

# Experimental and Computational Analyses of the Energetic Basis for Dual Recognition of Immunity Proteins by Colicin Endonucleases

Anthony H. Keeble<sup>1</sup>, Lukasz A. Joachimiak<sup>2</sup>, María Jesus Maté<sup>1</sup>, Nicola Meenan<sup>1</sup>, Nadine Kirkpatrick<sup>1</sup>, David Baker<sup>2</sup> and Colin Kleanthous<sup>1\*</sup>

<sup>1</sup>Department of Biology,  
University of York, Heslington,  
York YO10 5YW, UK

<sup>2</sup>Department of Biochemistry,  
Howard Hughes Medical  
Institute, University of  
Washington, HSB J559 Seattle,  
WA 98195, USA

Received 13 February 2008;  
received in revised form  
19 March 2008;  
accepted 25 March 2008  
Available online  
3 April 2008

Colicin endonucleases (DNases) are bound and inactivated by immunity (Im) proteins. Im proteins are broadly cross-reactive yet specific inhibitors binding cognate and non-cognate DNases with  $K_d$  values that vary between  $10^{-4}$  and  $10^{-14}$  M, characteristics that are explained by a 'dual-recognition' mechanism. In this work, we addressed for the first time the energetics of Im protein recognition by colicin DNases through a combination of E9 DNase alanine scanning and double-mutant cycles (DMCs) coupled with kinetic and calorimetric analyses of cognate Im9 and non-cognate Im2 binding, as well as computational analysis of alanine scanning and DMC data. We show that differential  $\Delta\Delta G$ s observed for four E9 DNase residues cumulatively distinguish cognate Im9 association from non-cognate Im2 association. E9 DNase Phe86 is the primary specificity hotspot residue in the centre of the interface, which is coordinated by conserved and variable hotspot residues of the cognate Im protein. Experimental DMC analysis reveals that only modest coupling energies to Im9 residues are observed, in agreement with calculated DMCs using the program ROSETTA and consistent with the largely hydrophobic nature of E9 DNase–Im9 specificity contacts. Computed values for the 12 E9 DNase alanine mutants showed reasonable agreement with experimental  $\Delta\Delta G$  data, particularly for interactions not mediated by interfacial water molecules.  $\Delta\Delta G$  predictions for residues that contact buried water molecules calculated using solvated rotamer models met with mixed success; however, we were able to predict with a high degree of accuracy the location and energetic contribution of one such contact. Our study highlights how colicin DNases are able to utilise both conserved and variable amino acids to distinguish cognate from non-cognate Im proteins, with the energetic contributions of the conserved residues modulated by neighbouring specificity sites.

© 2008 Elsevier Ltd. All rights reserved.

Edited by B. Honig

**Keywords:** protein–protein interactions; specificity; alanine scanning; double-mutant cycles; crystallography

\*Corresponding author. E-mail address: ck11@york.ac.uk.

Present addresses: A. H. Keeble, MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 0QH, UK; M.-J. Maté, Centro de Investigaciones Biológicas, Consejo Superior de Investigaciones Científicas, C/Ramiro de Maeztu 9, Madrid 28040, Spain.

Abbreviations used: E9 DNase, endonuclease domain of colicin E9; Im2, immunity protein for colicin E2 DNase; Im9, immunity protein for colicin E9 DNase; IPE, immunity protein exosite; ITC, isothermal titration calorimetry; DMC, double-mutant cycle;  $\Delta\Delta C_{kin}^{obs}$ , observed change in the free energy of binding determined kinetically;  $k_{on}$ , association rate constant;  $k_{off}$ , dissociation rate constant;  $\Delta\Delta G^{calc}$ , calculated change in the free energy of binding;  $\Delta\Delta G_{H_2O}^{calc}$ , calculated change in the free energy of binding incorporating a solvated rotamer model;  $\Delta\Delta G_{ITC}^{obs}$ , observed change in the free energy of binding determined by ITC;  $\Delta\Delta G_{int}^{obs}$ , observed interaction energy obtained from DMC analysis;  $\Delta\Delta G_{int}^{calc}$ , calculated interaction energy obtained from DMC analysis.

## Introduction

Most biological processes involve the recognition of one protein by another. High-throughput screening of protein–protein interactions in prokaryotic and eukaryotic proteomes has shown cells to be a conglomeration of interconnected complexes, the principal components of which are ‘protein hubs,’ nodes with a larger-than-average number of interaction partners.<sup>1</sup> Hubs tend to come in two flavours, with recent studies indicating that each has distinct evolutionary trajectories; hubs that simultaneously accommodate multiple partners at distinct binding sites appear to evolve slowly, while those that distinguish multiple binding partners at a single interaction site in a specific mutually exclusive fashion evolve more quickly.<sup>2</sup> This latter group of protein–protein interaction complexes can be further subdivided as those that bind unrelated partners at the binding site, such as the Fc hinge region of human immunoglobulin G,<sup>3</sup> and those that distinguish members of a highly related protein family, which is typical of protein inhibitors of enzymes and intra- and intercellular signalling complexes.<sup>4,5</sup> The present work addressed the fundamental problem of specificity in this latter type of complex, where, using protein engineering and X-ray crystallography in combination with kinetic, thermodynamic and computational analyses, we dissected the specific recognition of the colicin E9 endonuclease (E9 DNase) by cognate and non-cognate immunity (Im) proteins. Colicin DNase–Im protein complexes are a powerful model system with which to study discrimination at a single binding site since specificity in this family of closely related complexes spans the thermodynamic stability of all known protein–protein interactions.<sup>6</sup>

The E colicin DNases E2, E7, E8 and E9 are 15-kDa C-terminal domains of larger bacterial toxins produced by strains of *Escherichia coli* under times of environmental stress that kill competing cells through random degradation of genomic DNA.<sup>7,8</sup> Cell suicide is prevented by the co-production of small Im proteins (Im2, Im7, Im8 and Im9) that form 1:1 complexes with the corresponding DNase domains, thereby neutralising their enzymatic activity in the host.<sup>6,9</sup> One of the distinguishing features of these complexes is that the inhibitors bind at an Im protein exosite (IPE) on the DNase that, in contrast to the majority of other protein inhibitors of enzymes, results in inhibition through steric and electrostatic occlusion of nucleic acid substrates.<sup>10,11</sup> Colicin DNases share a high degree of sequence identity (~65%) and structural conservation and so cleave DNA by the same basic mechanism. Importantly, however, much of the sequence variation is confined to residues 72–98 in these 134 amino acid domains, which coincides with the location of the IPE (Fig. 1a) and explains how Im proteins have evolved to recognise specific colicin DNases.<sup>6</sup>

Previous work from our laboratory had shown that Im proteins, which share ~50% sequence

identity, are specific, high-affinity inhibitors of colicin DNases ( $K_d = 10^{-14}$ – $10^{-15}$  M), with cognate complexes binding  $10^6$ - to  $10^{10}$ -fold tighter than non-cognate complexes.<sup>14–16</sup> Non-cognate Im protein binding, typically in the micromolar to nanomolar range, has been attributed to a conserved hotspot on the Im protein that is accommodated at the sequence-variable DNase IPE largely through backbone hydrogen bonds and van der Waals interactions.<sup>13</sup> Im proteins are helical proteins that bind the DNase through a ‘dual-recognition’ mechanism; a long specificity helix (helix II) preferentially stabilises cognate binding, while a shorter conserved helix (helix III) forms a hotspot that stabilises cognate and non-cognate complexes alike and is responsible for the basal DNase cross-reactivity of Im proteins.<sup>16,17,18</sup> Although neither protein undergoes substantial structural change on binding, their association is a special case of a conformational change-type mechanism where cognate and non-cognate complexes alike form an electrostatically steered encounter complex centred on the conserved residues of helix III of the Im protein, followed by rigid-body rotation, allowing optimal alignment of specificity residues across the interface.<sup>19</sup> Such a mechanism has strong similarities to recent theories of anchor residue binding in protein–protein interactions, as described by Rajamani *et al.*,<sup>20</sup> where conserved polar or aromatic anchor residues, as are found in helix III of Im9, bind to pre-organised pockets on the binding partner to form the encounter complex and neighbouring latch (specificity) residues contribute to forming the final high-affinity complex.

While the kinetics of binding are conserved between the different DNase–Im protein complexes,<sup>16,19</sup> isothermal titration calorimetry (ITC) has shown that the underlying thermodynamics of binding vary significantly.<sup>21</sup> Specificity is enthalpically driven and coupled to a reduction in side-chain or backbone flexibility in the cognate complexes of colicins E2, E7 and E8. In contrast, changes in flexibility appear to play a lesser role in Im9 binding to the colicin E9 DNase, where instead specificity is entropically driven and coupled to the loss of hydrophobic hydration.<sup>21</sup>

In addition to studies aimed at investigating the kinetics and thermodynamics of colicin DNase–Im protein complexes, the system has also served as a model in a number of computational and evolutionary studies aimed at understanding specificity in protein–protein recognition. Goh and Cohen showed how evolutionary trace methods could correctly assign Im protein specificity to colicin DNases.<sup>22</sup> Bernath *et al.*<sup>23</sup> devised a novel *in vitro* compartmentalisation strategy for investigating the directed evolution of Im protein specificity of colicin DNases and showed that this correlates with those isolated naturally. Baker and co-workers redesigned colicin–Im protein specificity through *de novo* computational methods, showed that the structure of the redesigned complex matched closely that predicted and demonstrated



binding, we conducted an alanine scan across the surface of the enzyme, guided by the 1.7-Å crystal structure of the E9 DNase–Im9 complex,<sup>13</sup> and determined changes in the free energy of binding using kinetic methods ( $\Delta\Delta G_{\text{kin}}^{\text{obs}}$ ). The E9 DNase IPE is composed predominantly of a single contiguous epitope (residues 72–98) but includes Arg54 at the periphery of the Im protein binding site. Arg54 contacts substrate DNA and is required for catalysis; its burial by Im9 largely accounts for the inhibition of DNA binding by the acidic Im protein.<sup>26</sup> Thirteen residues were selected (Materials and Methods) and mutated singly to alanine (Fig. 1b). All were purified with the exception of E9 DNase Y83A, which was not expressed to any detectable level in bacteria and so could not be purified. The 12 mutants were folded and adjudged not to be significantly destabilised by the mutations as deduced by intrinsic tryptophan fluorescence, far-UV CD spectroscopy and thermal denaturation experiments, all of which were similar to the wild-type protein (Materials and Methods). Association ( $k_{\text{on}}$ ) and dissociation ( $k_{\text{off}}$ ) rate constants for both cognate (Im9) and non-cognate (Im2) E9 DNase–Im protein complexes were determined as described previously,<sup>14,15,19</sup> from which the equilibrium dissociation constant ( $K_{\text{d}}$ ) for each complex was obtained using Eq. (1).

