

Following evolutionary paths to protein-protein interactions with high affinity and selectivity

Kalia Bernath Levin^{1,2}, Orly Dym³, Shira Albeck³, Shlomo Magdassi², Anthony H Keeble^{4,5}, Colin Kleanthous⁴ & Dan S Tawfik¹

How do intricate multi-residue features such as protein-protein interfaces evolve? To address this question, we evolved a new colicin-immunity binding interaction. We started with Im9, which inhibits its cognate DNase ColE9 at 10⁻¹⁴ M affinity, and evolved it toward ColE7, which it inhibits promiscuously ($K_d > 10^{-8}$ M). Iterative rounds of random mutagenesis and selection toward higher affinity for ColE7, and selectivity (against ColE9 inhibition), led to an ~10⁵-fold increase in affinity and a 10⁸-fold increase in selectivity. Analysis of intermediates along the evolved variants revealed that changes in the binding configuration of the Im protein uncovered a latent set of interactions, thus providing the key to the rapid divergence of new Im7 variants. Overall, protein-protein interfaces seem to share the evolvability features of enzymes, that is, the exploitation of promiscuous interactions and alternative binding configurations via ‘generalist’ intermediates, and the key role of compensatory stabilizing mutations in facilitating the divergence of new functions.

The evolutionary routes and mechanisms that led to today’s proteins are usually hypothesized from sequence and structure comparisons. However, contemporary proteins are separated by numerous amino acid exchanges, most of which are unrelated to changes in structure or function. Understanding how new proteins evolve therefore remains a challenge¹. Divergence routes leading to new enzyme, vision proteins, or receptor specificities have been studied previously^{2–6}. These revealed how individual mutations, and the early mutation in particular, affected the evolving function¹, whether the mutational paths are largely additive or restricted by epistasis⁴, and how the newly evolving function trades off with the original function (‘generalist’ or ‘specialist’ intermediates)^{7–9}. Here we explore the divergence of a new protein-protein interaction, and we investigate the unraveled routes through which new high-affinity and high-selectivity protein-protein interactions can evolve.

We studied members of the colicin-immunity family. Colicins (ColEs) are bacterial SOS-induced DNA nuclease toxins¹⁰. To protect the producing cell, each ColE is coexpressed with a cognate, high-affinity inhibitor, dubbed immunity protein (Im), that is subsequently removed. The ColE–Im family has been under strong selection for divergence¹¹, and it therefore provides an example for rapidly evolving pairs of tight and selective protein-protein interactions¹². Analysis of the existing *Escherichia coli* ColE–Im pairs led to the ‘dual recognition’ hypothesis, by which new ColE–Im interfaces diverged while maintaining a few key interacting residues that are conserved throughout the family and that serve as a common anchoring point (the ‘conserved hotspot’). In contrast, other parts of the interface (the

‘variable region’) changed substantially to obtain high affinity and selectivity. The conserved hotspot enables weak cross-reactivity with noncognate partners ($K_d = 10^{-9}$ M to 10⁻⁴ M), in contrast to the high-affinity cognate binding ($\leq 10^{-14}$ M). However, the binding interface of each partner comprises about a dozen contacting residues that are in themselves interconnected^{12–14}. So how would a new ColE–Im interface evolve while maintaining binding capability throughout, in particular, when both high affinity and selectivity are needed?

In the laboratory evolution experiment described here, we began with wild-type Im9 and evolved it toward the inhibition of ColE7 nuclease. Successive rounds of mutagenesis, recombination and selection for inhibition of ColE7, and then for selectivity (that is, against inhibition of ColE9), gave rise to a series of new, *in vitro*-evolved Im7 proteins (nevIm7s) that showed increasing levels of inhibition potency toward ColE7. We could thereby monitor the sequence, structure and functional changes that lead to inhibitors of ColE7 with high affinity and selectivity. And, we could recapitulate *in vitro* certain aspects related to the divergence of the natural homologs of this family¹¹.

RESULTS

In vitro evolution of Im7 variants

In previous work, we selected ColE7 inhibitors by *in vitro* compartmentalization (IVC), in which emulsion droplets serve as artificial, cell-like compartments¹⁵. In each round of selection, a library of ~10¹⁰ Im variants was compartmentalized, at an average occupancy of one gene per droplet, together with an inactive ColE7 DNase. Following the transcription and translation of the Im genes within

¹Department of Biological Chemistry, Weizmann Institute of Science, Rehovot, Israel. ²Casali Institute of Applied Chemistry, The Hebrew University of Jerusalem, Givat Ram, Jerusalem, Israel. ³Israel Structural Proteomics Center, Weizmann Institute of Science, Rehovot, Israel. ⁴Department of Biology (Area 10), University of York, York, UK. ⁵Present address: Departments of Biochemistry and Infection, Immunity and Inflammation, University of Leicester, Leicester, UK. Correspondence should be addressed to D.S.T. (tawfik@weizmann.ac.il) or O.D. (orly.dym@weizmann.ac.il).

Received 7 April; accepted 4 August; published online 13 September 2009; doi:10.1038/nsmb.1670

Table 1 Inhibition potency, binding affinities and stability of Im variants

Variant ^a	Sequence ^b								<i>In vivo</i> inhibition potency and affinity					
	Helix I-II connecting loop				Helix II				ColE7			ColE9		
	24	26	27	28	34	37	41	55	Urea denaturation D_{50} (M)	<i>In vivo</i> protection ColE7 (M) ^c	K_d (M) ^d	ColE7/ColE9	<i>In vivo</i> protection ColE9 (M) ^c	K_d (M) ^d
Wild-type Im9	N	D	T	S	V	V	E	Y	4.9	$(3-10) \times 10^{-10}$	5.6×10^{-8e}	0.4×10^{-6}	$>10^{-4}$	2.4×10^{-14}
Im9 V34D										$(0.3-30) \times 10^{-9}$			$>10^{-4}$	
R5-4			A	T					4.2	$(3-10) \times 10^{-9}$	n.d. ^f		$>10^{-4}$	n.d. ^f
R5-10	D	N	S						3.25	$(1-10) \times 10^{-9}$	$<10^{-8f}$		$>10^{-4}$	n.d. ^f
R5-11			A	T	D				g	$(1-10) \times 10^{-9}$	$<10^{-8f}$		10^{-7} to 10^{-4}	n.d. ^f
R8-1	D		A	T					2.2	$(1-30) \times 10^{-9}$	n.d.		10^{-8} to 10^{-4}	n.d.
R8-6	D		A	T	D				g	$(0.1-30) \times 10^{-8}$	n.d.		3×10^{-7} to 10^{-4}	n.d.
R8-6 E41G	D		A	T	D		G			$(1-10) \times 10^{-9}$	n.d.		$(0.3-30) \times 10^{-8}$	
R8-7				T	D	I			3.7	3×10^{-9} to 10^{-4}	2.73×10^{-11}	0.34	$>10^{-4}$	9.3×10^{-12d}
R8-7 E41G				T	D	I	G			10^{-8} to 10^{-4}	n.d.		3×10^{-7} to 10^{-4}	
R8-8	D	N	A		D	I		W	n.d. ^h	10^{-7} to 10^{-4}	3.85×10^{-12}	4.5	$>10^{-4}$	1.46×10^{-11d}
R8-8 E41G	D	N	A		D	I	G	W		3×10^{-7} to 10^{-4}			3×10^{-7} to 10^{-4}	
R8-4	D		A	T	D	I			3.0	3×10^{-8} to 10^{-4}	n.d.		$>10^{-4}$	n.d.
R12-2	D		A	T	D	I	G		2.4	3×10^{-7} to 10^{-4}	$\sim 4 \times 10^{-12}$	33	3×10^{-8} to 10^{-4}	$\sim 1.3 \times 10^{-10}$
R12-4	D	N	A	T	D	I	G	W	n.d. ^h	$>10^{-4}$	$\leq 10^{-12}$	≥ 70	3×10^{-8} to 10^{-4}	$\sim 0.7 \times 10^{-10}$
R12-13	D	E	A	T	D	I		W	n.d. ^h	$>10^{-4}$	$\leq 10^{-12}$	≥ 14	$>10^{-4}$	$\sim 1.4 \times 10^{-11}$
Wild-type Im7	K	N	V	A	D	L	V	Y	2.5	$>10^{-4}$	$\sim 10^{-14}$	$\sim 10^{10}$	$(0.3-3) \times 10^{-9}$	$\sim 10^{-4e}$

