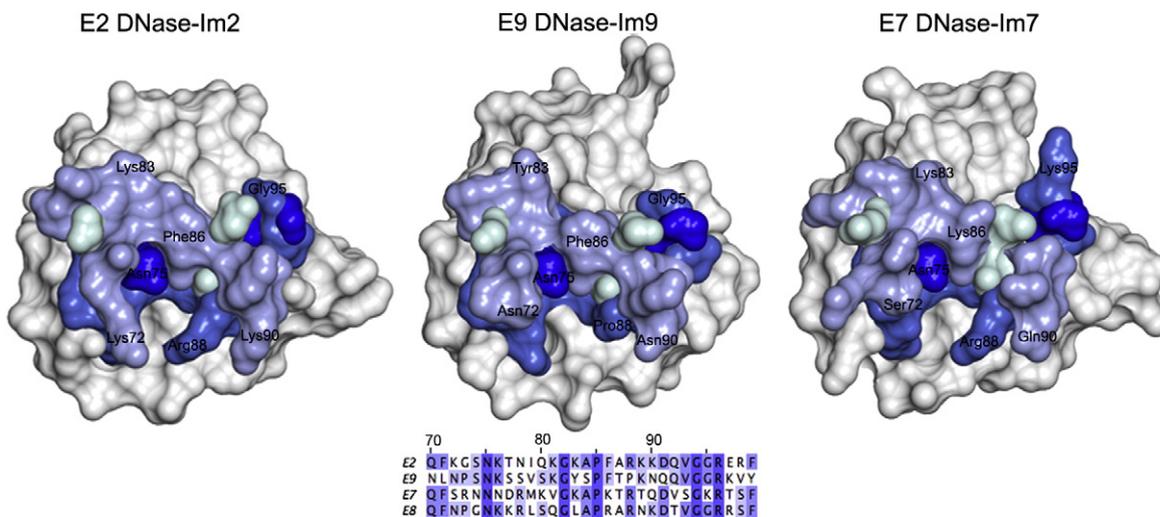


Fig. 3. Im proteins bury overlapping but distinct surfaces in their complexes with colicin DNases. Surface representations of colicin DNase IPEs for colicins E2 (present work), E9, and E7 showing regions that are buried by their respective cognate Im proteins and colored according to sequence conservation (dark blue, conserved; blue, conserved in three colicin DNase–Im complexes; light blue, conserved in two colicin DNase–Im complexes; cyan, variable residues). Shown below the structures is a sequence alignment of the contiguous 30-residue IPEs for all four colicin DNases, with the color scheme corresponding to that in the structures.

DNase–Im2 complex, the cleft is bound by DNase residues Lys72 and Arg88; in E9 DNase–Im9, these residues are bound by Asn72 and Pro88; and in E7 DNase–Im7, the residues are bound by Ser72 and Arg88. In each case, the phenyl ring of Tyr55 only makes clear van der Waals interactions with the residue at position 88.

Interfacial water molecules mediate the stability and specificity of the colicin E2 DNase–Im2 complex

A total of 14 water molecules mediate hydrogen bonds between the E2 DNase and Im2, of which seven are almost completely buried at the interface (Table 4). In addition to the three conserved waters described below, an additional water molecule is conserved only in the complexes of Im2 with the E2 and E9 DNases. The remainder all mediate interactions between the residues implicated in the specificity of the colicin E2 DNase–Im2 complex, the roles of some of which are detailed below.

Since most PPIs take place in aqueous environments, water is invariably involved in the thermodynamics of PPIs, but to what extent interfacial water molecules resolved in crystal structures are involved in stabilizing complexes and/or mediating specificity remains controversial.^{1,58,59} A number of authors have noted that tightly packed hotspot regions of PPIs tend to be devoid of interfacial waters, implying that water entropy effects provide one of the thermodynamic driving forces for complex formation and, indeed, this has been used to predict the location of hot spots.⁶⁰ It is striking

therefore that the conserved helix III hot spot of the colicin E2 DNase–Im2 complex, centering on Asp51, Tyr54, and Tyr55, involves side-chain or main-chain hydrogen bonds with water molecules (W32, W87, and W88) that are conserved in both cognate and noncognate colicin DNase–Im complexes alike (Fig. 4). Moreover, the same water molecules are also present in the structure of the unbound E9 DNase (PDB ID: 1fsj). Hence, these structurally resolved waters are indeed involved in stabilizing colicin DNase–Im complexes, significantly increasing the number of interfacial hydrogen bonds around the hotspot residues. Given their strategic placement between the helix III hotspot residues of the Im protein and the collar of the IPE bow tie, it is also conceivable that, unlike direct hydrogen bonds that would place restrictions on rotation, they act as pivot points for the rotation of the Im protein on the DNase surface, thereby facilitating the docking of specificity residues along helix II.

Overview of specificity interactions in DNase–Im protein complexes

In the following, we summarize the specificity interactions that distinguish E2 from E9 DNase binding by their respective Im proteins Im2 and Im9, since more complete biophysical data are available for these complexes. Of the four cognate and noncognate complexes that can be formed, only that between the E2 DNase and Im9 has not been structurally characterized (repeated crystallization experiments have failed to yield diffraction-quality crystals). We therefore generated a model for this