$$K_{\text{d}} = k_{\text{off}}/k_{\text{on}} \quad (1)$$

The  $k_{\text{on}}$  data for the 12 E9 DNase alanine mutants binding Im9 and those binding Im2 are shown in Tables 1 and 2, respectively. The alanine mutations generally decreased values of  $k_{\text{on}}$  by less than 3-fold, with the largest decrease accompanying removal of

charged groups. Notable exceptions are the acceleration of the  $k_{\text{on}}$  values for several mutants (N72A, N75A and F86A) binding Im2. Given that this rate enhancement occurs upon truncation to alanine of both polar (e.g., Asn72) and non-polar (Phe86) residues, the origin is likely to be due to geometric rather than electrostatic factors, consistent with the studies by Janin<sup>29</sup> and Schreiber.<sup>30</sup> Hence, in these instances, the mutations likely remove a restriction to alignment of the E9 DNase and Im2 during association. In contrast to the  $k_{\text{on}}$  data, the  $k_{\text{off}}$  data for the 12 E9 DNase alanine mutants in complex with cognate Im9 and non-cognate Im2 were affected by up to 1000-fold (Tables 1 and 2, respectively), indicating that  $k_{\text{off}}$  provides the kinetic discrimination between cognate and non-cognate binding, consistent with previous mutagenesis studies on Im proteins.<sup>14–16,19</sup>

### Experimental and computational determination of $\Delta\Delta G$ for Im9 binding E9 DNase IPE alanine mutants

Kinetically derived equilibrium dissociation constants were used to obtain values of  $\Delta\Delta G$  ( $\Delta\Delta G_{\text{kin}}^{\text{obs}}$ ) for all E9 DNase alanine mutants binding Im9 (Table 1), according to Eq. (2), and mapped onto the E9 DNase–Im9 complex in Fig. 2.

$$\Delta\Delta G = RT \ln \left( K_{\text{d}}^{\text{mutant}} / K_{\text{d}}^{\text{wild type}} \right) \quad (2)$$

The data highlight a series of important characteristics about the energetics of the DNase in its recognition of the cognate Im9. First, values for

**Table 1.** Kinetic and thermodynamic parameters for E9 DNase alanine mutants binding the cognate Im9

Im9 complex	$k_{\text{on}} \times 10^7$ ( $\text{M}^{-1} \text{s}^{-1}$ ) <sup>a</sup>	$k_{\text{off}}$ ( $\text{s}^{-1}$ ) <sup>b</sup>	$K_{\text{d,kinetic}}$ (M) <sup>c</sup>	$\Delta\Delta G_{\text{kin}}^{\text{obs}}$ ( $\text{kcal mol}^{-1}$ ) <sup>d</sup>	$\Delta\Delta G^{\text{calc}}$ ( $\text{kcal mol}^{-1}$ ) <sup>e</sup>	$\Delta\Delta G_{\text{H}_2\text{O}}^{\text{calc}}$ ( $\text{kcal mol}^{-1}$ ) <sup>f</sup>
Wild type	7.24 ( $\pm 0.21$ )	$2.36 (\pm 0.14) \times 10^{-6}$	$3.26 (\pm 0.28) \times 10^{-14}$			
R54A	6.52 ( $\pm 0.34$ )	$3.54 (\pm 0.27) \times 10^{-5}$	$5.43 (\pm 0.66) \times 10^{-13}$	1.7	1.8	1.7
N72A	5.50 ( $\pm 0.14$ )	$1.28 (\pm 0.04) \times 10^{-5}$	$2.33 (\pm 0.13) \times 10^{-13}$	1.1	1.7	−0.7
S74A	8.05 ( $\pm 0.1$ )	$1.75 (\pm 0.21) \times 10^{-6}$	$2.17 (\pm 0.28) \times 10^{-14}$	−0.2	−0.2	−3.8
N75A	10 ( $\pm 0.6$ )	$1.68 (\pm 0.22) \times 10^{-4}$	$1.68 (\pm 0.30) \times 10^{-12}$	2.3	0.6	1.3
S77A	7.18 ( $\pm 0.37$ )	$1.58 (\pm 0) \times 10^{-6}$	$2.20 (\pm 0.11) \times 10^{-14}$	−0.2	0.0	0.6
S78A	7.76 ( $\pm 0.23$ )	$1.02 (\pm 0.01) \times 10^{-6}$	$1.31 (\pm 0.05) \times 10^{-14}$	−0.5	−0.1	−0.1
S84A	6.91 ( $\pm 0.04$ )	$1.87 (\pm 0) \times 10^{-6}$	$2.71 (\pm 0.01) \times 10^{-14}$	−0.1	0.1	0.1
F86A	7.91 ( $\pm 0.49$ )	$1.80 (\pm 0.2) \times 10^{-3}$	$2.28 (\pm 0.37) \times 10^{-11}$	3.9	2.9	4.0
T87A	5.73 ( $\pm 0.39$ )	$2.44 (\pm 0.05) \times 10^{-6}$	$4.26 (\pm 0.35) \times 10^{-14}$	0.2	0.0	−0.1
Q92A	7.78 ( $\pm 0.28$ )	$1.59 (\pm 0.15) \times 10^{-6}$	$2.04 (\pm 0.26) \times 10^{-14}$	−0.3	0.9	1.6
K97A	3.42 ( $\pm 0.4$ )	$3.05 (\pm 0.1) \times 10^{-5}$	$8.92 (\pm 1.1) \times 10^{-13}$	2.0	0.3	−0.8
V98A	7.09 ( $\pm 0$ )	$1.45 (\pm 0.06) \times 10^{-5}$	$2.05 (\pm 0.09) \times 10^{-13}$	1.1	0.2	0.3

All experiments were carried out in 50 mM Mops buffer, pH 7.0, containing 200 mM NaCl and 1 mM DTT at 25 °C and represent the average of duplicate experiments, usually carried out on the same preparation of protein. Errors (shown in parentheses) were generally in the range of 5%–30% for duplicate measurements.

<sup>a</sup> E9 DNase–Im9  $k_{\text{on}}$  values were determined by stopped-flow fluorescence under pseudo-first-order conditions.

<sup>b</sup> E9 DNase–Im9  $k_{\text{off}}$  values were determined by radioactive subunit exchange kinetics as described by Wallis *et al.*<sup>15</sup>

<sup>c</sup> Kinetically derived equilibrium dissociation constant ( $K_{\text{d,kinetic}}$ ) values calculated according to Eq. (1). The value determined in this study for wild-type Im9 is very close to that published previously ( $2.4 \times 10^{-14}$  M).<sup>15</sup>

<sup>d</sup> Observed free energy change as a result of alanine substitution relative to the wild-type E9 DNase–Im9 complex from kinetically determined equilibrium dissociation constants according to Eq. (2).<sup>17</sup>

<sup>e</sup> Calculated free energy change as a result of alanine substitution relative to the wild-type E9 DNase–Im9 for each of the 12 purified E9 DNase alanine mutants using the program ROSETTA.<sup>27</sup>

<sup>f</sup> Calculated free energy change as above but incorporating solvated rotamers.<sup>28</sup>

**Table 2.** Kinetic and thermodynamic parameters for E9 DNase alanine mutants binding the non-cognate Im2

Im2 complex	$k_{\text{on}} \times 10^7$ ( $\text{M}^{-1} \text{s}^{-1}$ ) <sup>a</sup>	$k_{\text{off}}$ ( $\text{s}^{-1}$ ) <sup>b</sup>	$K_{\text{d,kinetic}}$ ( $\text{M}$ ) <sup>c</sup>	$K_{\text{d,ITC}}$ ( $\text{M}$ ) <sup>d</sup>	$\Delta\Delta G_{\text{kin}}^{\text{obs}}$ ( $\text{kcal mol}^{-1}$ ) <sup>e</sup>	$\Delta\Delta G_{\text{ITC}}^{\text{obs}}$ ( $\text{kcal mol}^{-1}$ ) <sup>f</sup>
Wild type	6.90 ( $\pm 0.21$ )	0.90 ( $\pm 0.04$ )	$1.30 (\pm 0.09) \times 10^{-8}$	$1.50 (\pm 0.04) \times 10^{-7}$		
R54A	7.85 ( $\pm 0.34$ )	6.92 ( $\pm 0.24$ )	$8.82 (\pm 0.65) \times 10^{-8}$	$4.20 (\pm 0.3) \times 10^{-7}$	1.1	0.6
N72A	17.7 ( $\pm 0.14$ )	10.83 ( $\pm 1.25$ )	$6.12 (\pm 0.75) \times 10^{-8}$	$3.40 (\pm 0.15) \times 10^{-7}$	0.9	0.5
S74A	8.71 ( $\pm 0.10$ )	0.27 ( $\pm 0.02$ )	$3.10 (\pm 0.03) \times 10^{-9}$	$4.0 (\pm 0.7) \times 10^{-7}$	-0.9	0.6
N75A	35.8 ( $\pm 0.60$ )	37.59 ( $\pm 1.22$ )	$1.05 (\pm 0.05) \times 10^{-7}$	$1.27 (\pm 0.1) \times 10^{-6}$	1.2	1.3
S77A	7.13 ( $\pm 0.37$ )	0.33 ( $\pm 0.02$ )	$4.63 (\pm 0.49) \times 10^{-9}$	$9.0 (\pm 0.1) \times 10^{-8}$	-0.6	-0.3
S78A	6.41 ( $\pm 0.23$ )	0.65 ( $\pm 0.03$ )	$1.01 (\pm 0.08) \times 10^{-8}$	$1.4 (\pm 0.15) \times 10^{-7}$	-0.2	0.0
S84A	7.18 ( $\pm 0.04$ )	0.80 ( $\pm 0.04$ )	$1.11 (\pm 0.06) \times 10^{-8}$	$1.4 (\pm 0.1) \times 10^{-7}$	-0.1	0.0
F86A	58.6 ( $\pm 0.49$ )	54.93 ( $\pm 2.62$ )	$9.37 (\pm 0.52) \times 10^{-8}$	$7.4 (\pm 0.2) \times 10^{-7}$	1.2	0.9
T87A	3.26 ( $\pm 0.39$ )	1.03 ( $\pm 0.01$ )	$3.16 (\pm 0.36) \times 10^{-8}$	$2.2 (\pm 0.13) \times 10^{-7}$	0.5	0.2
Q92A	6.90 ( $\pm 0.28$ )	0.41 ( $\pm 0.01$ )	$5.94 (\pm 0.37) \times 10^{-8}$	$1.2 (\pm 0.1) \times 10^{-7}$	-0.5	-0.1
K97A	2.26 ( $\pm 0.23$ )	0.68 ( $\pm 0.01$ )	$3.01 (\pm 0.29) \times 10^{-8}$	$5.8 (\pm 0.7) \times 10^{-7}$	0.5	0.8
V98A	6.60 ( $\pm 0$ )	1.05 ( $\pm 0.09$ )	$1.59 (\pm 0.14) \times 10^{-8}$	$3 (\pm 0.25) \times 10^{-7}$	0.1	0.4

Experimental conditions and error analysis are as those for Im9 binding, described in Table 1.