^aThe numbering of the nevIm7 variants (for example, R8-7) indicates the round of mutation and selection from which the variant was isolated (in this example, round 8) and an arbitrary sequential number of the variant in the series analyzed (in this example, the seventh variant). ^bShown are all sequence changes observed at the variable region, which spans the loop connecting helices $\alpha 1$ and $\alpha 2$ (residues 24–30), helix $\alpha 2$ (residues 31–44) and the conserved hotspot (residues 51–56). The complete amino acid sequences are available in **Supplementary Figure 1**.

^c*In vivo* protection levels were determined with *E. coli* cells expressing the denoted Im variant, treated with increasing concentrations of either ColE7 or ColE9 (10^{-4} M to 10^{-11} M)¹⁹. The numbers indicate the maximal concentrations of ColEs under which growth remained unaffected under basal expression levels (no IPTG induction; first number within the value range) and full induction (1 M IPTG; second number). For a detailed description of these results, see **Supplementary Table 1**. ^dBinding parameters were determined by stopped-flow measurements³⁹ or SPR²⁵ (**Supplementary Fig. 5**). The K_d values were calculated from the kinetic parameters (k_{on} and k_{off}). A detailed analysis is provided in **Supplementary Table 1**. ^eBinding parameters from ref. 19.

^fThe dissociation rates for round 5 were off-range for SPR: dissociation from ColE7 was too fast, and too slow from ColE9. The higher affinity toward ColE7 relative to Im9 was therefore confirmed by competition experiments. These indicate that variants R5-10 and R5-11 show more than ten-fold higher affinity than Im9, because at a ratio of 1:8 with wild-type Im9 they were still able to displace Im9 (**Supplementary Fig. 2**). ^gDenaturation seems to be noncooperative and could not be fitted to a sigmoidal model (**Supplementary Fig. 6**). ^hThese variants carry the Trp55 mutation, which prevented detection of denaturation by fluorescence emission.

the droplets, the ColE7 was activated by the delivery of cobalt ions. Gene variants encoding an active inhibitor survived ColE7 digestion. We amplified these genes by PCR, mutated them and then subjected them to another round of selection. We gradually increased the selection pressure by augmenting the rate of ColE7 digestion, and after eight rounds of mutation and selection we isolated the library variants with higher ColE7 inhibition than the Im9 starting point¹⁵.

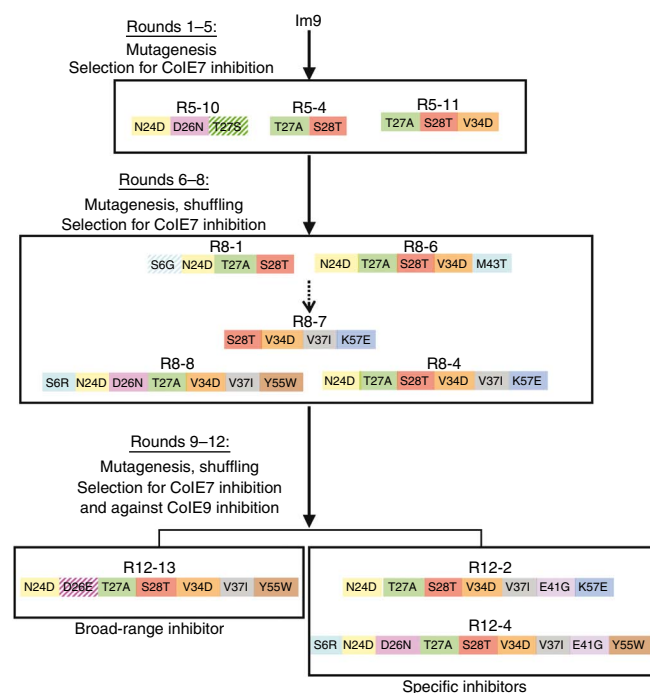
As described here, we subsequently performed a few more rounds of mutation and selection in the presence of increasing concentrations of ColE7, but these failed to appreciably increase affinity or reduce the cross-reactivity of the evolving variants toward ColE9. We therefore devised a dual selection pressure—for both affinity and selectivity. In addition to ColE7, we added a large excess of the ColE9 H103A mutant, which has no nuclease activity but binds the Im protein with the same affinity as wild-type ColE9 (ref. 16). The ColE9 mutant acted as a bait to attract Im9 binders. And so, only Im variants that selectively bind ColE7 over ColE9 could protect their encoding gene.

We followed up these three additional rounds of dual selections (beyond the eight rounds described previously)¹⁵ with a twelfth round that comprised an *in vivo* screen aimed at isolating Im variants that confer protection to *E. coli* at the highest concentration of ColE7 (ref. 17). The *in vivo* protection assay was also applied to monitor within a cellular context the activity of evolving variants from all rounds of selection and, thereby, provide a measure for the progress of the evolutionary process. The assay included the modulation of Im protein levels by varying concentrations of

the expression inducer IPTG, thus testing the nuclease's inhibitory concentration at basal Im levels (no IPTG) versus the medium and highest levels (0.05 mM and 1 mM IPTG, respectively)¹⁸. As demonstrated earlier¹⁹, the levels of *in vivo* protection with no induction are linearly correlated with $\log K_d$ for Im7 variants with affinities of 10^{-8} M to 10^{-11} M. Beyond that range, for affinities of $\sim 10^{-11}$ M to 10^{-14} M, we observed full protection at the highest nuclease levels ($\leq 10^{-4}$ M). The assay indicated that, by the end of the twelfth round, several evolved variants showed inhibition levels of ColE7 that were identical to that of wild-type Im7, within the plateau range of $\leq 10^{-11}$ M, for example, variant R12-4 and R12-13 (**Table 1**; variant nomenclature is as follows: R12-13 relates to variant 13 isolated from round 12). However, at basal expression levels (with no IPTG), most variants seemed inferior to the wild type in their ability to inhibit ColE7, for example, the maximal ColE7 concentration tolerated by variant R12-2 was 3×10^{-7} M, compared with $\geq 10^{-4}$ M for wild-type Im7. The selectivity of most twelfth round variants was also improved, as indicated by lower protection levels toward ColE9 (for example, variants R12-2 and R12-4). However, we also identified cross-reactive variants that strongly protect against ColE9 (for example, variant R12-13).