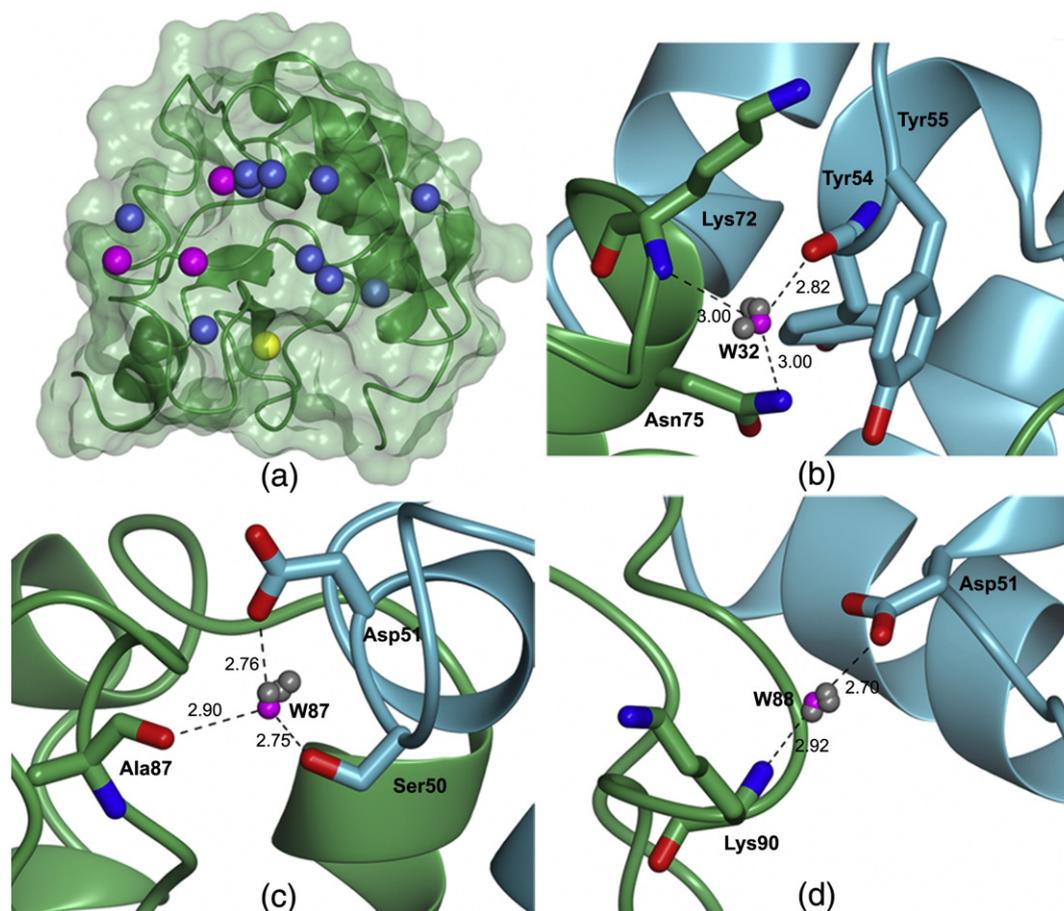


Fig. 4. Conserved and variable water molecules at the colicin E2 DNase–Im2 complex interface. (a) Fourteen water molecules mediate interprotein hydrogen bonds (blue, magenta, and yellow) at the E2 DNase–Im2 interface. Only the structure of the E2 DNase is shown. Four waters occupy conserved positions in the E2 DNase–Im2, E9 DNase–Im9, E9 DNase–Im2 (noncognate), and E7 DNase–Im7 complexes. Three waters mediate hydrogen bonds in all four complexes (magenta), while one mediates hydrogen bonds in the E2 DNase–Im2 and E9 DNase–Im2 complexes (yellow). (b–d) Hydrogen-bond networks of conserved water molecules (magenta; equivalent to those in (a)) that mediate interactions between E2 DNase (green) and Im2 (cyan) residues. Gray spheres represent equivalent water molecules in the structures of all other colicin DNase–Im protein complexes determined to date.

complex using the program Rosetta (see [Materials and Methods](#)) in order to make comparisons with the two cognate complexes and with one noncognate complex. Although the buried surface areas in all four complexes are comparable, the computed binding energy is less favorable in the E2 DNase–Im9 model (-23 Rosetta energy units) than in the cognate complexes (~ -31 Rosetta energy units), likely reflecting its poorer charge and shape complementarity being leading to its affinity being 7–8 orders of magnitude lower.

At the core of the interfaces of all four DNase–Im complexes ([Figs. 3 and 5](#)) is residue 86, which is a phenylalanine in both colicins E2 and E9 (but lysine and arginine in E7 and E8, respectively). Although Phe86 is conserved in the two DNases and, in each case, the aromatic ring packs against Tyr54 in the Im protein, it nevertheless makes differential contribu-

tions to Im protein binding specificity. In studies focused on the E9 DNase, a Phe86Ala mutation had a much greater effect on Im9 *versus* Im2 binding ($\Delta\Delta G_{\text{binding}} \sim 4$ and 1 kcal/mol, respectively³⁵). This differential effect stems from the residues surrounding the Phe86–Tyr54 pair in each complex. In the E9 DNase–Im9 complex, hydrophobic residues (E9 DNase Val98, Im9 Leu33, Val37, and Val34) surround the pair, while in the E2 DNase–Im2 complex, charge/hydrophilic residues predominate (E2 DNase Val98, Im2 Asp33 and Asn34). In the case of E2 DNase–Im2, the aliphatic carbon chain of E2 DNase Arg98 stacks against Phe86, the charged guanidinium group hydrogen bonding to Im2 Asn34 across the interface ([Fig. 5a](#)). In the noncognate E9 DNase–Im2 complex, the side-chain positions of the residues surrounding the central specificity pair remain essentially as they appear in