<sup>a</sup> E9 DNase-Im2  $k_{\text{on}}$  values were determined by stopped-flow fluorescence under second-order conditions.

<sup>b</sup> E9 DNase-Im2  $k_{\text{off}}$  values were determined by stopped-flow fluorescence competition experiments (Materials and Methods).

<sup>c</sup> Kinetically derived equilibrium dissociation constants (Eq. (1)).

<sup>d</sup> Equilibrium dissociation constants determined by ITC (Materials and Methods).

<sup>e</sup> Changes in the free energy of binding Im2 by E9 DNase alanine mutants where  $K_{\text{d}}$  was determined kinetically.

<sup>f</sup> Changes in the free energy of binding Im2 by E9 DNase alanine mutants where  $K_{\text{d}}$  was determined by ITC.

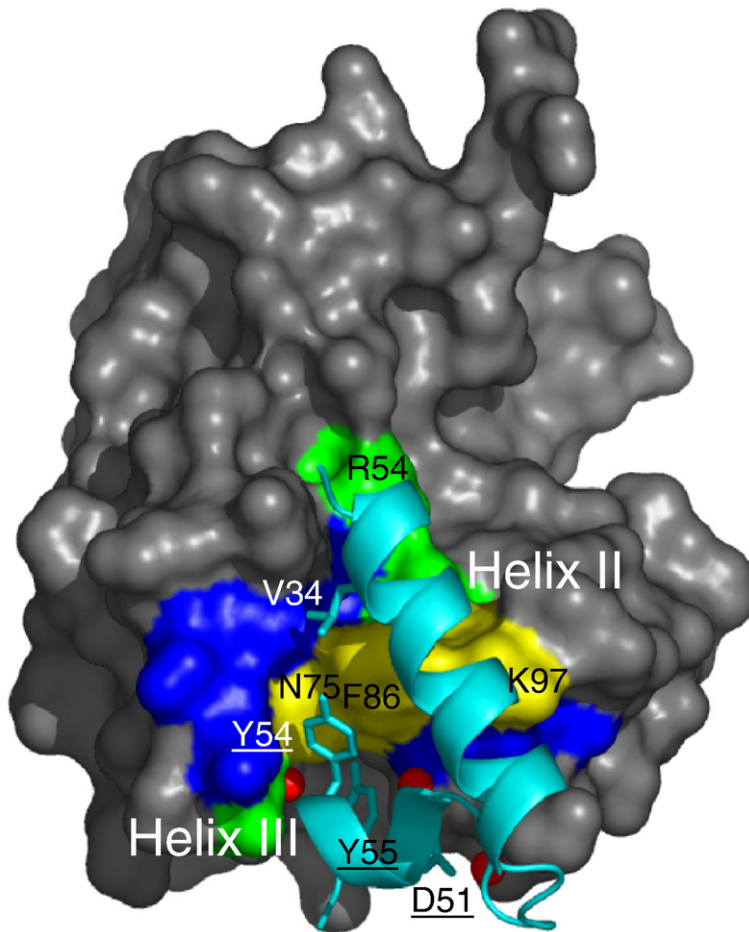
$\Delta\Delta G_{\text{kin}}^{\text{obs}}$  fall into three categories: those that affect binding by (1)  $<1$  kcal/mol, (2) 1–2 kcal/mol and (3) 2–4 kcal/mol (coloured blue, green and yellow, respectively, in Fig. 2), with half of the mutants analysed (S74A, S77A, S78A, S84A, T87A and Q92A) being in the first category and so having little or no effect on binding. These include the three residues described above that, although buried by Im9, make no direct contact with the protein. Second, the three most important residues (Asn75, Phe86 and Lys97) form a central belt on the surface of the enzyme that comprises the hotspot, consistent with other protein–protein interactions where the binding energy is not distributed evenly across the interface but localised.<sup>31–34</sup> Third, previous alanine scans of IPE contacting residues in Im9 showed that Im protein mutations, particularly of conserved hotspot residues from helix III, have a much greater impact on binding than do mutations of Im protein-contacting residues in the E9 DNase IPE reported here. Mutation of the three helix III hotspot residues in Im9 (Asp51, Tyr54 and Tyr55) generates  $\Delta\Delta G_{\text{kin}}^{\text{obs}} >5$  kcal/mol compared with  $<3$  kcal/mol for most DNase mutants, with the exception of E9 DNase F86A (3.9 kcal/mol). Im9 Asp51 is accommodated at the sequence-variable IPE of the DNase through direct and water-mediated hydrogen bonds to the backbone, whereas DNase IPE residues mostly engage in side-chain interactions.<sup>13</sup> Notwithstanding these differences, it is clear that the ‘hotter spot’ of Im9, in particular, Tyr54 and Tyr55, meets that of the E9 DNase IPE centred on Phe86 (Fig. 2).

Several authors have stressed the importance of validating alanine scanning data by computational approaches.<sup>35</sup> In the present work, we computed the effects of these alanine substitutions on the free energy of binding ( $\Delta\Delta G^{\text{calc}}$ ) for E9 DNase alanine mutants binding Im9 for direct comparison with  $\Delta\Delta G_{\text{kin}}^{\text{obs}}$  not only to further probe the physical basis

for binding but also as a test of the robustness of computational approaches using this model system. Computational strategies employing a simple physical model have successfully recapitulated experimental alanine scanning data for many protein interfaces<sup>36</sup> and have paved the way for developing tools aimed at the rational design of protein–protein complexes.<sup>24,25,37–39</sup> Here we applied the program ROSETTA<sup>27</sup> to compute binding free energies for a set of alanine-substituted models of the colicin DNase E9–Im9 complex. Agreement between the experimentally determined  $\Delta\Delta G_{\text{kin}}^{\text{obs}}$  and predicted  $\Delta\Delta G^{\text{calc}}$  is reasonable over the 12 residues on the E9 DNase side of the interface (Fig. 3; Table 1), with a linear fit yielding a correlation coefficient of 0.75.

These initial calculations did not take account of buried, interfacial water molecules. Indeed, prediction performance is dependent on whether a given residue is involved in a water-mediated contact; if the four residues that make water-mediated contacts are excluded from the mutation set (Asn72, Asn75, Gln92 and Lys97), the linear fit for the remaining eight mutations yields a correlation coefficient of 0.94. In the field of design, we have had success optimising the packing arrangement of non-polar residues comprising the core to thermostabilise a protein<sup>40</sup> and in the design of protein–protein interaction specificity by engineering a hydrogen bond network.<sup>24,25</sup> However, protein interfaces tend to also be solvated; to this end, it is important to develop tools able to capture the energetic contribution of water-mediated contacts. In all four cases of residues that engage in water-mediated hydrogen bond networks at the interface, the computed  $\Delta\Delta G^{\text{calc}}$  values are incorrectly assigned (Fig. 3a, data points coloured red).

To try to capture the loss of the water-mediated energetics, we used a solvated rotamer approach<sup>28</sup> to hydrate the interface and compute  $\Delta\Delta G_{\text{H}_2\text{O}}^{\text{calc}}$



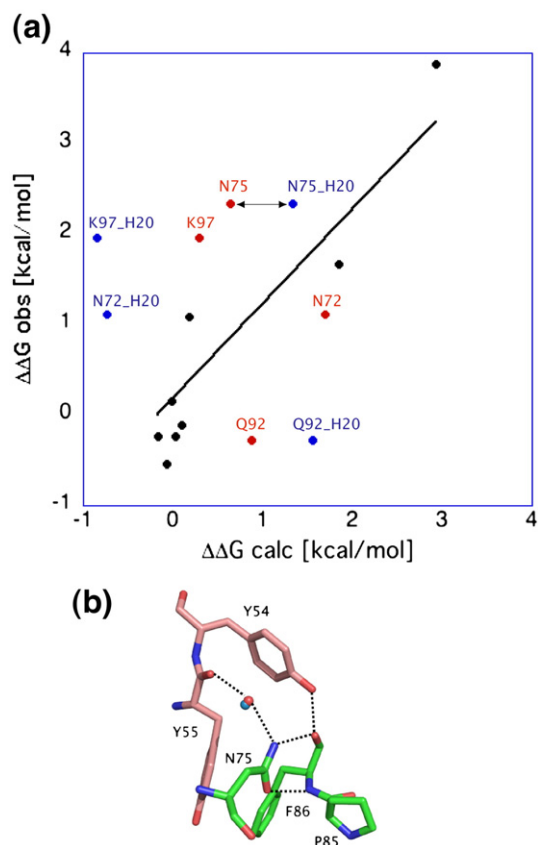
**Fig. 2.** The conserved hotspot of Im9 meets the variable hotspot of the E9 DNase. Molecular surface representation of the E9 DNase, coloured according to values of  $\Delta\Delta G$  for alanine mutants, in the context of the complex with Im9 (Table 1). Only two of the four helices of Im9 are shown for simplicity. Colour code for values of  $\Delta\Delta G_{\text{kin}}^{\text{obs}}$ : yellow, 2–4 kcal/mol; green, 1–2 kcal/mol; blue, <1 kcal/mol. With the exception of Arg54 and Asn75, the surface of the IPE is largely variable in the DNase family. By contrast, the three Im9 residues of helix III (shown as ribbon), Asp51, Tyr54 and Tyr55, are all conserved (underlined in the figure). Wallis *et al.*<sup>17</sup> showed previously that mutation of these residues to alanine generates much greater values of  $\Delta\Delta G$  (>5 kcal/mol). The docking of helix III into the concave cleft of the DNase IPE positions the specificity helix of the Im protein for the appropriate recognition of residue 86 on the DNase and to make additional specificity contacts, such as with Lys97. The correct positioning of helix II is thought to be through rotation about the conserved hotspot of the Im protein in the encounter complex, possibly mediated by three conserved water molecules, shown as red spheres in the figure.