Characterizing the newly evolved Im7 variants

Phenotyping by the *in vivo* protection assay of the 34 best variants of the twelfth round of selection indicated two groups of inhibitors: selective (low cross-reaction with ColE9) and cross-reactive variants. The sequences of 11 variants from these two groups indicated that the cross-reactive variants carried the set of mutations represented



by variant R12-13 (Fig. 1). The selective inhibitors carried sets of mutations represented by either variant R12-2 or R12-4, the primary difference between these variants being the mutations Y55W and D26N. All the selective variants carried the E41G mutation that was absent in the cross-reactive variants. Following phenotyping and sequencing, we also chose three representative nevIm7 variants from the fifth round, when the first clear signs of increased ColE7 inhibition were observed¹⁵, and five variants from the eighth round, the last round before the application of the dual, affinity-selectivity selection. Because the sampling of variants along the evolutionary process is sparse (only 3 rounds out of 12 were systematically sampled), and owing to the high mutation rate, the chosen variants are separated by more than one amino acid exchange. We therefore arranged the 11 representative variants by their order of appearance and mutational composition to generate a putative genealogy that represents a likely divergence path from wild-type Im9 (Fig. 1). The detailed characterization of these representative nevIm7 variants and the engineering and characterization of variants carrying specific mutations seen in these variants enabled us to analyze the route leading to the high-affinity and high-selectivity nevIm7 inhibitors. Specifically, we overexpressed and purified the variants listed in Figure 1 and Table 1 and analyzed their binding parameters and stability. We also solved the crystal structures of the final generation variants R12-2 and R12-13 in complex with ColE7 (Table 2; PDB 3GKL and PDB 3GJN, respectively). All attempts to crystallize round 8 variants failed. These complexes did not even precipitate, and when they were further concentrated ($\leq 500 \text{ mg ml}^{-1}$) they lost their native structures (data not shown). We observed similar phenomena in our attempts to crystallize noncognate complexes²⁰.

The early steps of divergence

In round 5, only one mutation (V34D), which was seen in the minority of variants, occurred at the core of the ‘variable region’ within helix $\alpha 2$, in a key contacting residue¹⁴. This mutation exchanged a hydrophobic hotspot interaction in the Im9–ColE9 complex (PDB 1BXI²¹) into an electrostatic interaction. As indicated by an engineered Im9 variant, this mutation can on its own improve ColE7 inhibition *in vivo*

Figure 1 The presumed genealogy of the *in vitro* evolution experiment. The scheme is based on the representative variants analyzed in detail. The order follows the order of rounds (only the fifth, eighth and twelfth rounds of mutation and selection were sampled), the *in vivo* protection levels (for variants within the same round) and the mutational content. Noted are mutations that were fixed during the divergence process; the complete sequences are available in **Supplementary Figure 1**.

by 10-fold to 30-fold (Table 1; detailed parameters for all variants are available in **Supplementary Table 1**). In the native ColE7–Im7 complex structure (PDB 7CEI²²), residue 34 interacts directly with Lys528 of ColE7 (ref. 22). A single interaction is possible in the evolved variants, in which Asp34 interacts with Lys528 and/or Lys537 of ColE7, although the poor electron density observed for the side chain of Asp34 prevented the determination of its correct rotamer.

All other mutations seen in round 5 variants (N24D, D26N, T27A and S28T) are located in the loop connecting $\alpha 1$ to $\alpha 2$, next to but not within the variable region. These residues are not involved in binding the colicins and were not previously implicated in the divergence of binding specificity. None of these residues, with the exception of that in position 26, adopted the same amino acid as that seen in wild-type Im7 (Supplementary Fig. 1). Even Asn26, which has a role in the native ColE7–Im7 interface²², does not seem to contact ColE7 in the nevIm7 variants. On their own, even two of these loop mutations result in a 10-fold to 30-fold increase in ColE7 inhibition (for example, variant R5-4; Table 1). *In vitro* assays confirmed that the higher binding affinities (Supplementary Fig. 2) and dissociation rate constant (k_{off}) measurements with ColE7 indicated >100-fold slower dissociation relative to the starting point Im9 for variant R8-1, which carries an additional mutation (N24D) relative to R5-4 (three loop mutations in total; Supplementary Table 1).

Although not in direct contact with ColE7, the mutations situated in the loop connecting helices $\alpha 1$ to $\alpha 2$ seem to affect the mode of binding of the nevIm7s. The overall structures of Im9, Im7 and

Table 2 Data collection and refinement statistics (molecular replacement)

	ColE7–R12-2	ColE7–R12-13
Data collection		
Space group	$P2_12_12_1$	$P2_1$
Cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	54.2, 67.4, 123.2	59.6, 53.7, 79.3
α , β , γ (°)	90.0, 90.0, 90.0	90.0, 105.6, 90.0
Resolution (Å)	50.0(2.2) *	50.0(2.5) *
R_{sym}	12.0(33.4)	14.3(27.6)
$I / \sigma I$	8.8(1.8)	8.6(2.3)
Completeness (%)	95.3(95.8)	94.2(83.9)
Redundancy	5.9(5.8)	3.4(2.5)
Refinement		
Resolution (Å)	50–2.2	50–2.5
No. reflections	21,366	15,100
$R_{\text{work}} / R_{\text{free}}$	23.4/27.9	25.16/30.11
No. atoms		
Protein	3,262	3,235
Water	10	—
B-factors		
Protein	41.17	41.25
Water	30	—
R.m.s. deviations		
Bond lengths (Å)	0.066	0.019
Bond angles (°)	4.6	1.95

*Values in parentheses are for highest-resolution shell.