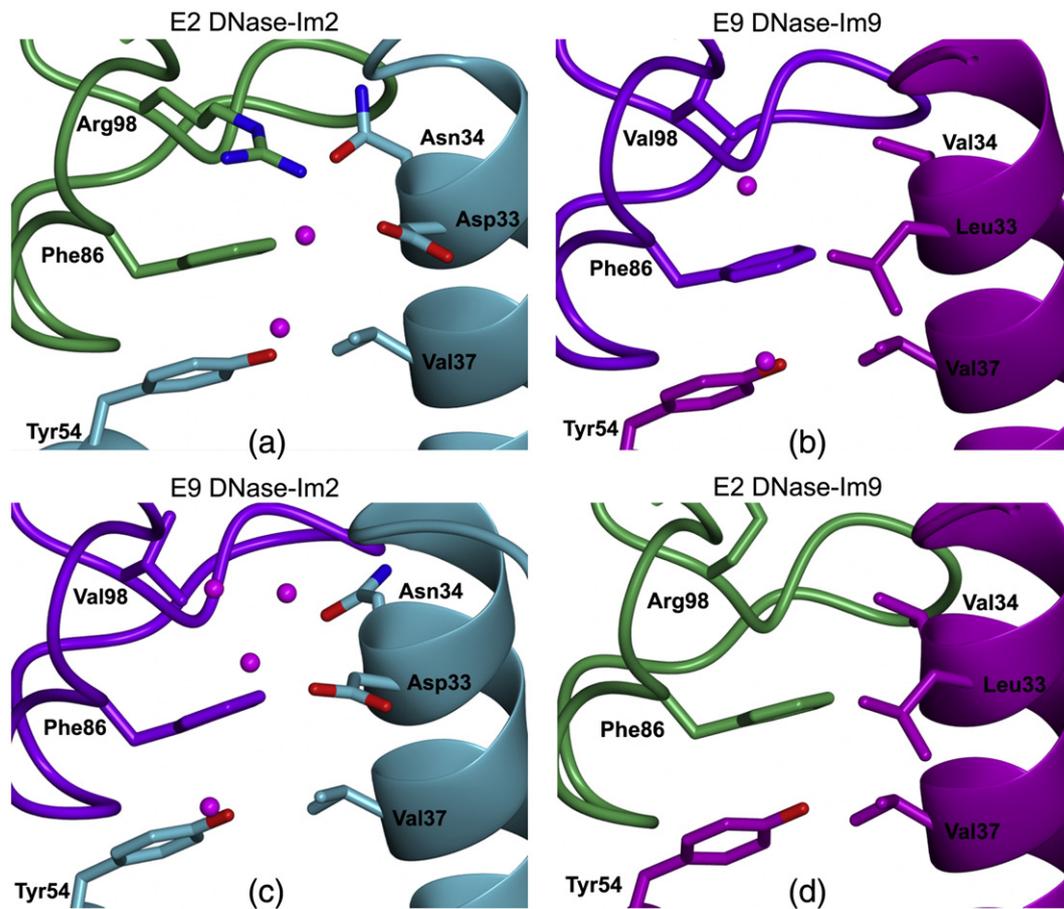


Fig. 5. Comparison of core specificity interactions in cognate and noncognate colicin E2 and E9 DNase–Im protein complexes. (a) E2 DNase–Im2 complex. (b) E9 DNase–Im9 complex (PDB ID: 1emv). (c) E9 DNase–Im2 complex (PDB ID: 2wpt). (d) Modeled E2 DNase–Im9 complex. The figure highlights the central specificity contact in all four complexes of the hotspot tyrosine of the Im protein (Tyr54) with the key specificity residue in each DNase (Phe86) and the surrounding specificity residues from helix II of the Im protein and the DNase, as well as interfacial water molecules. See the text for details.

the respective cognate complexes, except that additional water molecules fill the cavity left by the guanidinium moiety of E2 DNase Arg98, with frustration being a result of the forced colocalization of E9 DNase Phe86 and Im2 Asp33 (Fig. 5c).⁴² Frustration can be partially relieved by an alanine mutation at Im2 Asp33 (~100-fold improvement in binding³⁵), with mutation to leucine (the cognate residue in Im9) yielding a 10,000-fold increase in binding.³⁸ Our model of the E2 DNase–Im9 complex, which does not predict the placement of interfacial water molecules, indicates that the charged guanidinium group of E2 DNase Arg98 rotates away from the E9 DNase Phe86–Im9 Tyr54 pair, although we cannot preclude the possibility that this also does not change its position (potentially leading to a frustrated complex) and involves intervening water molecules (Fig. 5d).

While position 33 in this combination of complexes serves a key role in defining colicin DNase

PPI specificity,^{32,41,42} other regions within helix II have also been found to contribute to specificity.^{32,38}

In phage display experiments, where residues in helix II were randomly mutated in Im2 and selected for binding the E2 DNase, Arg38 and Glu41 were the next most selected residues after position 33, although their contribution to binding free energy appears small (<1 kcal/mol³²). Examination of the E2 DNase–Im2 structure reveals that Im2 Glu41 forms a single hydrogen bond with E2 DNase Lys89, while Im2 Arg38 forms bifurcated hydrogen bonds with E2 DNase Glu97 and Im2 Glu31 (Fig. 6a). In the E9 DNase–Im9 complex, Im9 Glu31 is too distant from the residue at position 38 (Thr38) to form an interaction, with its side chain rotating ~120° away from residues in helix II. E9 DNase Lys97 is within hydrogen-bonding distance of Im9 Glu41 (although the geometry is not ideal), while the side chain of E9 DNase Lys89 shifts position by 2.6 Å, leaving it unable to contact Im9 Glu41 directly (Fig. 6b). In the