(change in the free energy of binding incorporating a solvated rotamer model). The agreement between the experimentally determined  $\Delta\Delta G_{\text{kin}}^{\text{obs}}$  and predicted  $\Delta\Delta G_{\text{H}_2\text{O}}^{\text{calc}}$  is worse than  $\Delta\Delta G^{\text{calc}}$  over the same 12 residues (Table 1), with a linear fit yielding a correlation coefficient of 0.58 (data not shown). However, we were specifically interested whether our solvated rotamer approach improves our predictions for residues known to be involved in water-mediated interactions and, in particular, Asn72, Asn75, Gln92 and Lys97. The comparison results of  $\Delta\Delta G_{\text{H}_2\text{O}}^{\text{calc}}$  and  $\Delta\Delta G^{\text{obs}}$  values for these four residues are shown in Fig. 3a (data points coloured blue). In two cases (Asn72 and Lys97), the introduction of the localised water worsens the correlation by grossly underestimating the influence of the alanine mutation, while in the case of Gln92, it causes marginal overestimation. In the case of Asn75, the positioning of the water molecule is accurately predicted and the computed binding energy using a solvated model gives a  $\Delta\Delta G_{\text{H}_2\text{O}}^{\text{calc}}$  value of 1.33 kcal/mol (Fig. 3a, N75A\_H<sub>2</sub>O, blue), hence capturing some of the water-mediated energetics (Table 1). This asparagine extends a hydrogen bond network by donating to the backbone carbonyl of Phe86 and a tightly bound water molecule (W2)

while accepting from a backbone amide of Phe86 (Fig. 3b). We conclude that while the solvated rotamer approach can be useful in determining placement of functionally important water molecules that are tightly bound and buried (<2 hydrogen bonds and low *B*-factors) and their energetic contribution to binding can be quantitated, there is still a degree of variability in the success of this approach. Likely reasons for the failure of the solvated approach can be attributed to errors in the energy model and are compounded by uncertainties in the positions of the water molecules. Improved prediction of water placement at protein interfaces are likely to be achieved by using larger solvated rotamer libraries and changes in the solvation and water–van der Waals components of the energy function.

#### Differential effects of E9 DNase IPE alanine mutants on Im9 and Im2 binding reveal contributions from both conserved and variable residues to specificity

The  $K_d$  for Im2 binding the E9 DNase is  $10^{-7}$  M, at pH 7 in buffer containing 200 mM NaCl and at 25 °C, 7 orders of magnitude weaker than cognate

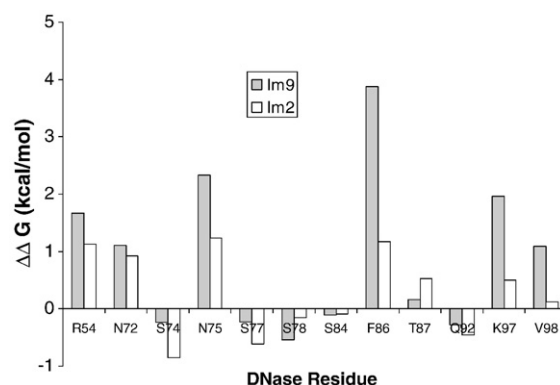


**Fig. 3.** (a) Calculated *versus* observed changes in  $\Delta\Delta G$  for alanine mutations of colicin E9 DNase IPE residues binding the cognate partner Im9. Line reflects a fit with a fixed zero intercept. Residues marked in red are reported to form water-mediated hydrogen bonds in the interface. Residues marked in blue (e.g., N75\_H2O) were computed using a solvated rotamer model. A linear fit to all 12 residues yields a correlation coefficient of 0.75; a linear fit to residues not involved in water-mediated contacts yields a correlation coefficient of 0.94. See the text for details. (b) Rotamer water prediction coordinating Asn75 in the native E9 DNase-Im9 Im protein complex. Figure shows a comparison of the native (red) and predicted (blue) water molecules that mediate the interfacial hydrogen bond between Asn75 of the E9 DNase and the backbone of Im9.

Im9 binding.<sup>19</sup> To determine whether our set of E9 DNase alanine mutants could provide information on how the enzyme discriminates between these close homologues, we obtained kinetically determined equilibrium dissociation constants and hence  $\Delta\Delta G_{\text{kin}}^{\text{obs}}$  for all E9 DNase alanine mutants binding Im2 (Table 2), determined according to Eq. (2). We showed recently that complications in the analysis of biphasic dissociation kinetics of E9 DNase-Im2 complexes have led to erroneous  $K_d$  determinations due to misassignment of the rate-limiting step for dissociation of the complex.<sup>19</sup> We therefore checked all kinetically derived values of  $K_d$  ( $K_{d,\text{kinetic}}$ ) by ITC ( $K_{d,\text{ITC}}$ ), from which  $\Delta\Delta G_{\text{kin}}^{\text{obs}}$  and  $\Delta\Delta G_{\text{ITC}}^{\text{obs}}$  (observed change in the free energy of binding determined by ITC) were obtained (Table 2). This analysis shows

that while kinetically and calorimetrically determined  $K_d$  values for Im2 mutants binding the E9 DNase show some variability (2- to 20-fold), there is nonetheless reasonable agreement between values of  $\Delta\Delta G^{\text{obs}}$  from the two approaches (<0.5-kcal/mol discrepancy for 11/12 mutations). Since there is no high-resolution structure for the E9 DNase-Im2 complex or indeed any non-cognate complex,  $\Delta\Delta G^{\text{calc}}$  was not computed.

Figure 4 compares values of  $\Delta\Delta G_{\text{kin}}^{\text{obs}}$  for Im9 and Im2 binding the 12 E9 DNase alanine mutants and illustrates a number of key points concerning the specific recognition of Im proteins. First, none of the 12 mutations affected Im2 binding by significantly more than 1 kcal/mol (Table 2), with three-quarters having a negligible effect. Second, two mutation sites on the E9 DNase (Arg54 and Asn72) yield similar changes in Im binding for Im2 and Im9, both of which are on the periphery of the IPE. Third, comparisons with Im9 data show clear evidence of differential binding effects at the four positions within the IPE (75, 86, 97 and 98; Fig. 4) and therefore identify these as specificity sites, three of which (Asn75, Phe86 and Lys97) form the central hotspot belt for the cognate complex. The identification of Lys97 and Val98 as specificity residues is consistent with their contacting helix II residues, with the specificity helix in Im9 (Glu41 and Val34, respectively) and with previous alanine scanning data of Im2 and Im9 showing that these positions specify E9 and not E2 DNase binding.<sup>16</sup> That these sites are of more general importance to colicin specificity comes from the data of Curtis and James,<sup>12</sup> who found that swapping residues 97 and 98 in colicin E9 DNase for those in colicin E8 resulted in stronger protection by Im8 *in vivo*. Fourth, more surprising are the context-dependent  $\Delta\Delta G_{\text{kin}}^{\text{obs}}$  effects seen for conserved DNase residues on Im protein specificity. Alanine mutation of Asn75, which is conserved in all colicin DNases and, as described above, only contacts the Im protein indirectly



**Fig. 4.** Differential effects of E9 DNase alanine mutants on  $\Delta\Delta G^{\text{obs}}$  for cognate Im9 and non-cognate Im2 binding. Histogram showing the change in the free energy of binding relative to wild-type E9 DNase of 12 alanine mutants binding Im9 (gray bars; see Table 1 for details) and Im2 (white bars; see Table 2 for details). In each case, data are the average of two independent experiments.

through a conserved water molecule (Fig. 3b), has a differential effect on Im9 *versus* Im2 binding. E9 DNase F86A shows the largest differential effect on Im2 and Im9 binding, yet Phe86 is conserved in the E2 and E9 DNases (Fig. 1a). Fifth, Fig. 4 suggests no role for Ser77 and Ser78 in defining the specificity of Im9 relative to Im2 for the E9 DNase, consistent with a lack of stabilising interactions with Im9 residues at these sites. These positions are however involved in mediating specificity in both the E7 and E8 complexes as revealed by crystallography and mutagenesis, respectively.<sup>10,12</sup> This implies that different combinations of residues are used by the different colicins to achieve specific recognition of their cognate Im proteins, which mirrors how Im proteins recognise the different colicin DNases.<sup>16</sup> In summary, specific recognition of Im9 relative to Im2 by the E9 DNase is largely due to a group of four residue contact points, with position 86 taking the predominant role. The expectation is that a similar distributive specificity mechanism will pertain to other cognate colicin DNase–Im protein complexes all having position 86 as their centrepoint (as seen in the colicin E7 DNase–Im7 complex),<sup>10</sup> albeit elaborated with different combinations of other specificity contacts.

### Structures of mutant E9 DNase–Im9 complexes reveal only localised structural changes

We determined the crystal structures of the four most destabilising alanine mutations in the E9 DNase (R54A, N75A, F86A and K97A), which largely define its specific recognition of Im9 relative to Im2, in complex with Im9 at or below 2-Å resolution (Fig. 5; Table 3). These structures were used to ascertain the extent to which the complex had been disrupted. RMSDs for a structural superposition of the four mutants with the wild-type complex based on the E9 DNase all had values <0.7 Å; structural changes were generally localised to flexible regions in the endonuclease, which have been described previously by Kuhlmann *et al.*,<sup>13</sup> and to the immediate vicinity of the mutation. Hence, truncation of these side chains to alanine did not propagate large-scale structural changes through the interface. The amount of solvent-accessible surface area buried at the mutant interfaces was similar to the wild-type complex (1500–1600 Å<sup>2</sup>). The other main observation from the four structures is that numerous water molecules flood into the cavities created by the alanine mutations in order to fill gaps at the interface, replace hydrogen bonding interactions of deleted heteroatoms or shield unpaired electrostatic interactions. None of these water molecules corresponds to water molecules previously observed in colicin DNase–Im protein complexes.