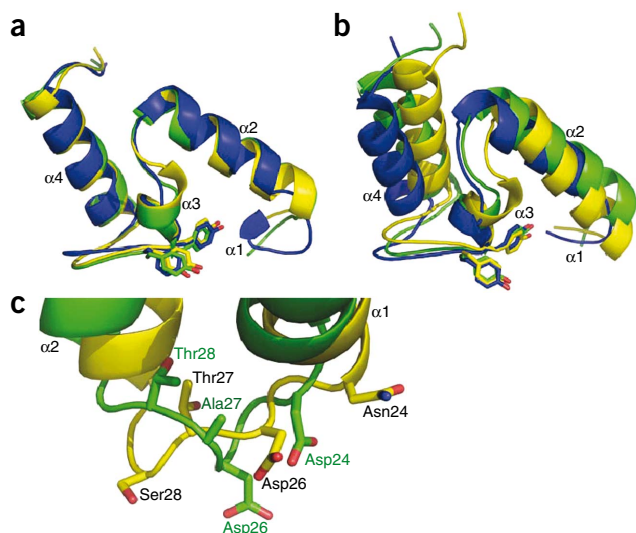


Figure 2 Binding configurations of Im7 and Im9 variants. Shown are backbone alignments of Im9 (yellow), evolved variant R12-2 (green) and Im7 (blue). Helix $\alpha 1$ (residues 1–25) was removed for clarity, and Tyr55–Tyr56, which serve as a common anchoring point¹⁴, are highlighted in sticks. **(a)** Alignment of the three immunities indicate their structural identity (r.m.s. deviation = 0.453 (Im9/R12-2); 0.988 (Im7/R12-2); 1.033 (Im9/Im7)). **(b)** Alignment of the complexed immunity proteins by aligning the bound colicins. Im9 and Im7 differ most substantially in their binding configurations (19°, with Tyr55–Tyr56 as the pivot)¹⁴, and the variant R12-2 is in an intermediate configuration between Im9 and Im7. **(c)** A zoom on the loop connecting helices $\alpha 1$ and $\alpha 2$ in Im9 (yellow, residue annotations in black) and evolved variant R12-2 (green, residue annotations in green). Shown are the deviations in the relative position of helices $\alpha 1$ and $\alpha 2$, in the loop's backbone and in the side chains of residues 24, 26 and 28.

variant R12-2 are essentially identical (**Fig. 2a**). However, they are considerably rotated in relation to the respective colicin partner. The conserved hotspot (positions 55 and 56) can be readily overlaid in all Im variants, native Im7, Im9 and nevIm7s. It acts as a pivot around which the Im proteins rotate relative to the bound colicin¹⁴. This rotation, which is 19° between wild-type Im9 and Im7 (**Fig. 2b**)¹⁴, results in different residues being presented to the colicin partner. The nevIm7 variants adopted an intermediate configuration between that of the Cole9–Im9 and Cole7–Im7 complexes (**Fig. 2b**), resulting in different residues making contact with the Cole7. Although the structured variants contain few mutations beyond the loop mutations, the cause for this shift seems to be the three or four mutations in the loop connecting helices $\alpha 1$ to $\alpha 2$ (**Fig. 2c**), most notably, residue 24. In Im9, this residue is in the C-terminal part of helix $\alpha 1$. But in both variants R12-2 and R12-13, position 24 was mutated from asparagine to aspartate and became part of the loop connecting helices $\alpha 1$ and $\alpha 2$, thus extending the loop by 1 residue. Indeed, in Im7, this loop contains an insertion of an additional residue (alanine, between residues 28 and 29; **Supplementary Fig. 1**).

The change in binding configuration triggered by mutations in the $\alpha 1$ – $\alpha 2$ loop seems to have capitalized on latent Im9–Cole7 interactions that are inaccessible in the original binding configuration of wild-type Im9, specifically, with the acidic residues Glu30 and Asp51 in Im9, and Asp31 and Asp52 in Im7. Glu30 of Im9 makes a salt bridge

to Cole9's Arg54 (**Fig. 3a**), whereas Asp51 has a role only in backbone or water-mediated interactions. In the Im7–Cole7 complex, these residues also make salt bridges: Asp31 of Im7 to Cole7's Arg520 and Lys525, and Asp52 to Arg 530 and Thr531 (**Fig. 3b**). Site-directed mutagenesis of these residues indicated that removal of these interactions has a deleterious effect on formation of the cognate complex of Im7 and Im9 (>10⁴-fold reduction in affinity for Im9 and a significant effect on *in vivo* protection for the Im7–Cole7 complex)^{23,24}. These contacts cannot promote noncognate binding because the configuration of the Cole9–Im9 and Cole7–Im7 complexes differ by 19°. However, owing to its new binding configuration, variant R12-2 seems to make new contacts via Glu30 and Asp51. Similarly, variant R12-2 seems to form new salt bridges via Glu30 with Cole7's Arg520, and via Asp51, which seems to interact with Gln532 and Thr531 of Cole7 (**Fig. 3c**). The Im residues Glu30 and Asp51, which remained unchanged, can also mediate Cole9 binding, thus explaining how variants of the fifth round could gain considerable advantage in inhibiting Cole7 even though their inhibition of Cole9 remained unchanged within the plateau range of the *in vivo* assay (**Table 1**).

Toward higher Cole7 affinity

As the selection progressed, up to four different mutations in the $\alpha 1$ – $\alpha 2$ connecting loop were combined with the V34D mutation situated in $\alpha 2$ (present in all but one of the eight round variants). However, these mutations also led to a severe loss of stability, as indicated by urea denaturation measurements: variants R5-4 and R5-10 showed a gradual decrease in *D*₅₀ value (denaturation mid-point, that is, the concentration of urea leading to 50% denaturation) from 4.9 M

Figure 3 The latent multipotency of Im9. Shown are ribbon presentations of the different complexes. Changes in the binding configuration of the evolved variants led to increased affinity toward Cole7 by re-aligning pre-existing Im9 residues. Shown are residues 30 (31 in Im7) and 51 (52 in Im7) in the complexes of Im9, Im7 and the variant R12-2. Potential salt bridge distances of ≤ 3.5 Å are shown as dashed black lines. **(a)** Glu30 of Im9 (teal) is involved in double salt bridge with Arg54 of Cole9 (orange), whereas Asp51 makes no direct side chain–side chain interactions¹⁴. **(b)** Asp31 of Im7 (blue) is involved in cluster of salt bridge bonds to Arg520 and Lys525 of Cole7 (orange), whereas Asp52 of Im7 is within hydrogen bond distance to Thr531 and Arg530 of Cole7 (ref. 22). **(c)** The side chain of Glu30 in the variant R12-2 (green) is shifted and forms a double salt bridge to Arg520 of Cole7 (orange). Asp51 is within hydrogen bond distance to Gln532 of Cole7. However, the side chains of Lys525 and Arg530, which have a key role in salt bridge contacts with Glu30 and Asp51, respectively, in the structure of the Cole7–Im7 complex (**b**) adopt a different conformation that eliminates these contacts in variant R12-2. The figures depicting the structures were prepared using PyMOL (<http://pymol.sourceforge.net>).