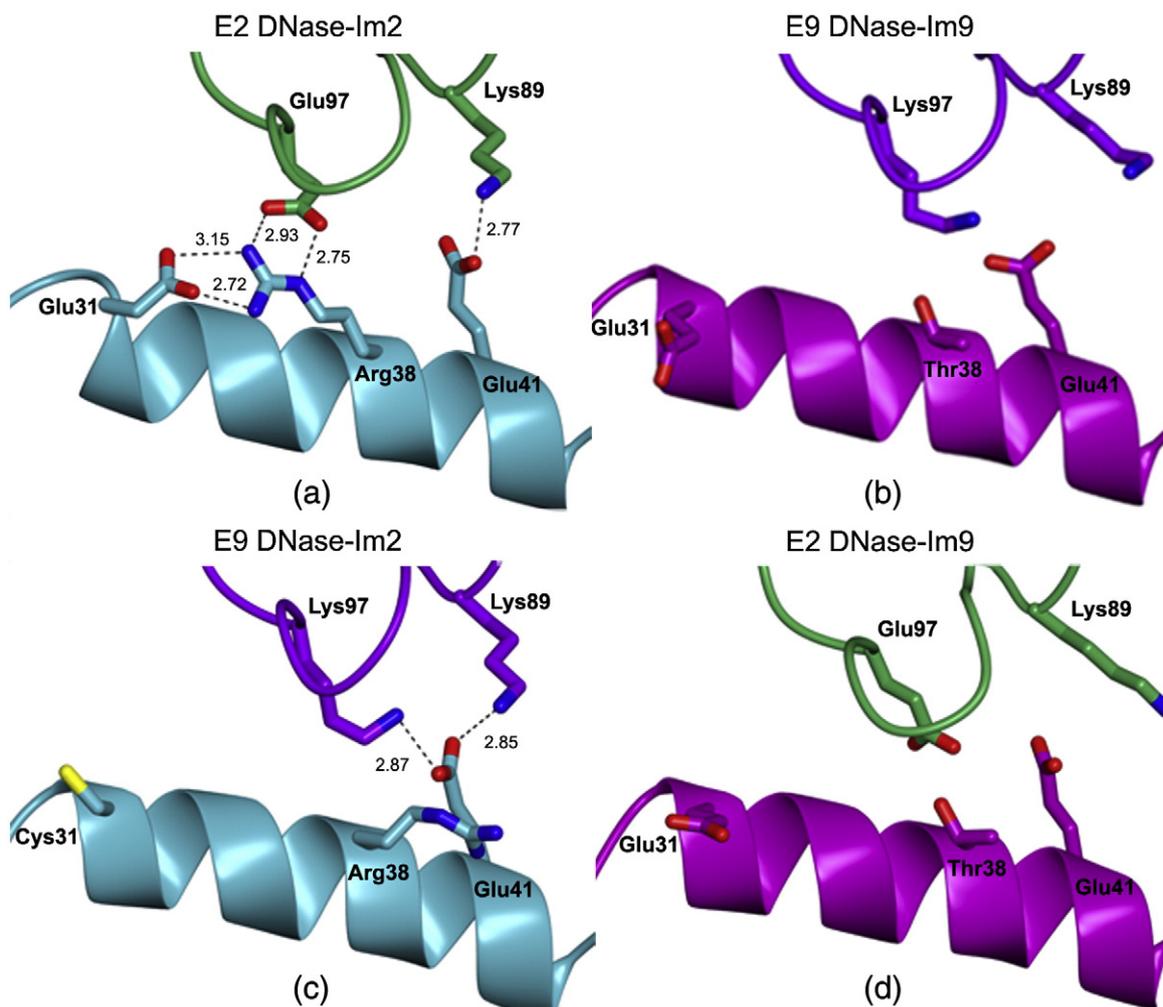


Fig. 6. Peripheral specificity contacts in cognate and noncognate colicin E2 and E9 DNase–Im protein complexes. (a) E2 DNase–Im2 complex. (b) E9 DNase–Im9 complex (PDB ID: 1emv). (c) E9 DNase–Im2 complex (PDB ID: 2wpt). (d) Modeled E2 DNase–Im9 complex. The figure highlights hydrogen-bonding interactions between Im protein helix II residues within the four cognate and noncognate DNase complexes. See the text for details.

noncognate E9 DNase–Im2 complex, a repulsive charge interaction causes Im2 Arg38 to rotate away from E9 DNase Lys97 and to face the solvent, while Im2 Glu41, similar to cognate complexes, hydrogen bonds E9 DNase Lys89 and Lys97 (Fig. 6c). In our model of the noncognate E2 DNase–Im9 complex, the incongruous electrostatics in this region of helix II are immediately apparent; the negatively charged side chain of E2 DNase Glu97 faces Im9 Thr38 and Glu41 (Fig. 6d), likely contributing to discrimination between these complexes. In summary, the hydrophobic and electrostatic complementarity surrounding the Phe86–Tyr54 contact in the center of PPI, along with contributions from intervening water molecules and residues at the C-terminal end of helix II of the Im protein, helps sculpt specific colicin DNase binding by Im proteins.

The central roles of Asp33 and water in defining Im2 specificity for colicin E2 DNase

The varied contributions of helix II residues to the stabilization of the cognate complex is one of the distinguishing features of Im2/Im9 binding their specific colicin DNases. In Im9, Leu33, Val34, and Val37, which surround the DNase Phe86–Im Tyr54 specificity contact (Fig. 5b), make small but similar contributions to stabilization relative to the hotspot residues of Asp51, Tyr54, and Tyr55, where alanine mutations destabilize the complex by >5 kcal/mol,³⁹ and are only modestly selected for in Im9 phage display experiments.³² In Im2, however, an Asp33Ala mutation destabilizes the complex almost as much as one of the conserved hotspot residues, and aspartic acid is very strongly selected for in Im2 phage display experiments.³² Given its clear

importance to the specific binding of the colicin E2 DNase, it is surprising that Asp33 forms no direct hydrogen bonds or salt bridges with the enzyme, as anticipated by these earlier studies. Instead, our structure of the E2 DNase–Im2 complex shows that Asp33 coordinates a multitude of specificity contacts with the E2 DNase that require subtle changes in Im2 conformation and involve water molecules. The loss of these contacts explains the loss in binding free energy when the residue is mutated to alanine and its strong selection in phage display experiments.

Im2 undergoes minor but critical changes in structure upon binding its cognate partner E2 DNase, but not the noncognate E9 DNase. In particular, the C-terminus of Im2 helix I and its adjoining loop I become reconfigured upon binding to the enzyme (Fig. 7a), adopting a conformation that is not represented by the solution ensemble of unbound Im2 determined previously by NMR spectroscopy (PDB ID: 2no8). Moreover, binding of Im2 to the E2 DNase requires repacking of its hydrophobic core, with Phe18 and Phe40 (from helices I and II, respectively) adopting substantially different conformations (data not shown). Such repacking is not observed in the

cognate E9 DNase–Im9 complex. Formation of the E2 DNase-bound conformation of Im2 also involves the loss of two backbone hydrogen bonds at the C-terminal end of helix I in free Im2 (Lys21–Ala25 and Cys23–Glu26), reorientation of loop I, and formation of a new intramolecular hydrogen bond between the backbone atoms of Ile22 and Ala25 (Fig. 7b). This relatively minor structure change, aided by repacking of the hydrophobic core, shortens helix I and leads to the displacement of Ala25 by $>5 \text{ \AA}$ to become part of loop I, where, along with the main-chain oxygen of Gly27, it hydrogen bonds the amino group of the E2 DNase specificity residue Lys83 (Fig. 7c). A similar loop I backbone interaction is seen in the E9 DNase–Im9 complex,⁴¹ but in this instance, the DNase residue is Tyr83 (not Lys83), and loop I does not undergo significant changes in conformation in order to accommodate it. Remarkably, the new orientation of Im2 loop I in its complex with the E2 DNase is stabilized by the key specificity residue in Im2, Asp33, which forms a new hydrogen bond with the main-chain nitrogen of Gly27 (Fig. 8a). Uniquely, Asp33 becomes the centerpiece of a network of 12 hydrogen bonds that connects specificity and conserved hotspot residues across