#### E9 DNase F86A–Im9 complex

The mutation of the key specificity determinant Phe86 for alanine generates a cavity in the interface

that is filled by two water molecules, W35 and W153, with both water molecules completely buried at the interface (accessibility values calculated with AREAIMOL of 0.0 for W35 and 2.0 for W153). As well as being hydrogen bonded to each other, W36 is also hydrogen bonded to W59 and the hydroxyl of Im9 Tyr54. The stacking interaction made by Tyr54 to the phenyl ring of Phe86 in the native complex is only slightly rotated in the mutant, its orientation stabilised by the hydrogen bond to W35. It is interesting to note that the additional water molecules are not present in the E7 DNase–Im7 complex in which Phe86 is replaced by a lysine residue.

#### E9 DNase N75A–Im9 complex

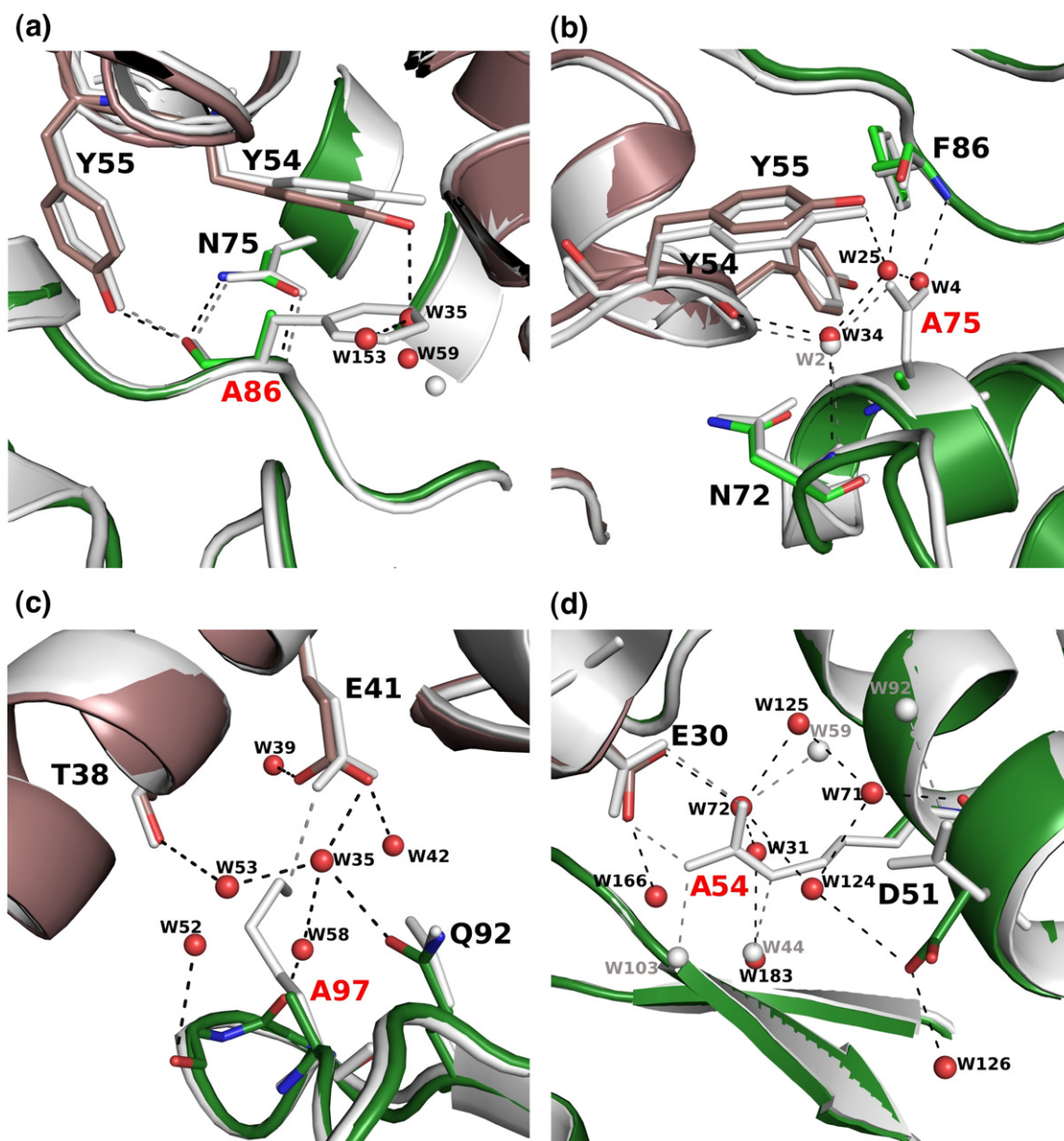
Two water molecules substitute for the side chain of the residue, W4 and W25, both having almost zero accessibility to bulk solvent. In the native complex, Asn75 is hydrogen bonded to W2, one of the three conserved water molecules at the interface, which mediates hydrogen bonds between the DNase domain and Im protein. W2 is present in the mutant structure (W34) and is hydrogen bonded to W25.

#### E9 DNase K97A–Im9 complex

The side chain of Lys97 in the wild-type complex is in a salt bridge with Im9 Glu41, one of only two salt bridges at the interface. Substitution of the lysine for alanine results in four water molecules, W35, W52, W53 and W58, entering the interface, none of which is found in the E7 DNase–Im7 complex where this specificity residue is replaced by a threonine. The accessibility values of these four water molecules are slightly higher than those of the water molecules in the other mutant complexes (accessibility values of 3, 10, 2 and 11, respectively). Other water molecules are also present in the mutant complex but have much higher accessibility indices, and so we do not consider them here. In the absence of Lys97, Im9 Glu41 is slightly rotated and hydrogen bonded to three water molecules: W35, W39 and W42. This mutant was the only one to reveal long-range structural changes wherein Arg54 from the enzyme swivels away from its salt bridge partner, Im9 Glu30, in order to form a salt bridge with a sulfate ion that was present in the crystallisation buffer. No other structure change was evident between Lys97 and Arg54, which are 14 Å apart, suggesting that the reorientation of Arg54 is solely due to the fortuitous location of the sulfate ion.

#### E9 DNase R54A–Im9 complex

The second salt bridge at the interface of the wild-type complex is formed between the side chain of E9 DNase Arg54 and Im9 Glu30. In the R54A mutant, the side chain of the arginine is replaced by four water molecules that are hydrogen bonded to other water molecules and to Im9 Glu30, which is almost in its native conformation. The side chain of E9



**Fig. 5.** Crystal structures of E9 DNase alanine mutants in complex with Im9. See Table 3 for details of refinement statistics. (a) E9 DNase F86A. (b) E9 DNase N75A. (c) E9 DNase K97A. (d) E9 DNase R54A. See the text for details of the structural changes in the immediate vicinity of each mutation and the changes in bound water. All mutant structures (green, with additional water molecules shown in red) are overlaid with the wild-type structure (gray).

DNase Asp51 is rotated in this mutant and hydrogen bonded to two water molecules.

#### Experimental and computational DMC analyses of Im protein residues surrounding the key E9 DNase specificity determinant Phe86

The close interconnections between stabilising hotspot residues of Im9 with its own specificity residues and those of E9 DNase Phe86 prompted us to use DMC analysis to investigate the energetic linkage of these residues at the core of the protein–protein interaction. We also investigated these linkages for Im2 binding E9 DNase in a non-cognate context to determine their importance to specificity.

DMCs have been used extensively for investigating the architecture and organisation of protein–protein complexes, where both super- and subadditivity of alanine mutations have been recorded.<sup>4,41–44</sup> In the current analysis, the interaction energy ( $\Delta\Delta G_{\text{int}}$ ) between two side chains was calculated using Eq. (3), which is the difference between the sum of  $\Delta\Delta G$  for each individual alanine mutant and that for the double mutant.

$$\Delta\Delta G_{\text{int}} = (\Delta\Delta G_{\text{DNase}}^{\text{E9}} + \Delta\Delta G_{\text{bind}}^{\text{Im}}) - \Delta\Delta G_{\text{bind}}^{\text{E9 DNase, Im}} \quad (3)$$

In general, mutation of interface residues that are not in close proximity ( $>7 \text{ \AA}$ ) will tend to produce

**Table 3.** Data collection and structure refinement statistics for complexes of E9 DNase alanine mutants binding Im9

	E9 DNase R54A–Im9	E9 DNase N75A–Im9	E9 DNase F86A–Im9	E9 DNase K97A–Im9
<i>Data collection</i>				
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Unit cell dimensions				
$a$ (Å)	36.28	36.28	36.05	36.38
$b$ (Å)	77.92	77.93	77.49	78.61
$c$ (Å)	41.40	41.72	41.81	40.21
$\beta$ (°)	100.07	100.06	100.21	99.35
Resolution (Å)	30.0–2.0	40.0–1.6	40.0–1.6	30.0–1.8
Unique reflections	15,357	29,843	29,183	19,052
Completeness (%) <sup>a</sup>	98.4 (98.4)	98.9 (98.9)	97.8 (85.4)	91.9 (91.9)
Multiplicity <sup>a</sup>	4.3 (4.3)	7.6 (4.8)	3.9 (3.1)	2.8 (1.9)
$R_{\text{meas}}$ (%) <sup>a</sup>	7.3 (33.5)	7.6 (39.3)	6.9 (34.3)	5.7 (25.1)
$I/\sigma(I)$ <sup>a</sup>	18.0 (5.8)	22.2 (3.9)	17.4 (4.1)	13.9 (3.9)
<i>Refinement</i>				
$R$ -factor (%) <sup>b</sup>	16.55	16.83	16.43	17.4
$R_{\text{free}}$ (%) <sup>b</sup>	22.86	19.80	20.31	22.8
RMSDs from ideal				
Bond lengths (Å)	0.018	0.017	0.019	0.019
Bond angles (°)	1.57	1.71	1.77	1.81
No. of protein atoms	1703	1706	1735	1762
No. of water molecules	212	329	300	218
Average protein $B$ -factor (Å <sup>2</sup> )	24.56	14.08	12.50	12.68
Average water $B$ -factor (Å <sup>2</sup> )	31.03	24.63	28.76	33.83
Ramachandran plot (%)				
Most favoured	95.1	96.7	95.6	94.7
Additional allowed	4.4	3.8	2.8	4.3
Generously allowed	0.5	0.5	1.6	1.0
Disallowed	0	0	0	0

<sup>a</sup> Numbers in parentheses correspond to the highest-resolution shell.