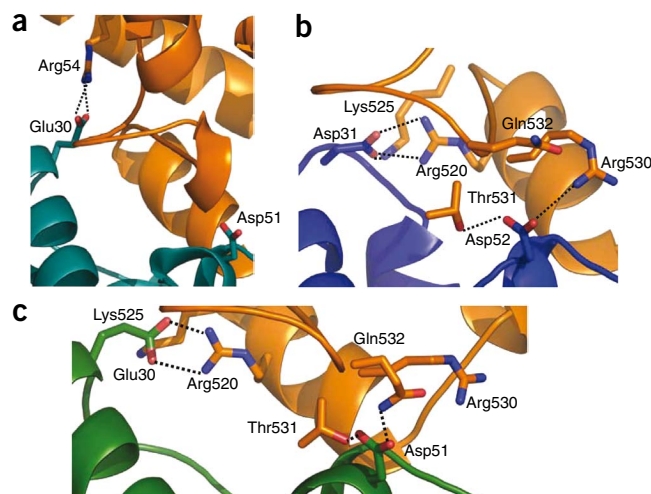
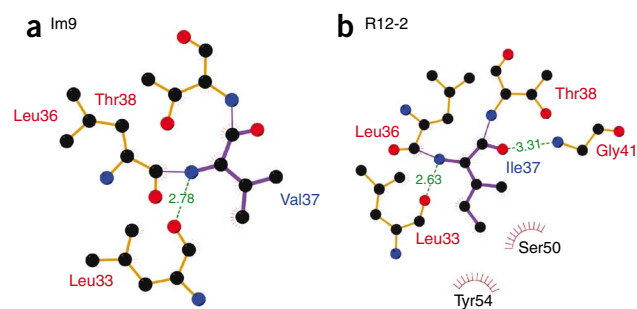


Figure 4 Intramolecular stabilizing interactions by residue 37 in wild-type Im9 (**a**, Val37) and the evolved variant R12-2 (**b**, Ile37). In the evolved variant, Ile37 interacts with two additional residues, Tyr54 and Ser50. Tyr54 is a binding hotspot and has a key role in the interaction with colicins. Adopting the correct orientation in the evolving Im variants might therefore be crucial (**Supplementary Fig. 3**). In variant R12-2, Ile37 also forms a hydrogen bond with Gly41 and might thereby have enabled the appearance of the selectivity mutation E41G. The figure was created with Ligplot³⁸. Hydrogen bonds are presented as dashed green lines, and the interatomic distances are shown in angstroms. Residues that form van der Waals contacts with the Ile37 and Val37 are depicted as labeled arcs with radial spokes that point toward the ligand atoms with which they interact.



for the Im9 starting point to 3.25 M, whereas variants R5-11 and R8-6 showed noncooperative unfolding, and variant R8-1 showed a D_{50} of 2.2 M (**Table 1**). Substantial reduction in stability can reduce the levels of soluble, functional inhibitor and may account for the impaired *in vivo* inhibition levels toward both ColE7 and ColE9 (see variants R5-11, R8-1 and R8-6; **Table 1** and **Supplementary Table 1**). Stability seemed to be regained once the mutation V37I was introduced (for example, variant R8-7 has a D_{50} of 3.7 M). This mutation seemed to improve packing of $\alpha 1$ against $\alpha 3$ and to interact with Tyr54 of the ‘conserved hotspot’ (**Fig. 4**). Its presence led to an increase in ColE7’s *in vivo* protection to levels of 10^{-7} M at low expression and up to 10^{-4} M at high expression levels (**Table 1**, variants R8-7 and R8-8), probably owing to higher levels of soluble Im protein.

The binding parameters of the purified Im variants were measured by stopped-flow kinetics and surface plasmon resonance (SPR) (**Supplementary Tables 1** and **2**). The k_{on} values (measured for two round 8 variants) toward ColE7 increased by ~eight-fold, and the fast binding rates of the nevIm7 evolved variants ($\sim 8 \times 10^8 \text{ M}^{-1} \text{ s}^{-1}$) became markedly similar to wild-type that of Im7. However, most of the increase in ColE7 binding comes, as expected, from decreases in k_{off} , from 5.34 s^{-1} in Im9 to $\sim 10^{-3} \text{ s}^{-1}$ in the best round 12 variants (R12-4, R12-13). Overall, the evolutionary process gave rise to a $> 5 \times 10^4$ increase in affinity, and the K_d of the best variants was $< 10^{-12}$ M (**Table 1**). The affinity measurements correlate with the *in vivo* protection levels, as no further advantage is observed *in vivo* to variants with K_d values of below 10^{-11} (ref. 19). Indeed, the best variants show *in vivo* protection levels that are indistinguishable from those of wild-type Im7 (**Table 1** and **Supplementary Table 1**). Notable exceptions can be seen in relation to stability. For example, variants R8-6 and R8-7 have similar binding parameters (k_{off} values by SPR of 0.064 s^{-1} and 0.067 s^{-1} , respectively), but the *in vivo* protection levels of R8-6 at high expression levels (1 mM IPTG) are substantially lower (3×10^{-7} versus 10^{-4} M ColE7 for R8-7; **Table 1**). This difference probably stems from the limited stability of R8-6, which results in low levels of soluble, functional protein. Indeed, variant R8-6 showed noncooperative urea unfolding. In contrast, variant R8-7, which carries the stabilizing V37I mutation, shows cooperative unfolding, with a D_{50} of 3.7 M.

The only mutation observed in the conserved hotspot is Y55W in variants 4 and 13 of round 12. This mutation is somewhat unexpected as it involves two nonsynonymous substitutions within the same codon (TAC to TGG). However, comparison of the structures of the evolved variants with the native complexes indicates that this mutation had almost no impact on the conserved hotspot (**Supplementary Fig. 3**), as previously reported²⁵.

Acquisition of selectivity

Round 8 variants had comparable affinity toward their original target, ColE9, and their new target, ColE7 (**Table 1**, $K_d \text{ ColE7} / K_d \text{ ColE9} \approx 1$). The dual selection for affinity and selectivity led to the fixation of a

single mutation (E41G) that increased selectivity toward ColE7 and favored E7 binding over E9 ($K_d \text{ ColE7} / K_d \text{ ColE9}$ values of 33 up to 70; **Table 1**). Residue 41 mediates a key Im9 binding interaction with Lys97 of ColE9 but has no role in the Im7–ColE7 complex. Indeed, comparison of the structures of the variant with (R12-2; Gly41) and without (R12-13; Glu41) this mutation showed that it disrupts ColE9 binding with no apparent effect on the binding configuration with ColE7 (**Supplementary Fig. 4**). A substantial effect on binding affinity (34-fold decrease in affinity²⁴), with no effect on binding configuration, was seen when the salt bridge between Lys97 of ColE9 and Glu41 of Im9 was disrupted via the E41A mutation (PDB 1FR2)¹⁴.