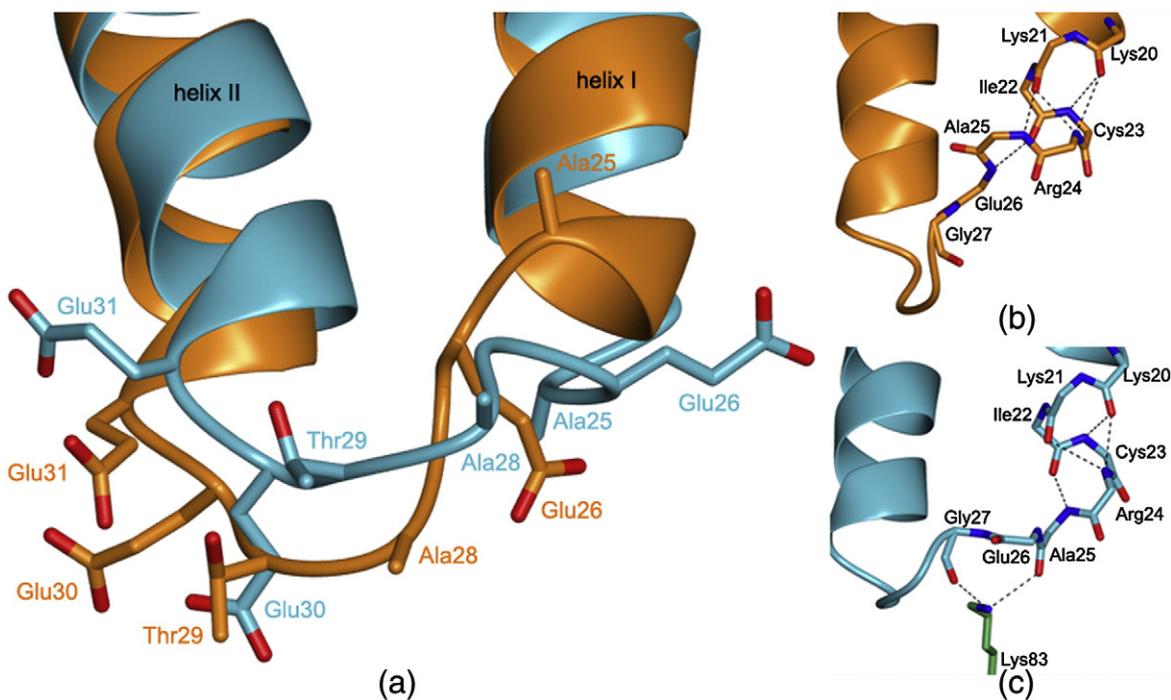


Fig. 7. Localized conformational changes in helix I and loop I of Im2 following binding to the E2 DNase enable backbone interactions with a DNase specificity residue. (a) Superposition of Im2 (PDB ID: 2no8; orange), determined by NMR spectroscopy, with the bound conformation of Im2 (cyan) showing the reorientation of loop I caused by E2 DNase binding. (b) Structure of unbound Im2, as in (a), detailing hydrogen bonds at the C-terminus of helix I. (c) Structure of E2 DNase-bound Im2, as in (a), showing the reorganization of hydrogen bonds at the base of helix I and the main-chain interaction of loop I residues with the specificity residue Lys83 on the E2 DNase.

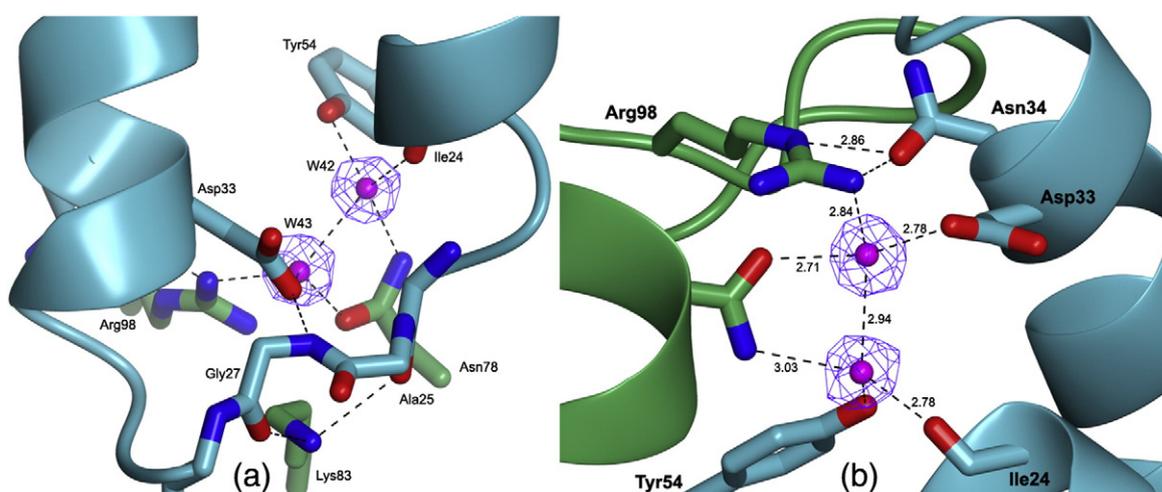


Fig. 8. The key Im2 specificity residue Asp33 in helix II organizes the protein's specificity interactions but does not engage the colicin E2 DNase directly. (a) Im2 Asp33 (cyan) helps stabilize the bound conformation of loop I by forming a hydrogen bond with the backbone of the nitrogen of Gly27. Also shown are the E2 DNase (green) specificity residue Arg98 and interfacial water molecules that contribute to specificity. (b) Alternate view of the figure in (a) showing the hydrogen-bond network surrounding the specificity waters W42 and W43 (magenta balls; electron density map $2F_o - F_c$ shown at 1σ) and involving Asp33 and other specificity residues from both E2 DNase (Asn78 and Arg98) and Im2 (Asn34).

the interface (Fig. 8a and b). In addition to its interaction with loop I, it also forms a hydrogen bond with one of two water molecules (W43) that are hydrogen bonded to other Im2 residues (the backbone carbonyl of Ile24, Asn34, and Tyr54) and E2 DNase residues Asn78 and Arg98, with the latter also hydrogen bonded to Im2 Asn34. We conclude that E2 DNase–Im2 specificity is the result of a complex network of hydrogen bonds at the PPI that interconnects specificity sites with conserved hotspot residues, all mediated by structured water molecules.

In summary, the present structure adds a new dimension to our understanding of specificity among ultra-high-affinity PPIs that will be challenging to incorporate into protein design methodology. First and foremost, water molecules contribute to forming the hot spot of the E2 DNase–Im2 interface, which dominates the binding free energy of the complex. But waters also play an essential role in sculpting the specificity of the complex, which is not so evident in other colicin DNase–Im protein complexes where, as in the case of E9 DNase–Im9, hydrophobic contacts largely govern specificity. Importantly, the study highlights how a single amino acid (Im2 Asp33) not only plays a pivotal role in destabilizing noncognate complexes⁴² but also plays an indirect but nonetheless essential role in the coordination of specificity interactions with its cognate partner that involve water molecules. It is this combination of factors that allows the E2 DNase to selectively bind Im2 over Im9 by almost 8 orders of magnitude.