<sup>b</sup>  $R$ -factor =  $\sum_{hkl} ||F_{\text{obs}}| - k|F_{\text{calc}}|| / \sum_{hkl} |F_{\text{obs}}|$ ;  $R_{\text{free}}$ , the same for a test set of 5% reflections not used during refinement.

simple additive effects in DMCs, whereas complex patterns of additivity can result in cases where residues are in direct contact, signifying cooperative effects.<sup>41,44,45</sup> DMCs have the advantage over single alanine mutant studies in that they allow individual interactions to be quantified even if side chains interact with more than one other residue. They are also less sensitive to artifacts arising from mutations destabilising the structure of proteins (rather than binding directly) as these effects cancel out since they are present in both single- and double-mutant  $\Delta\Delta G$  values. Lastly, we compared experimental

values of  $\Delta\Delta G_{\text{int}}$  (observed interaction energy obtained from DMC analysis,  $\Delta\Delta G_{\text{int}}^{\text{obs}}$ ) with values calculated using ROSETTA (calculated interaction energy obtained from DMC analysis,  $\Delta\Delta G_{\text{int}}^{\text{calc}}$ ) for the cognate complex, as described in Materials and Methods and presented in Table 4, to determine whether the program is able to capture interaction energies for residues surrounding Phe86.

Considering first DMC data for the cognate complex, the measured and calculated values of  $\Delta\Delta G_{\text{int}}$  for the E9 F86A mutant binding Im9 L33A, Im9 V34A, Im9 V37A, Im9 Y54A and Im9 Y55A are

**Table 4.** DMC analysis of E9 DNase F86A binding Im9 alanine mutants

Double-mutant complex	$k_{\text{on}} \times 10^7$ (M <sup>-1</sup> s <sup>-1</sup> ) <sup>a</sup>	$k_{\text{off}}$ (s <sup>-1</sup> ) <sup>b</sup>	$K_{\text{d,kinetic}}$ (M) <sup>c</sup>	$\Delta\Delta G_{\text{kin}}^{\text{obs}}$ (kcal mol <sup>-1</sup> ) <sup>d</sup>	$\Delta\Delta G_{\text{int}}^{\text{obs}}$ (kcal mol <sup>-1</sup> ) <sup>e</sup>	$\Delta\Delta G_{\text{int}}^{\text{calc}}$ (kcal mol <sup>-1</sup> ) <sup>f</sup>
E9 F86A–Im9 L33A	5.83 (±0.4)	0.15 (±0.01)	$2.57 (\pm 0.4) \times 10^{-9}$	6.9	0.4	0.4
E9 F86A–Im9 V34A	5.79 (±0.1)	$4.97 (\pm 0.23) \times 10^{-3}$	$8.58 (\pm 0.6) \times 10^{-11}$	4.8	1.6	0.3
E9 F86A–Im9 V37A	5.44 (±0.04)	$1.7 (\pm 0.1) \times 10^{-3}$	$3.13 (\pm 0.1) \times 10^{-11}$	4.3	1.3	0.5
E9 F86A–Im9 Y54A	3.93 (±0.05)	3.18 (±0.02)	$8.09 (\pm 0.16) \times 10^{-8}$	8.9	−0.2	1.1
E9 F86A–Im9 Y55A	8.07 (±0.04)	3.00 (±0.04)	$3.72 (\pm 0.07) \times 10^{-8}$	8.4	0.1	0.0

Experimental conditions and error analysis are as those described in Table 1. Dissociation constant data for wild-type E9 DNase binding Im9 mutants L33A, V34A, V37A, Y54A and Y55A were taken from Wallis *et al.*<sup>17</sup>

<sup>a</sup> E9 DNase F86A–Im9  $k_{\text{on}}$  values were determined as described in Table 1.

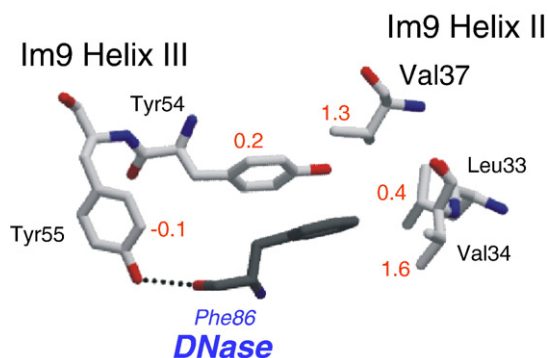
<sup>b</sup> E9 DNase F86A–Im9  $k_{\text{off}}$  values were determined as described in Table 1.

<sup>c</sup> Kinetically determined equilibrium dissociation constants.

<sup>d</sup> Change in the free energy of binding E9 DNase F86A for Im9 alanine mutants determined kinetically.

<sup>e</sup> Observed  $\Delta\Delta G_{\text{int}}$  values determined according to Eq. (3).

<sup>f</sup> Calculated values of  $\Delta\Delta G_{\text{int}}$  determined as described in Materials and Methods.



**Fig. 6.** DMC analysis of Im9 alanine mutants binding E9 DNase F86A indicating that interaction energies at the core of the interface are largely additive. See Table 4 and the text for details. Values in red are interaction energies ( $\Delta\Delta G_{\text{int}}^{\text{obs}}$ , according to Eq. (3)) for residues in close proximity to the E9 DNase specificity residue Phe86 determined by DMC analysis. Residues from helix II (that help define specificity) and helix III (that stabilise the complex) of Im9 form a hydrophobic cage that accepts the specificity residue of the enzyme.

tabulated in Table 4, with  $\Delta\Delta G_{\text{int}}^{\text{obs}}$  data presented in Fig. 6. Both experimental and computational values of  $\Delta\Delta G_{\text{int}}$  for interactions centred on the specificity hotspot residue E9 DNase Phe86 are small and positive ( $\sim 1.3$ – $1.6$  kcal/mol), which indicates that in each case the single mutations have a greater effect than the double mutant but not by much relative to experimental error. Interestingly, the largest interaction energies recorded for the phenyl ring of E9 DNase Phe86 are toward the specificity sites of Im9 and not Tyr54, which presents Phe86 toward the specificity sites, or Tyr55, which closes off solvent access to the core of the interface.

$\Delta\Delta G_{\text{int}}^{\text{calc}}$  correctly predicted the additive effects between Phe86 and Leu33 in helix II and Tyr55 in helix III and the mildly superadditive effects between Phe86 and Val37 on Im9 (Table 4). However, ROSETTA did not capture the magnitude of the additivity for the E9 DNase Phe86–Im9 Val34 interaction. Lastly, for the interaction between Phe86 and Tyr54, the experimental data indicate modest or no cooperativity, whereas our calculations suggest this to be superadditive, likely reflecting the overestimation of the alanine contribution. We note that values of  $\Delta\Delta G_{\text{int}}^{\text{obs}}$  for hotspot residues in the colicin E9 DNase–Im9 complex are much smaller than those measured for other high-affinity nuclease–inhibitor complexes, such as barnase–barstar, where values of up to 7 kcal/mol have been recorded.<sup>41</sup> This reflects the differences in chemistries at the respective hotspots of these complexes, with localised non-polar interactions predominating at the core of the E9 DNase–Im9 complex but highly cooperative charge–charge interaction networks dominating at the barnase–barstar interface. The preponderance of van der Waals contacts for the specificity residue Phe86 at the core of this protein–protein interaction also explains the ability of

ROSETTA to correctly predict these interaction energies with reasonable accuracy.

DMC data for the non-cognate Im2 protein residues binding E9 DNase Phe86 reveal clear differences from those for Im9 (Table 5). No binding could be detected for either of the tyrosine substitutions of Im2, reiterating the importance of these conserved hotspots to non-cognate binding. In contrast to Im protein specificity sites in the cognate complex, which show mildly positive DMC values, those in Im2 are predominantly negative, highlighting the destabilising influences of these Im2 side chains to E9 DNase binding that are alleviated on substitution to alanine. In summary, DMC data for Im protein binding the E9 DNase Phe86Ala mutant emphasise the role Im protein residues 33 and 37 take in discriminating the key specificity determinant at position 86 in cognate and non-cognate DNases.

## Conclusions

Discrimination of related proteins at a single binding site is a common occurrence at hubs in intracellular protein–protein interaction networks. Dual recognition is a thermodynamic mechanism that underpins such discrimination, first described for colicin DNase–Im protein complexes<sup>13,17</sup> but also seen in other protein–protein interactions.<sup>46,47</sup> A conserved hotspot, which is a common feature of protein–protein interactions,<sup>34,48</sup> from the Im protein (helix III) provides broad colicin DNase cross-reactivity, while neighbouring residues (helix II) confer specificity. What remain unresolved are the thermodynamic basis for cognate Im protein recognition by the enzyme IPE, which is variable in

**Table 5.** DMC analysis of E9 DNase F86A binding Im2 alanine mutants

Double-mutant complex	$K_d$ (M) <sup>a</sup>	$\Delta\Delta G_{\text{ITC}}^{\text{obs}}$ (kcal mol <sup>-1</sup> ) <sup>b</sup>	$\Delta\Delta G_{\text{int}}^{\text{obs}}$ (kcal mol <sup>-1</sup> ) <sup>c</sup>
E9 F86A–Im2 D33A	$1.20 (\pm 0.22) \times 10^{-7}$	-0.3	-1.7
E9 F86A–Im2 N34A	$1.47 (\pm 0.08) \times 10^{-6}$	1.2	1.4
E9 F86A–Im2 V37A	$2.87 (\pm 0.15) \times 10^{-5}$	2.9	-0.3
E9 F86A–Im2 Y54A	n/d	n/d	n/d
E9 F86A–Im2 Y55A	n/d	n/d	n/d

“n/d” represents values not determined due to lack of any measured binding for the indicated combination of mutants.

<sup>a</sup> Equilibrium dissociation constants were determined by ITC. Dissociation constants for Im2–E9 DNase wild type, Im2–E9 DNase F86A, Im2 D33A–E9 DNase wild type, Im2 N34A–E9 DNase wild type and Im2 V37A–E9 DNase wild type used in the analysis were determined by ITC as  $2.07 (\pm 0.29) \times 10^{-7}$  M,  $1.44 (\pm 0.10) \times 10^{-6}$  M,  $1 \times 10^{-9}$  M,  $2.3 (+0.29) \times 10^{-6}$  M and  $2.56 (\pm 0.11) \times 10^{-6}$  M, respectively (data not shown).