The E41G mutation also had other effects related to stability. Its introduction to variant R8-6 (whose stability is already compromised because of the $\alpha 1$ – $\alpha 2$ loop mutations and V34D) reduced both ColE9 and ColE7 *in vivo* inhibition, in particular at 1mM IPTG. This suggests that E41G decreases stability and soluble expression levels, thus resulting in lower protection levels against both colicins. In contrast, introduction of E41G into variant R8-7 and, in particular, to R8-8 reduced ColE9 inhibition with no deleterious effect on ColE7 inhibition (**Table 1** and **Supplementary Table 1**). Unlike variant R8-6, variants R8-7 and R8-8 carry the V37I mutation, which seems to stabilize the G41E mutant (other potentially stabilizing mutations such as K57E and S6R appeared in round 8 and round 12 variants). Indeed, the crystal structures indicate that the evolving Im9 variants carrying Ile37 make two additional intramolecular hydrophobic interactions that seem to stabilize the specificity mutation of E41G and pack against two other crucial Im residues (**Fig. 4**).

DISCUSSION

We describe a transition from the Im9 starting point, involving eight mutations overall, leading to an increase of $\sim 10^5$ -fold in binding affinity and inhibition potency *in vivo* and a $> 10^8$ -fold increase in selectivity (for other recently described divergence trajectories, see refs. 1–6,26,27).

The ‘dual recognition’ mechanism drawn from comparisons of natural members of the colicin-immunity family^{12,24} is the key to this transition. The conserved hotspot (helix $\alpha 3$ residues 54–55) provided the starting point via the promiscuous binding of the noncognate ColE7 and was maintained throughout the process. The new binding specificity was mediated via mutations in the variable region, and in particular in the helix $\alpha 2$ residues 34, 37 and 41. However, the new binding specificity occurred primarily by exploitation of latent interactions, without changing the sequence of the Im-contacting residues (**Figs. 2** and **3**). The first mutations induced changes in noncontacting residues within the loop connecting $\alpha 1$ to $\alpha 2$. These changes (four amino acid exchanges in a loop spanning six residues) seem to have enabled a rotation around the fixed pivot of the conserved hotspot, and thus uncovered a new set of interactions (**Fig. 3**). The observation that early mutations occurred in noncontacting residues is reminiscent

of the 'evolution' of antibody responses, in which affinity maturation is often mediated by mutations in noncontacting residues that fix the conformation of the binding loops²⁸. Similarly, many mutations that induce changes in enzyme specificity occur in second-shell residues and affect the conformation of active site loops²⁹.

The only selection pressure applied up to the eighth round has been for increased ColE7 inhibition. Maintenance of the original ColE9 inhibition in the absence of selective pressure to do so reflects the inherent multispecific, 'generalist' nature of these intermediate variants and the involvement of multiple binding configurations in mediating the promiscuous, evolving function. These features are reminiscent of what has been observed with the evolution of new enzyme functions^{7–9,29} and in receptors^{2,5,29,30}. The diminishment of the original binding configuration that mediates ColE9 binding came after the application of selection against ColE9 inhibition, via the disruption of a key ColE9–Im9 interaction (the E41G mutation). Thus, selectivity and the divergence of a 'specialist' followed a dual selection for the new target and against the old one³⁰.

What is the number of possible mutational solutions for a given evolutionary challenge¹⁰. The nevIm7 variants represent a solution that differs fundamentally from the native Im7, in sequence, in the overall binding configuration and in the contacts made with ColE7. Thus, in the ColE–Im system, at least two nonredundant solutions are available for one challenge. That said, the nevIm7 variants unexpectedly show distinct characteristics of wild-type Im7, including low stability and fast k_{on} rates^{31,32} (Supplementary Table 1). It should also be noted that the nevIm7 variants have an affinity for ColE that is 100-fold weaker than that of wild-type Im7, and they are 10⁶-fold less selective. Whether the nevIm7s can further evolve and match wild-type Im7, and whether completion of the process will result in complete convergence with wild-type Im7, is currently unknown.

As is the case with enzyme evolution, loss of stability resulting from new-function mutations, and compensation via stabilizing mutations, is a major factor^{5,33–37}. The earliest mutations seem to be largely additive: V34D, and several different mutations in the $\alpha 1$ – $\alpha 2$ connecting loop, provide ColE7 protection on their own and increased protection when combined (Table 1). They do, however, lead to a major loss of stability. The specificity-changing mutation E41G further decreases stability, and its emergence seems to be dependent on the presence of the stabilizing mutation V37I (Fig. 4). The introduction of E41G to variants that lack V37I resulted in reduced ColE7 protection levels, probably owing to reduced expression levels (Table 1). Indeed, E41G was not observed without V37I in >20 sequences acquired for variants from rounds 5, 8 and 12, although the opposite was common (for example, variants R8-7 and R8-8). A similar phenomenon dubbed 'conformational epistasis' was observed in the divergence of corticoid receptors—mutations stabilizing the binding site enabled the subsequent accumulation of specificity-switching mutations⁵. It seems, therefore, that mutations that endow new functions are generally destabilizing, regardless of whether the function is enzymatic, ligand binding or protein binding, and stabilizing, compensatory mutations constitute a key step in the evolutionary trajectories of all types of protein functions^{34,35}. Evolution of Im proteins might be particularly restricted because several Im residues that are ColE contacting are also involved in folding intermediates³².

Overall, our results indicate striking similarities between the divergence paths of new functions in enzymes, receptors and protein-protein interactions. These include the exploitation of promiscuous interactions and of alternative binding conformations (or configurations, as is the case here), 'generalist' intermediates capable of efficient binding of both the original and the new target, loss of stability upon gain of a new

function and the key role of compensatory stabilizing mutations. These similarities exist despite the fact that protein-protein binding surfaces are relatively flat (as opposed to deep pockets made of long, flexible loops, as is the case for enzymes, antibodies and many receptors) and are often based on rigid secondary-structure elements such as helices. Thus, the same general principles seem to underline the evolution of different proteins with varied functions and folds²⁹.

METHODS

Methods and any associated references are available in the online version of the paper at <http://www.nature.com/nsmb/>.

Accession codes. Protein Data Bank: Coordinates for ColE7–R12-2 and ColE7–R12-13 have deposited with accession codes 3GKL and 3GJN, respectively.

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

ACKNOWLEDGMENTS

We gratefully acknowledge financial support by the Israel Ministry of Science and Technology, the EU training network ProSA and the Sasson and Marjorie Peress Philanthropic Fund. We are grateful to G. Schreiber, O. Cohavi and M. Harel for their help in affinity measurements. C.K. acknowledges the UK Biotechnology and Biological Sciences Research Council for funding.

AUTHOR CONTRIBUTIONS

K.B.L. performed the selections and variant characterization; O.D. and S.A. crystallized and solved the structures; S.M. participated in the development of the emulsion selections; A.H.K. and C.K. performed stopped-flow measurements; D.S.T. designed the experiments and analyzed the data; K.B.L., O.D. and D.S.T. wrote the manuscript.

Published online at <http://www.nature.com/nsmb/>.

Reprints and permissions information is available online at <http://npg.nature.com/reprintsandpermissions/>.