Materials and Methods

Protein expression and purification

DNA sequence encoding the E2 DNase–Im2 complex was cloned in tandem into the expression vector pET21d (Novagen) in such a way that the E2 sequence was followed by a 2-bp frame shift and Im2, which contained a C-terminal noncleavable His₆ tag. The E2 DNase–Im2 complex was expressed in the *E. coli* strain BL21 DE3 pLysS and purified essentially as previously described.^{51,61} Cultures were grown in LB at 37 °C until OD_{600} = 0.6–0.8, induced with 1 mM IPTG, and left shaking overnight. Harvested cells were stored at –20 °C. Pellet from the 2-L cell culture was resuspended in 20 ml of lysis buffer [50 mM Tris (pH 7.5), 50 mM NaCl, 10 mM imidazole, and 1 mM MgCl₂], lysed by sonication, and centrifuged at 13,000 rpm. The supernatant was loaded onto 4 ml of Ni-NTA beads (Qiagen) equilibrated with lysis buffer, and the protein was eluted (monitoring absorbance at 280 nm) with buffer containing 50 mM Tris (pH 7.5), 50 mM NaCl, and 500 mM imidazole and then dialyzed overnight against 50 mM Tris (pH 7.5) and 500 mM NaCl. The protein sample was subsequently loaded onto a Superdex75 16/60 gel-filtration column (GE Healthcare) equilibrated in 50 mM Tris (pH 7.5) and 500 mM NaCl and eluted in the same buffer, and then fractions were pooled, dialyzed against 50 mM Tris (pH 7.5), and stored at –20 °C. The purity of the complex was determined by Coomassie-stained SDS-PAGE gel to be >99%. The mass of each protein was determined by electrospray ionization mass spectrometry (Technology Facility, York), in each case the observed mass being within 2 Da of the expected mass (Im2, 11,054 Da; E2 DNase, 15,322 Da).

Crystallization and structure determination

Crystallization trials with the E2 DNase–Im2 complex at 25 mg/ml (extinction coefficient, $22,460 \text{ M}^{-1} \text{ cm}^{-1}$) were performed using hanging-drop vapor diffusion. Crystals were obtained in 0.1 M MMT (DL-malic acid, 4-morpholineethanesulfonic acid, and Tris–NaOH mixed in a 1:2:2 molar ratio) at pH 7.0 and 27% polyethylene glycol 1500. A single crystal from the crystallization drop was directly transferred into liquid nitrogen. Single-wavelength X-ray diffraction data containing 360 images were collected from a single crystal at 100 K at European Synchrotron Radiation Facility beamline ID23-2 using a MARMOSAIC 225 CCD detector. Crystal-to-detector distance was kept at 210.7 mm, with an oscillation range of 0.5° . The crystal belonged to space group $P2_12_12$ with unit cell dimensions $a=121.81 \text{ \AA}$, $b=53.28 \text{ \AA}$, and $c=32.78 \text{ \AA}$. Recorder images were processed with XDS.⁶² Reflection intensities were processed with COMBAT and scaled with SCALA⁶³ from the CCP4 program suite.⁶⁴ The structure was determined by molecular replacement with the program MOLREP⁶⁵ using the E9 DNase–Im9 complex structure as search model (PDB ID: 1emv). The solution contained one E2 DNase–Im2 complex in the asymmetric unit. The molecular replacement solution was used as preliminary model for ARP/wARP,⁶⁶ and refinement was carried out using the program REFMAC5.⁶⁷ The structure was visualized and rebuilt into electron density using the program Coot,⁶⁸ and the stereochemistry of the model was evaluated with the program MolProbity.⁶⁹ Data collection and refinement statistics are shown in Table 1. Atomic coordinates and structural amplitudes have been deposited in the PDB (PDB ID: 3u43).

Model building the noncognate E2 DNase–Im9 complex

Our model for the E2 DNase–Im9 complex was constructed by superimposing the DNases of the E2 DNase–Im2 and E9 DNase–Im9 complexes and by extracting an E2 DNase–Im9 hybrid. The Rosetta protocol FastRelax was then used to relax the structure by conducting eight iterations of full side-chain repacking and all-atom minimization over side-chain, backbone, and rigid-body degrees of freedom. Thirty separate trajectories were run, out of which the best model was selected by computed for binding energy.

Acknowledgements

We thank Renata Kaminska (York) and Berni Strongitharm (deceased) of the Technology Facility for mass spectrometry data. This work was supported by the Biotechnology and Biological Sciences Research Council of the UK (grant BB/G020671/1) and carried out with the support of the European Synchrotron Radiation Facility.

References

1. Reichmann, D., Rahat, O., Cohen, M., Neuvirth, H. & Schreiber, G. (2007). The molecular architecture of protein–protein binding sites. *Curr. Opin. Struct. Biol.* **17**, 67–76.
2. Schreiber, G. & Keating, A. E. (2011). Protein binding specificity versus promiscuity. *Curr. Opin. Struct. Biol.* **21**, 50–61.
3. Karanicolas, J. & Kuhlman, B. (2009). Computational design of affinity and specificity at protein–protein interfaces. *Curr. Opin. Struct. Biol.* **19**, 458–463.
4. Kortemme, T., Joachimiak, L. A., Bullock, A. N., Schuler, A. D., Stoddard, B. L. & Baker, D. (2004). Computational redesign of protein–protein interaction specificity. *Nat. Struct. Mol. Biol.* **11**, 371–379.
5. Bolon, D. N., Grant, R. A., Baker, T. A. & Sauer, R. T. (2005). Specificity versus stability in computational protein design. *Proc. Natl Acad. Sci. USA*, **102**, 12724–12729.
6. Grigoryan, G., Reinke, A. W. & Keating, A. E. (2009). Design of protein–interaction specificity gives selective bZIP-binding peptides. *Nature*, **458**, 859–864.
7. Havranek, J. J. & Harbury, P. B. (2003). Automated design of specificity in molecular recognition. *Nat. Struct. Biol.* **10**, 45–52.
8. Karanicolas, J., Corn, J. E., Chen, I., Joachimiak, L. A., Dym, O., Peck, S. H. *et al.* (2011). A *de novo* protein binding pair by computational design and directed evolution. *Mol. Cell*, **42**, 250–260.
9. Potapov, V., Reichmann, D., Abramovich, R., Filchitski, D., Zohar, N., Ben Halevy, D. *et al.* (2008). Computational redesign of a protein–protein interface for high affinity and binding specificity using modular architecture and naturally occurring template fragments. *J. Mol. Biol.* **384**, 109–119.
10. Skerker, J. M., Perchuk, B. S., Siryaporn, A., Lubin, E. A., Ashenberg, O., Goulian, M. & Laub, M. T. (2008). Rewiring the specificity of two-component signal transduction systems. *Cell*, **133**, 1043–1054.
11. Fleishman, S. J., Whitehead, T. A., Ekiert, D. C., Dreyfus, C., Corn, J. E., Strauch, E. M. *et al.* (2011). Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. *Science*, **332**, 816–821.
12. Yajima, A., van Brussel, A. A., Schripsema, J., Nukada, T. & Yabuta, G. (2008). Synthesis and stereochemistry–activity relationship of small bacteriocin, an autoinducer of the symbiotic nitrogen-fixing bacterium *Rhizobium leguminosarum*. *Org. Lett.* **10**, 2047–2050.
13. Yosef, E., Politi, R., Choi, M. H. & Shifman, J. M. (2009). Computational design of calmodulin mutants with up to 900-fold increase in binding specificity. *J. Mol. Biol.* **385**, 1470–1480.
14. Baan, R. A., Frijmann, M., van Knippenberg, P. H. & Bosch, L. (1978). Consequences of a specific cleavage in situ of 16-S ribosomal RNA for polypeptide chain elongation. *Eur. J. Biochem.* **87**, 137–142.
15. Bonsor, D. A. & Sundberg, E. J. (2011). Dissecting protein–protein interactions using directed evolution. *Biochemistry*, **50**, 2394–2402.
16. Kleantous, C. (2000). *Protein–Protein Recognition*. *Frontiers in Molecular Biology*. Oxford University Press, Oxford, UK.