<sup>b</sup> Observed change in the free energy of binding for double-mutant combinations determined by ITC.

<sup>c</sup> Observed  $\Delta\Delta G_{\text{int}}$  values determined according to Eq. (3).

sequence, and how non-cognate Im proteins are distinguished.

We addressed these questions using a combination of experimental and computational alanine scanning and DMC analysis.  $\Delta\Delta G^{obs}$  and  $\Delta\Delta G^{calc}$ , determined using ROSETTA, for Im9 binding 12 IPE alanine mutants showed reasonable agreement. The inclusion of solvated rotamer models for ROSETTA calculations correctly predicted the location and energetic contribution of one important conserved water-mediated interaction between an E9 DNase residue (Asn75) and Im9, although this was not replicated for other water-mediated interactions. Surprisingly,  $\Delta\Delta G$ s for E9 DNase IPE mutants are substantially smaller than those for Im9, particularly for conserved Im9 residues where many of the contacts are to the DNase backbone.<sup>17</sup> Four E9 DNase residues comprise the hotspot for the IPE, forming a belt across the centre of the interface, three of which are also the points of greatest discrimination between Im9 and Im2. The alanine mutation with the greatest observed and predicted effects on  $\Delta\Delta G$  for Im9 binding was E9 DNase Phe86. Phe86 is the centrepoint of the interface and is coordinated by the conserved hotspot of Im9 (Tyr54 and Tyr55) along with its hydrophobic specificity residues (Val34 and Val37). The two tyrosines form a clamp that docks helix III of the Im protein onto the DNase IPE cleft and form the rotation axis that expedites rigid-body rotations of the Im protein over the DNase surface in the conformational search for specificity contacts.<sup>13,19</sup> Consistent with the hydrophobic nature of the contacts at the centre of the interface, experimental and computational DMC analyses suggested only minor additivity effects that were nonetheless significantly altered in the non-cognate Im2 complex, further supporting the role of these contacts in conferring specificity. Although Phe86 is a key specificity site in colicin DNases (Arg and Lys in E8 and E7, respectively), phenylalanine is nevertheless conserved between the E2 and E9 DNases, suggesting that the energetic contribution of this residue is context dependent, the context being the surrounding specificity residues. A similar effect is seen for Asn75, whose water-mediated interaction is conserved in colicin DNase–Im protein complexes,<sup>13</sup> which differentially stabilises the Im9 complex over that of Im2. Additional discrimination of Im9 over Im2 is accomplished through neighbouring specificity contacts, principally through Lys97 and Val98 on the DNase contacting sites on helix II on the Im protein.

In conclusion, colicin DNase specificity for Im proteins is determined by both conserved and variable amino acids within the IPE, with the binding free energy contribution of conserved contact points modulated by neighbouring variable specificity sites. Such a mechanism has similarities with the framework effects of constant heavy regions on the affinity and specificity of antibodies for antigens,<sup>49</sup> although, in contrast to what we observe in colicin DNase–Im protein complexes, these effects are indirect.

## Materials and Methods

### Selection of E9 DNase IPE residues for alanine mutagenesis

Thirteen IPE residues (Arg54, Asn72, Ser74, Asn75, Ser77, Ser78, Ser84, Tyr83, Phe86, Thr87, Gln92, Lys97 and Val98) were selected for alanine mutagenesis based on three criteria: (1) significant burial upon Im9 binding (change in accessible surface area,  $>10 \text{ \AA}^2$ ); (2) readily identifiable contacts with Im9 in the high-resolution structure (1.7 Å) of the E9 DNase–Im9 complex<sup>13</sup>; and (3) implication in Im protein specificity from previous studies.<sup>12</sup> Some residues that become buried in the complex with Im9 but whose side chains make no discernible contact with the inhibitor (Ser74, Ser84 and Thr87) were included. With the exception of Arg54, all IPE residues reside within a single linear epitope in the enzyme (residues 72–98), many of which have previously been implicated in conferring Im protein specificity (Ser77, Ser78, Tyr83, Lys97 and Val98).<sup>12</sup> Arg54 was also mutated because, although not part of the linear epitope of the IPE, it forms a salt bridge with Im9 Glu30, a helix II residue previously shown to be important for E9 DNase binding.<sup>17</sup> The 13 residues selected for mutagenesis are underlined in the DNase domain sequence alignment (Fig. 1a) and shown on the molecular surface of the E9 DNase IPE (Fig. 1b).

### Site-directed mutagenesis of the E9 DNase

Plasmid pRJ353 (based on pET21d from Novagen), containing the gene for the wild-type E9 DNase and Im9,<sup>50</sup> was used as the template for the mutagenesis, which was carried out by the QuikChange PCR mutagenesis method (Stratagene), with primers purchased from MWG Biotech (Germany). Mutations were confirmed by sequencing (MWG Biotech). In all cases, the identity of purified proteins was confirmed by electrospray ionisation mass spectrometry.

### Protein purification and quantification

Purifications of E9 DNase, Im2, Im9 and tritiated Im9 ( $[^3\text{H}]\text{Im9}$ ) have all been described previously.<sup>14–16,51</sup> Protein concentrations were by absorbance at 280 nm using molar absorption coefficients of  $17,550 \text{ M}^{-1} \text{ cm}^{-1}$  for the E9 DNase domain and  $11,400 \text{ M}^{-1} \text{ cm}^{-1}$  for Im proteins determined previously.<sup>16</sup>

### Biophysical characterisation of E9 DNase alanine mutants

Preliminary assessment of the effect of alanine substitution on the stability of the protein was undertaken of all purified mutants by intrinsic tryptophan fluorescence emission spectroscopy and far-UV CD spectroscopy, as previously reported for colicin E9 DNase domain,<sup>51</sup> and shown to be indistinguishable from wild type. In addition, intrinsic tryptophan fluorescence spectroscopy was used to follow the temperature denaturation of each protein, from which was obtained the thermal denaturation midpoint.<sup>51</sup> For all mutants, the thermal denaturation midpoint was similar to the wild-type protein.

Fluorescence spectra of 1  $\mu\text{M}$  wild-type and alanine mutant E9 DNases in 50 mM triethanolamine, pH 7.5 and

25 °C, were recorded on a Spex Fluoromax 3. Tryptophan fluorescence was excited at 295 nm with fluorescence emission monitored from 300 to 450 nm. In all cases, the peak emission wavelength was close to 333 nm, typical of a folded protein and close to that previously observed for the E9 DNase<sup>51</sup> under similar conditions.

CD spectra of metal-free E9 DNase alanine mutants (typically at a concentration of 10 µM) were taken after dialysis into 10 mM potassium phosphate, pH 7.5, overnight at 4 °C. Spectra were recorded every nanometer from 190 to 260 nm in a Jasco J-810 spectropolarimeter in a 1-cm pathlength cell, with 10 measurements averaged to produce each final spectrum. In all cases, spectra recorded at 25 °C resembled closely those previously observed for the wild-type E9 DNase.<sup>51</sup>

Thermal denaturation profiles of mutant proteins were also followed by far-UV CD spectroscopy. Proteins were dialysed into 10 mM potassium phosphate, pH 7.5, and put into a stirred 1-cm pathlength cell within a Jasco J-810 spectropolarimeter. Spectra were recorded as described above every 5 °C from 15 to 75 °C, and the wavelength of maximum change (typically 198 or 203 nm) was used to quantify the change in CD signal with temperature. Fitted  $T_m$  values were close to those previously observed for the E9 DNase.<sup>51</sup>

### E9 DNase–Im protein dissociation constant ( $K_d$ )

All experiments were carried out in 50 mM Mops buffer, pH 7.0, containing 200 mM NaCl and 1 mM DTT at 25 °C. The bimolecular rate constant ( $k_{on}$ ) for E9 DNase–Im protein association was measured by stopped-flow fluorescence using pseudo-first-order or second-order binding kinetics as appropriate as described by Wallis *et al.*<sup>14</sup> and Li *et al.*<sup>16</sup> The  $k_{off}$  was determined by radioactive subunit exchange for slow  $k_{off}$  values ( $<10^{-3} s^{-1}$ ) or from fluorescence chase stopped-flow experiments for fast  $k_{off}$  values ( $>10^{-3} s^{-1}$ ) as described by Li *et al.*<sup>16</sup> and Keeble and Kleanthous.<sup>19</sup> The equilibrium dissociation constant was calculated from the ratio of the  $k_{off}$  and  $k_{on}$  ( $K_d = k_{off}/k_{on}$ ). In the case of non-cognate E9 DNase–Im2 complexes, all equilibrium  $K_d$  data were verified by ITC, as described by Keeble and Kleanthous.<sup>19</sup>

### Protein crystallography of E9 DNase mutants in complex with Im9

Crystals for the four mutants of E9 (R54A, N75A, F86A and K97A) in complex with Im9 were grown by the hanging-drop vapour-diffusion method. The crystallisation condition was 2.9 M ammonium sulfate and 100 mM Bis–Tris, pH 5.5, with diffraction-quality crystals obtained by iterative seeding in fresh equilibrated drop. Crystals were flash-cooled using sodium malonate as cryo-protectant. A data set at 1.8-Å resolution was collected at the beamline 10.1 at the Daresbury Synchrotron Radiation Source (Daresbury, England) for the mutant complex E9 K97A–Im9, while the other three mutant complex data sets were collected using the beamline ID14.1 at the European Synchrotron Radiation Facility (Grenoble, France). All data sets were integrated and scaled using MOSFLM and SCALA.<sup>52</sup> Table 3 shows the final statistics for each of the four structures. All crystals belonged to space group  $P2_1$ , as per the wild-type complex, with one molecule in the asymmetric unit. The structures were solved by molecular replacement using the program MOLREP<sup>53</sup> and the structure of the

wild-type E9 DNase–Im9 complex (Protein Data Bank code 1EMV) as the model. Structures were refined using REFMAC,<sup>54</sup> and rebuilding was done with Coot.<sup>55</sup> The final electron density map allowed us to identify most of the residues from the sequence. For the E9 DNase K97A–Im9 complex, six molecules of ammonium sulfate were modeled; for each of the other three complexes, three molecules of malonic acid were identified, all originating from the crystallisation conditions or cryo-protectant.