- Dean, A.M. & Thornton, J.W. Mechanistic approaches to the study of evolution: the functional synthesis. *Nat. Rev. Genet.* **8**, 675–688 (2007).
- Bridgham, J.T., Carroll, S.M. & Thornton, J.W. Evolution of hormone-receptor complexity by molecular exploitation. *Science* **312**, 97–101 (2006).
- Miller, S.P., Lunzer, M. & Dean, A.M. Direct demonstration of an adaptive constraint. *Science* **314**, 458–461 (2006).
- Weinreich, D.M., Delaney, N.F., Depristo, M.A. & Hartl, D.L. Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* **312**, 111–114 (2006).
- Ortlund, E.A., Bridgham, J.T., Redinbo, M.R. & Thornton, J.W. Crystal structure of an ancient protein: evolution by conformational epistasis. *Science* **317**, 1544–1548 (2007).
- Yokoyama, S., Tada, T., Zhang, H. & Britt, L. Elucidation of phenotypic adaptations: molecular analyses of dim-light vision proteins in vertebrates. *Proc. Natl. Acad. Sci. USA* **105**, 13480–13485 (2008).
- Aharoni, A. *et al.* The 'evolvability' of promiscuous protein functions. *Nat. Genet.* **37**, 73–76 (2005).
- Matsumura, I. & Ellington, A.D. *In vitro* evolution of β -glucuronidase into a β -galactosidase proceeds through non-specific intermediates. *J. Mol. Biol.* **305**, 331–339 (2001).
- Khersonsky, O., Roodveldt, C. & Tawfik, D.S. Enzyme promiscuity: evolutionary and mechanistic aspects. *Curr. Opin. Chem. Biol.* **10**, 498–508 (2006).
- Riley, M.A. Molecular mechanisms of colicin evolution. *Mol. Biol. Evol.* **10**, 1380–1395 (1993).
- Riley, M.A. Molecular mechanisms of bacteriocin evolution. *Annu. Rev. Genet.* **32**, 255–278 (1998).
- Kleanthous, C. & Walker, D. Immunity proteins: enzyme inhibitors that avoid the active site. *Trends Biochem. Sci.* **26**, 624–631 (2001).
- Reichmann, D., Rahat, O., Cohen, M., Neuvirth, H. & Schreiber, G. The molecular architecture of protein-protein binding sites. *Curr. Opin. Struct. Biol.* **17**, 67–76 (2007).
- Kühlmann, U.C., Pommer, A.J., Moore, G.R., James, R. & Kleanthous, C. Specificity in protein-protein interactions: the structural basis for dual recognition in endonuclease colicin-immunity protein complexes. *J. Mol. Biol.* **301**, 1163–1178 (2000).
- Bernath, K., Magdassi, S. & Tawfik, D.S. Directed evolution of protein inhibitors of DNA-nucleases by *in vitro* compartmentalization (IVC) and nano-droplet delivery. *J. Mol. Biol.* **345**, 1015–1026 (2005).
- Maté, M.J. & Kleanthous, C. Structure-based analysis of the metal-dependent mechanism of H-N-H endonucleases. *J. Biol. Chem.* **279**, 34763–34769 (2004).

17. Li, W., Dennis, C.A., Moore, G.R., James, R. & Kleanthous, C. Protein-protein interaction specificity of Im9 for the endonuclease toxin colicin E9 defined by homologue-scanning mutagenesis. *J. Biol. Chem.* **272**, 22253–22258 (1997).
18. Wallis, R. *et al.* *In vivo* and *in vitro* characterization of overproduced colicin E9 immunity protein. *Eur. J. Biochem.* **207**, 687–695 (1992).
19. Li, W. *et al.* Highly discriminating protein-protein interaction specificities in the context of a conserved binding energy hotspot. *J. Mol. Biol.* **337**, 743–759 (2004).
20. Keeble, A.H. & Kleanthous, C. The kinetic basis for dual recognition in colicin endonuclease-immunity protein complexes. *J. Mol. Biol.* **352**, 656–671 (2005).
21. Kleanthous, C. *et al.* Structural and mechanistic basis of immunity toward endonuclease colicins. *Nat. Struct. Biol.* **6**, 243–252 (1999).
22. Ko, T.P., Liao, C.C., Ku, W.Y., Chak, K.F. & Yuan, H.S. The crystal structure of the DNase domain of colicin E7 in complex with its inhibitor Im7 protein. *Structure* **7**, 91–102 (1999).
23. Lu, F.M., Yuan, H.S., Hsu, Y.C., Chang, S.J. & Chak, K.F. Hierarchical order of critical residues on the immunity-determining region of the Im7 protein which confer specific immunity to its cognate colicin. *Biochem. Biophys. Res. Commun.* **264**, 69–75 (1999).
24. Wallis, R. *et al.* 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 (1998).
25. Kortemme, T. *et al.* Computational redesign of protein-protein interaction specificity. *Nat. Struct. Mol. Biol.* **11**, 371–379 (2004).
26. Hiniker, A. *et al.* Laboratory evolution of one disulfide isomerase to resemble another. *Proc. Natl. Acad. Sci. USA* **104**, 11670–11675 (2007).
27. O'Maille, P.E. *et al.* Quantitative exploration of the catalytic landscape separating divergent plant sesquiterpene synthases. *Nat. Chem. Biol.* **4**, 617–623 (2008).
28. Foote, J. & Milstein, C. Conformational isomerism and the diversity of antibodies. *Proc. Natl. Acad. Sci. USA* **91**, 10370–10374 (1994).
29. Tokuriki, N. & Tawfik, D.S. Protein dynamism and evolvability. *Science* **324**, 203–207 (2009).
30. Collins, C.H., Leadbetter, J.R. & Arnold, F.H. Dual selection enhances the signaling specificity of a variant of the quorum-sensing transcriptional activator LuxR. *Nat. Biotechnol.* **24**, 708–712 (2006).
31. Ferguson, N., Capaldi, A.P., James, R., Kleanthous, C. & Radford, S.E. Rapid folding with and without populated intermediates in the homologous four-helix proteins Im7 and Im9. *J. Mol. Biol.* **286**, 1597–1608 (1999).
32. Friel, C.T., Smith, D.A., Vendruscolo, M., Gsponer, J. & Radford, S.E. The mechanism of folding of Im7 reveals competition between functional and kinetic evolutionary constraints. *Nat. Struct. Mol. Biol.* **16**, 318–324 (2009).
33. Tokuriki, N., Stricher, F., Serrano, L. & Tawfik, D.S. How protein stability and new functions trade off. *PLoS Comput. Biol.* **4**, e1000002 (2008).
34. Bershtein, S., Goldin, K. & Tawfik, D.S. Intense neutral drifts yield robust and evolvable consensus proteins. *J. Mol. Biol.* **379**, 1029–1044 (2008).
35. Wang, X., Minasov, G. & Shoichet, B.K. Evolution of an antibiotic resistance enzyme constrained by stability and activity trade-offs. *J. Mol. Biol.* **320**, 85–95 (2002).
36. Tomatis, P.E. *et al.* Adaptive protein evolution grants organismal fitness by improving catalysis and flexibility. *Proc. Natl. Acad. Sci. USA* **105**, 20605–20610 (2008).
37. Bloom, J.D., Labthavikul, S.T., Otey, C.R. & Arnold, F.H. Protein stability promotes evolvability. *Proc. Natl. Acad. Sci. USA* **103**, 5869–5874 (2006).
38. Wallace, A.C., Laskowski, R.A. & Thornton, J.M. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng.* **8**, 127–134 (1995).
39. Wallis, R. *et al.* Protein-protein interactions in colicin E9 DNase-immunity protein complexes. 2. Cognate and noncognate interactions that span the millimolar to femtomolar affinity range. *Biochemistry* **34**, 13751–13759 (1995).