17. Kastriitis, P. L., Moal, I. H., Hwang, H., Weng, Z., Bates, P. A., Bonvin, A. M. & Janin, J. (2011). A structure-based benchmark for protein-protein binding affinity. *Protein Sci.* **20**, 482–491.
18. Cascales, E., Buchanan, S. K., Duche, D., Kleanthous, C., Lloubes, R., Postle, K. *et al.* (2007). Colicin biology. *Microbiol. Mol. Biol. Rev.* **71**, 158–229.
19. Keeble, A. H., Maté, M. J. & Kleanthous, C. (2005). HNH endonucleases. In (Belfort, M., Derbyshire, V., Stoddard, B. & Wood, D., eds), *Homing Endonucleases and Inteins*, 16, pp. Springer-Verlag, Berlin, Germany.
20. Kuhlmann, U. C., Moore, G. R., James, R., Kleanthous, C. & Hemmings, A. M. (1999). Structural parsimony in endonuclease active sites: should the number of homing endonuclease families be redefined? *FEBS Lett.* **463**, 1.
21. Pommer, A. J., Cal, S., Keeble, A. H., Walker, D., Evans, S. J., Kuhlmann, U. C. *et al.* (2001). Mechanism and cleavage specificity of the H-N-H endonuclease colicin E9. *J. Mol. Biol.* **314**, 735.
22. Kleanthous, C., Kuhlmann, U. C., Pommer, A. J., Ferguson, N., Radford, S. E., Moore, G. R. *et al.* (1999). Structural and mechanistic basis of immunity toward endonuclease colicins. *Nat. Struct. Biol.* **6**, 243–252.
23. Kleanthous, C. & Walker, D. (2001). Immunity proteins: enzyme inhibitors that avoid the active site. *Trends Biochem. Sci.* **26**, 624–631.
24. Goh, C. S. & Cohen, F. E. (2002). Co-evolutionary analysis reveals insights into protein-protein interactions. *J. Mol. Biol.* **324**, 177–192.
25. Montalvao, R. W., Cavalli, A., Salvatella, X., Blundell, T. L. & Vendruscolo, M. (2008). Structure determination of protein-protein complexes using NMR chemical shifts: case of an endonuclease colicin-immunity protein complex. *J. Am. Chem. Soc.* **130**, 15990–15996.
26. Janin, J. (2010). Protein-protein docking tested in blind predictions: the CAPRI experiment. *Mol. Biosyst.* **6**, 2351–2362.
27. Baron, R., Wong, S. E., de Oliveira, C. A. & McCammon, J. A. (2008). E9-Im9 colicin DNase-immunity protein biomolecular association in water: a multiple-copy and accelerated molecular dynamics simulation study. *J. Phys. Chem. B*, **112**, 16802–16814.
28. Bernath, K., Magdassi, S. & Tawfik, D. S. (2005). Directed evolution of protein inhibitors of DNase by *in vitro* compartmentalization (IVC) and nano-droplet delivery. *J. Mol. Biol.* **345**, 1015–1026.
29. Joachimiak, L. A., Kortemme, T., Stoddard, B. L. & Baker, D. (2006). Computational design of a new hydrogen bond network and at least a 300-fold specificity switch at a protein-protein interface. *J. Mol. Biol.* **361**, 195–208.
30. Levin, K. B., Dym, O., Albeck, S., Magdassi, S., Keeble, A. H., Kleanthous, C. & Tawfik, D. S. (2009). Following evolutionary paths to protein-protein interactions with high affinity and selectivity. *Nat. Struct. Mol. Biol.* **16**, 1049–1055.
31. Keeble, A. H., Kirkpatrick, N., Shimizu, S. & Kleanthous, C. (2006). Calorimetric dissection of colicin DNase-immunity protein complex specificity. *Biochemistry*, **45**, 3243–3254.
32. Li, W., Keeble, A. H., Giffard, C., James, R., Moore, G. R. & Kleanthous, C. (2004). Highly discriminating protein-protein interaction specificities in the context of a conserved binding energy hotspot. *J. Mol. Biol.* **337**, 743–759.
33. 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 non-cognate interactions that span the millimolar to femtomolar affinity range. *Biochemistry*, **34**, 13751.
34. Wallis, R., Moore, G. R., James, R. & Kleanthous, C. (1995). Protein-protein interactions in colicin E9 DNase-immunity protein complexes: 1. Diffusion-controlled association and femtomolar binding for the cognate complex. *Biochemistry*, **34**, 13743–13750.
35. Keeble, A. H., Joachimiak, L. A., Mate, M. J., Meenan, N., Kirkpatrick, N., Baker, D. & Kleanthous, C. (2008). Experimental and computational analyses of the energetic basis for dual recognition of immunity proteins by colicin endonucleases. *J. Mol. Biol.* **379**, 745–759.
36. Keeble, A. H. & Kleanthous, C. (2005). The kinetic basis for dual recognition in colicin endonuclease-immunity protein complexes. *J. Mol. Biol.* **352**, 656–671.
37. Li, W., Dennis, C. A., Moore, G. R., James, R. & Kleanthous, C. (1997). Protein-protein interaction specificity of Im9 for the endonuclease toxin colicin E9 defined by homologue-scanning mutagenesis. *J. Biol. Chem.* **272**, 22253–22258.
38. 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.
39. Wallis, R., Leung, K. Y., Osborne, M. J., James, R., Moore, G. R. & Kleanthous, C. (1998). Specificity in protein-protein recognition: conserved Im9 residues are the major determinants of stability in the colicin E9 DNase-Im9 complex. *Biochemistry*, **37**, 476.
40. 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. *Structure*, **7**, 91–102.
41. 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.
42. Meenan, N. A., Sharma, A., Fleishman, S. J., Macdonald, C. J., Morel, B., Boetzel, R. *et al.* (2010). The structural and energetic basis for high selectivity in a high-affinity protein-protein interaction. *Proc. Natl Acad. Sci. USA*, **107**, 10080–10085.
43. Halperin, I., Wolfson, H. & Nussinov, R. (2004). Protein-protein interactions; coupling of structurally conserved residues and of hot spots across interfaces. Implications for docking. *Structure*, **12**, 1027–1038.
44. Ma, B., Elkayam, T., Wolfson, H. & Nussinov, R. (2003). Protein-protein interactions: structurally conserved residues distinguish between binding sites and exposed protein surfaces. *Proc. Natl Acad. Sci. USA*, **100**, 5772–5777.
45. Shiomi, D., Zhulin, I. B., Homma, M. & Kawagishi, I. (2002). Dual recognition of the bacterial chemoreceptor