### Computational alanine scan on interfaces

Binding free energy changes upon alanine mutation were calculated using previously described equations.<sup>36</sup> For modeling side-chain conformational changes on complex formation and mutation, side chains were represented as rotamers on a fixed backbone template. Rotamers were taken from a backbone-dependent library by Bower *et al.*,<sup>56</sup> with additional rotamers added by rotations around the  $\lambda_1$  and  $\lambda_2$  angles by 5°–20° and extra rotamers for  $\lambda_3$  and  $\lambda_4$  angles as described by Dahiyat and Mayo.<sup>57</sup> The X-ray coordinates of the native side chains at each position were included in the library for both the complex and the isolated partners. All residues having at least one side-chain atom within a sphere of 5-Å radius of the site of mutation were repacked. All other amino acid side chains were left in the conformations observed in the parent crystal structures. Energies were computed for each rotamer with the constant part of the molecule (the template backbone and all unchanged side chains) and for all pairwise rotamer–rotamer combinations by using the free energy function described by Kortemme and Baker.<sup>36</sup> Global optimisation of side-chain conformations was performed using a Monte Carlo simulated annealing procedure,<sup>27</sup> in which a move consisted of the replacement of a randomly picked side-chain rotamer at a single position by another rotamer from the library.

### Application of the model to computational alanine scanning on protein interfaces

The energy function was parametrised using a data set of experimentally determined stability changes taken from the PROTHERM database as described previously.<sup>36</sup> The parametrised free energy function was used to compute  $\Delta\Delta G^{calc}$  and  $G_{int}^{calc}$  for the E9 DNase–Im9 Im protein complex and compared with the experimentally measured changes in binding energy on alanine mutagenesis and experimentally derived interaction energies from the DMC analysis.

### Solvated rotamer approach

For the solvated rotamer approach, we started with the backbone-dependent library compiled by Bower *et al.*<sup>56</sup> and added water molecules at allowed backbone and polar side-chain atoms as described by Jiang *et al.*<sup>28</sup> Energies were computed using the free energy function described by Kortemme and Baker<sup>36</sup> with an additional protein–water interaction composed of a Lennard–Jones repulsive potential and a water-mediated hydrogen bonding potential. Global energy optimisation was performed as in the work of Kuhlman and Baker.<sup>27</sup> The

free energy function was parametrised as previously described by Kortemme and Baker.<sup>36</sup>

### Accession numbers

All mutant complex structures have been deposited in the Protein Data Bank with the following coordinate/structure factor codes: 2vlq/r2vlqsf for F86A, 2v1n/r2v1nsf for N75A, 2vlo/r2vlof for K97A and 2vlp/r2vlpsf for R54A.

### Acknowledgements

This work was funded by Biotechnology and Biological Sciences Research Council grants (C.K.) and a grant from the National Institutes of Health (L.J. and D.B.). We thank Ansgar Pommer and Ulrike Kuhlmann (Magdeburg) for their help and advice during the early stages of this work.

### References

- Han, J. D., Bertin, N., Hao, T., Goldberg, D. S., Berriz, G. F., Zhang, L. V. *et al.* (2004). Evidence for dynamically organized modularity in the yeast protein–protein interaction network. *Nature*, **430**, 88–93.
- Kim, P. M., Lu, L. J., Xia, Y. & Gerstein, M. B. (2006). Relating three-dimensional structures to protein networks provides evolutionary insights. *Science*, **314**, 1938–1941.
- DeLano, W. L., Ultsch, M. H., de Vos, A. M. & Wells, J. A. (2000). Convergent solutions to binding at a protein–protein interface. *Science*, **287**, 1279–1283.
- Laskowski, M., Jr., Qasim, M. A. & Yi, Z. (2003). Additivity-based prediction of equilibrium constants for some protein–protein associations. *Curr. Opin. Struct. Biol.* **13**, 130–139.
- Pawson, T. & Nash, P. (2003). Assembly of cell regulatory systems through protein interaction domains. *Science*, **300**, 445–452.
- Kleanthous, C. & Walker, D. (2001). Immunity proteins: enzyme inhibitors that avoid the active site. *Trends Biochem. Sci.* **26**, 624–631.
- 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.
- James, R., Penfold, C. N., Moore, G. R. & Kleanthous, C. (2002). Killing of *E. coli* cells by E group nuclease colicins. *Biochimie*, **84**, 381–389.
- Kleanthous, C. & Pommer, A. J. (2000). Protein–protein recognition. In (Kleanthous, C., ed.), Oxford University Press, Oxford, UK.
- 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.
- 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.
- Curtis, M. D. & James, R. (1991). Investigation of the specificity of the interaction between colicin E9 and its immunity protein by site-directed mutagenesis. *Mol. Microbiol.* **5**, 2727–2733.
- 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.
- 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.
- 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.
- 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.
- 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**, 476–485.
- 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.
- Keeble, A. H. & Kleanthous, C. (2005). The kinetic basis for dual recognition in colicin endonuclease–immunity protein complexes. *J. Mol. Biol.* **352**, 656–671.
- Rajamani, D., Thiel, S., Vajda, S. & Camacho, C. J. (2004). Anchor residues in protein–protein interactions. *Proc. Natl. Acad. Sci. USA*, **101**, 11287–11292.
- Keeble, A. H., Kirkpatrick, N., Shimizu, S. & Kleanthous, C. (2006). Calorimetric dissection of colicin DNase–immunity protein complex specificity. *Biochemistry*, **45**, 3243–3254.
- Goh, C. S. & Cohen, F. E. (2002). Co-evolutionary analysis reveals insights into protein–protein interactions. *J. Mol. Biol.* **324**, 177–192.
- Bernath, K., Magdassi, S. & Tawfik, D. S. (2005). Directed evolution of protein inhibitors of DNA-nucleases by *in vitro* compartmentalization (IVC) and nano-droplet delivery. *J. Mol. Biol.* **345**, 1015–1026.
- 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.
- 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.
- Maté, M. J. & Kleanthous, C. (2004). Structure-based analysis of the metal-dependent mechanism of H–N–H endonucleases. *J. Biol. Chem.* **279**, 34763–34769.
- Kuhlman, B. & Baker, D. (2000). Native protein sequences are close to optimal for their structures. *Proc. Natl. Acad. Sci. USA*, **97**, 10383–10388.
- Jiang, L., Kuhlman, B., Kortemme, T. & Baker, D. (2005). A “solvated rotamer” approach to modeling water-mediated hydrogen bonds at protein–protein interfaces. *Proteins*, **58**, 893–904.

29. Janin, J. (1997). The kinetics of protein–protein recognition. *Proteins*, **28**, 153–161.
30. Schreiber, G. (2002). Kinetic studies of protein–protein interactions. *Curr. Opin. Struct. Biol.* **12**, 41–47.
31. Clackson, T. & Wells, J. A. (1995). A hot spot of binding energy in a hormone–receptor interface. *Science*, **267**, 383–386.
32. Bogan, A. & Thorn, K. S. (1998). Anatomy of hot spots in protein interfaces. *J. Mol. Biol.* **280**, 1–9.
33. Chakrabarti, P. & Janin, J. (2002). Dissecting protein–protein recognition sites. *Proteins*, **47**, 334–343.
34. 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.
35. DeLano, W. L. (2002). Unraveling hot spots in binding interfaces: progress and challenges. *Curr. Opin. Struct. Biol.* **12**, 14–20.
36. 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.
37. 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. *Nat. Struct. Biol.* **9**, 621–627.
38. Shifman, J. M. & Mayo, S. L. (2002). Modulating calmodulin binding specificity through computational protein design. *J. Mol. Biol.* **323**, 417–423.
39. Havranek, J. J. & Harbury, P. B. (2003). Automated design of specificity in molecular recognition. *Nat. Struct. Biol.* **10**, 45–52.
40. Dantas, G., Kuhlman, B., Callender, D., Wong, M. & Baker, D. (2003). A large scale test of computational protein design: folding and stability of nine completely redesigned globular proteins. *J. Mol. Biol.* **332**, 449–460.
41. Schreiber, G. & Fersht, A. R. (1995). Energetics of protein–protein interactions: analysis of the barnase–barstar interface by single mutations and double mutant cycles. *J. Mol. Biol.* **248**, 478–486.
42. Goldman, E. R., Dall’Acqua, W., Braden, B. C. & Mariuzza, R. A. (1997). Analysis of binding interactions in an idiotope–antiidiotope protein–protein complex by double mutant cycles. *Biochemistry*, **36**, 49–56.
43. Chen, C. Z. & Shapiro, R. (1999). Superadditive and subadditive effects of “hot spot” mutations within the interfaces of placental ribonuclease inhibitor with angiogenin and ribonuclease A. *Biochemistry*, **38**, 9273–9285.
44. Reichmann, D., Rahat, O., Albeck, S., Meged, R., Dym, O. & Schreiber, G. (2005). The modular architecture of protein–protein binding interfaces. *Proc. Natl. Acad. Sci. USA*, **102**, 57–62.
45. 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.
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. Su, Z., Osborne, M. J., Xu, P., Xu, X., Li, Y. & Ni, F. (2005). A bivalent dissectional analysis of the high-affinity interactions between Cdc42 and the Cdc42/Rac interactive binding domains of signaling kinases in *Candida albicans*. *Biochemistry*, **44**, 16461–16474.
48. Keskin, O., Ma, B. & Nussinov, R. (2005). Hot regions in protein–protein interactions: the organization and contribution of structurally conserved hot spot residues. *J. Mol. Biol.* **345**, 1281–1294.
49. Torres, M. & Casadevall, A. (2008). The immunoglobulin constant region contributes to affinity and specificity. *Trends Immunol.* **29**, 91–97.
50. Garinot-Schneider, C., Pommer, A. J., Moore, G. R., Kleanthous, C. & James, R. (1996). Identification of putative active-site residues in the DNase domain of colicin E9 by random mutagenesis. *J. Mol. Biol.* **260**, 731–742.
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**, 27153–27160.
52. Collaborative Computational Project No. 4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **50**, 760–763.
53. Vagin, A. & Teplyakov, A. (1997). MOLREP: an automated program for molecular replacement. *J. Appl. Crystallogr.* **30**, 1022–1025.
54. 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.
55. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **60**, 2126–2132.
56. Bower, M. J., Cohen, F. E. & Dunbrack, R. L., Jr. (1997). Prediction of protein side-chain rotamers from a backbone-dependent rotamer library: a new homology modeling tool. *J. Mol. Biol.* **267**, 1268–1282.
57. Dahiyat, B. I. & Mayo, S. L. (1997). *De novo* protein design: fully automated sequence selection. *Science*, **278**, 82–87.