ONLINE METHODS

Selections and screens. We performed mutagenesis by error-prone PCR at an average rate of 3.5 mutations per gene, as described¹⁵. We carried out DNA shuffling twice, after the sixth and eleventh rounds of selection. The resulting gene libraries were selected by IVC in emulsions, as described¹⁵. ColE7 was translated *in situ* (at concentrations of $\leq 0.1 \mu\text{M}$)¹⁵, and the ColE9 H103A mutant¹⁶ was added to the cell-free translation mix at $0.7 \mu\text{M}$ concentration. We performed *in vivo* protection screens as described¹⁹. Briefly, we recloned the Im variants isolated from the twelfth round of IVC selection into the pTrc99a vector and expressed them in JM83 bacteria. Cultures were grown without IPTG at 37°C . We plated about 100 bacteria on LB-ampicillin plates and incubated them for 4 h at 37°C . We then poured 3 ml of 0.7% top agar, mixed with ampicillin and various ColE7–Im7 complexes, on top of the plates for overnight incubation at 37°C . Surviving colonies were picked for further analysis. The screen was repeated four times and the best 34 variants were analyzed further.

For determining the levels of *in vivo* protection (Table 1 and Supplementary Table 1), we cloned the newly evolved and wild-type Im variants into the IPTG-inducible expression plasmid pTrc99a (Pharmacia Biotech, Germany) and used them to transform *E. coli* JM83 cells. Cells were grown as lawns on LB-ampicillin agar plates without or with IPTG (0.05 mM or 1 mM) and spotted with ColE7 at different concentrations. After an overnight incubation we assessed cell death, as visualized in the form of plaques, and recorded the lowest concentration of ColE7 at which there was no plaque formation.

Purification of colicin DNases and Im variants. We carried out purification largely as described¹⁹. We cloned the Im protein variants into the pTrc99a vector and used it to transform *E. coli* JM83 bacteria. Immunity proteins expressed with a streptomycin (Strep) tag at the N terminus (denoted with 's' preceding the variant's annotation) were cloned in pET20 (NovaGen) and transformed into BL21(DE3). We performed purification by ion-exchange chromatography (HP-SP, GE healthcare) regardless of the Strep tag. We produced ColE proteins in several versions: for binding assays, as an inactive H103A mutant with a His₆ tag, or as an active nuclease coexpressed with the His₆-tagged cognate Im, and eluted from Ni-NTA by 6 M guanidinium-HCl^{19,25}. For *in vivo* toxicity assays, we purified the ColE–Im–His₆ construct¹⁹. SDS-PAGE analysis showed indicated that proteins were $\geq 90\%$ pure. Stability was determined by urea denaturation³¹.

Kinetic parameters. We performed stopped-flow and SPR using BIAcore 3000 measurements, as described¹⁹. ColE7 or ColE9 was attached to CM5 chips by amine coupling. We injected Im proteins at $30 \mu\text{l min}^{-1}$ flow at 100–800 nM concentrations in 50 mM MOPS, pH 7.5, 200 mM NaCl and 0.005% P20. The dissociation curves were fitted to single exponential curves with Kaleidograph (Synergy Software) (Supplementary Fig. 5 and Supplementary Table 2).

Structures. We obtained crystals of the ColE7–R12-2 complex at 19°C by the microbatch method under oil by using the Oryx6 robot (Douglas Instruments). The protein concentration used for crystallization was 32 mg ml^{-1} , and crystals were grown from a precipitating solution of 100 mM CHES, pH 9, and 30% (w/v) PEG 400. Crystals formed in space group $P2_12_12_1$ and contain two monomers in the asymmetric unit cell, with a V_m of $2.42 \text{ \AA}^3 \text{ Da}^{-1}$. Diffraction data to 2.2 \AA , from a single crystal, were collected at 100 K at the European Synchrotron Radiation Facility (ESRF) beamline ID23-2 at wavelength 0.873 \AA . We also obtained crystals of the ColE7–R12-13 complex at 19°C by the microbatch method under oil. The protein concentration used for crystallization was 40 mg ml^{-1} , and the crystals were grown from a precipitating solution of 100 mM CHES, pH 10.5, and 30% (w/v) PEG 400. Crystals formed in space group $P2_1$ and contain two monomers in the asymmetric unit cell, with a V_m of $2.37 \text{ \AA}^3 \text{ Da}^{-1}$. A complete data set up to 2.5 \AA was collected at 1.5418 \AA on a Rigaku R-AXIS IV++ imaging plate area detector mounted on a Rigaku RU-H3R generator with $\text{CuK}\alpha$ radiation focused by Osmic confocal mirrors. The diffraction data were indexed and integrated using the program HKL2000 (ref. 40). Integrated intensities were scaled using the program SCALEPACK⁴⁰. The structure factor amplitudes were calculated using TRUNCATE from the CCP4 program suite⁴¹. The refinement was carried out using the program, CCP4/Refmac5 (ref. 42). The model was built the electron density maps ($2F_o - F_c$ and $F_o - F_c$) using the program COOT⁴³. Water molecules were built into peaks greater than 3σ in the $F_o - F_c$ maps. The details of the refinement statistics of the ColE7–R12-2 and ColE7–R12-13 structures are presented in Table 2. All figures depicting structures were prepared using PyMol (DeLano Scientific). Finally, the ColE7–R12-2 and ColE7–R12-13 models were evaluated with the program MolProbity⁴⁴. Percentages of residues in the most favored, additionally allowed, generously allowed and disallowed regions of a Ramachandran plot are, respectively, 85.8%, 13.1%, 1.2% and 0.0% (ColE7–R12-2 complex) and 87.1%, 11.2%, 0.9% and 0.9% (ColE7–R12-13 complex).

40. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
41. French, S. & Wilson, K. On the treatment of negative intensity observations. *Acta Crystallogr. A* **34**, 517–525 (1978).
42. Murshudov, G.N., Vagin, A.A. & Dodson, E.J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240–255 (1997).
43. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
44. Davis, I.W. *et al.* MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**, W375–W383 (2007).