- by chemotaxis-specific domains of the CheR methyltransferase. *J. Biol. Chem.* **277**, 42325–42333.
46. Gilquin, B., Braud, S., Eriksson, M. A., Roux, B., Bailey, T. D., Priest, B. T. *et al.* (2005). A variable residue in the pore of Kv1 channels is critical for the high affinity of blockers from sea anemones and scorpions. *J. Biol. Chem.* **280**, 27093–27102.
 47. Lupardus, P. J., Birnbaum, M. E. & Garcia, K. C. (2010). Molecular basis for shared cytokine recognition revealed in the structure of an unusually high affinity complex between IL-13 and IL-13Ralpha2. *Structure*, **18**, 332–342.
 48. Kosloff, M., Travis, A. M., Bosch, D. E., Siderovski, D. P. & Arshavsky, V. Y. (2011). Integrating energy calculations with functional assays to decipher the specificity of G protein–RGS protein interactions. *Nat. Struct. Mol. Biol.* **18**, 846–853.
 49. Poornam, G. P., Matsumoto, A., Ishida, H. & Hayward, S. (2009). A method for the analysis of domain movements in large biomolecular complexes. *Proteins*, **76**, 201–212.
 50. Keeble, A. H., Hemmings, A. M., James, R., Moore, G. R. & Kleanthous, C. (2002). Multistep binding of transition metals to the H–N–H endonuclease toxin colicin E9. *Biochemistry*, **41**, 10234–10244.
 51. Pommer, A. J., Kuhlmann, U. C., Cooper, A., Hemmings, A. M., Moore, G. R., James, R. & Kleanthous, C. (1999). Homing in on the role of transition metals in the HNH motif of colicin endonucleases. *J. Biol. Chem.* **274**, 27153.
 52. van den Bremer, E. T., Jiskoot, W., James, R., Moore, G. R., Kleanthous, C., Heck, A. J. & Maier, C. S. (2002). Probing metal ion binding and conformational properties of the colicin E9 endonuclease by electrospray ionization time-of-flight mass spectrometry. *Protein Sci.* **11**, 1738.
 53. Doudeva, L. G., Huang, H., Hsia, K. C., Shi, Z., Li, C. L., Shen, Y. *et al.* (2006). Crystal structural analysis and metal-dependent stability and activity studies of the ColE7 endonuclease domain in complex with DNA/Zn²⁺ or inhibitor/Ni²⁺. *Protein Sci.* **15**, 269–280.
 54. Mate, M. J. & Kleanthous, C. (2004). Structure-based analysis of the metal-dependent mechanism of H–N–H endonucleases. *J. Biol. Chem.* **279**, 34763–34769.
 55. Schaller, K. & Nomura, M. (1976). Colicin E2 is DNA endonuclease. *Proc. Natl Acad. Sci. USA*, **73**, 3989–3993.
 56. Lawrence, M. C. & Colman, P. M. (1993). Shape complementarity at protein/protein interfaces. *J. Mol. Biol.* **234**, 946–950.
 57. Krissinel, E. & Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797.
 58. Lo Conte, L., Chothia, C. & Janin, J. (1999). The atomic structure of protein–protein recognition sites. *J. Mol. Biol.* **285**, 2177–2198.
 59. Rodier, F., Bahadur, R. P., Chakrabarti, P. & Janin, J. (2005). Hydration of protein–protein interfaces. *Proteins*, **60**, 36–45.
 60. Oshima, H., Yasuda, S., Yoshidome, T., Ikeguchi, M. & Kinoshita, M. (2011). Crucial importance of the water-entropy effect in predicting hot spots in protein–protein complexes. *Phys. Chem. Chem. Phys.* **13**, 16236–16246.
 61. Wallis, R., Reilly, A., Rowe, A., Moore, G. R., James, R. & Kleanthous, C. (1992). *In vivo* and *in vitro* characterization of overproduced colicin E9 immunity protein. *Eur. J. Biochem.* **207**, 687–695.
 62. Kabsch, W. (1988). Evaluation of single-crystal X-ray diffraction data from a position-sensitive detector. *J. Appl. Crystallogr.* **21**, 916–924.
 63. Evans, P. R. (1993). Data Collection and Processing. In *Proceedings of the CCP4 Study Weekend* (Sawyer, L., Isaacs, N. & Bailey, S., eds), pp. 114–122, Daresbury Laboratory, Warrington, UK.
 64. Collaborative Computational Project, No. 4. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **50**, 760–763.
 65. Vagin, A. & Teplyakov, A. (1997). MOLREP: an automated program for molecular replacement. *J. Appl. Crystallogr.* **30**, 1022–1025.
 66. Perrakis, A., Morris, R. & Lamzin, V. S. (1999). Automated protein model building combined with iterative structure refinement. *Nat. Struct. Biol.* **6**, 458–463.
 67. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **53**, 240–255.
 68. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **60**, 2126–2132.
 69. Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X. *et al.* (2007). MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**, W375–W383.