Positive selection and recombination: major molecular mechanisms in colicin diversification

Ying Tan and Margaret A. Riley

Colicins, a family of antimicrobial proteins produced by *Escherichia coli*, are one of the best characterized microbial systems for studying processes of molecular diversification. Recent studies employing DNA sequence comparisons and experimental evolution suggest that positive selection and recombination play dominant roles in colicin diversification. Recombination between distantly related colicins has repeatedly generated novel classes of colicins, while positive selection for novel colicin immunity systems produces further diversity among closely related colicins. Together, these forces have resulted in a surprisingly large and diverse class of antimicrobials. Colicins are thought to play an important role in the invasion of bacteria into novel habitats.

Ying Tan is at the Human Genetics Center, School of Public Health, University of Texas, PO Box 20334, Houston, TX 77225, USA; Margaret Riley is at the Dept of Ecology and Evolutionary Biology, Yale University, New Haven, CT 06520-8104, USA (riley@beagle.biology.yale.edu).

Producing antimicrobial compounds seems to be a generic phenomenon for most, if not all, bacteria[1]. Antimicrobials include toxins, bacteriolytic enzymes, bacteriophages, by-products of primary metabolic pathways, antibiotics and bacteriocins. One class of antimicrobials, bacteriocins, has received increasing attention because of the high levels of bacteriocin diversity observed, the widespread distribution of bacteriocins in bacteria, and the potential (and actual) use of bacteriocins in the food industry (as preservatives) and in the human health industry (as antibiotics)[1,2].

Bacteriocins are compounds produced by bacteria that inhibit or kill closely related species[1]. Their production occurs across all major groups of Eubacteria and the Archaebacteria[2,3]. Although the natural role of bacteriocins has not been determined, they appear to be involved in microbial invasion and/or defense.

Colicins constitute one group of extensively studied bacteriocins. They have served as a model system with which to explore the molecular mechanisms of bacteriocin diversification. Roughly 30% of naturally occurring *E. coli* have colicins, encoded on plasmids, and over 20 distinct types of colicins have been characterized[4–7]. Under conditions of stress, a small fraction of colicinogenic bacteria are induced to produce colicin proteins. Induction results in the rapid release of colicin into the environment. Colicin proteins recognize specific cell-surface receptors and are transported into neighboring cells[8–10].

Having gained access, colicins kill cells by one of the following four mechanisms: (1) formation of ion-permeable channels in the cytoplasmic membrane; (2) non-specific degradation of cellular DNA; (3) inhibition of protein synthesis through the specific cleavage of 16s rRNA; and (4) cell lysis resulting from inhibition of peptidoglycan synthesis[7,11].

Colicinogenic cells produce an immunity protein that provides protection against their own colicin. Immunity protein, which is constitutively expressed, recognizes its own colicin, binds to the C-terminal end, and inhibits killing. Only cells carrying the same colicin plasmid survive under conditions of colicin production.

Colicins are divided into two major classes based on function, the nuclease colicins (including DNase- and RNase-type E colicins) and the pore-forming colicins[11–13]. DNA and protein sequence comparisons suggest that the nuclease colicins represent a very closely related lineage of toxins, even though the lineage comprises both RNase and DNase killing functions. In contrast, the pore formers, although united in killing function, comprise a more heterogeneous group.

Here, we review recent progress in our understanding of the molecular mechanisms involved in the diversification of colicins. We focus on two mechanisms of diversification: positive selection and recombination. Recent data suggest that these are the major forces involved in the diversification of colicins. More interestingly, the same data suggest that the pore-forming and nuclease colicins evolve in dramatically different fashions.

Positive selection: a major force in the diversification of nuclease colicins

Positive selection has been invoked to explain the high level of nuclease colicin diversity[13,14]. The form of selection envisioned is based on an 'advantage when rare' scenario that involves two steps. First, a point mutation occurs in the immunity gene of a nuclease colicin that confers a broadened immunity function. Such a cell is immune to self, to its immediate ancestor and, perhaps, to several other colicins (Fig. 1a). This evolved immunity function will have an obvious advantage over the ancestral state under colicinogenic conditions and will be selectively retained in the population. While retained, a second, paired, mutation in the colicin gene results in a colicin (called a 'super-killer') to which its ancestor is no longer immune (Fig. 1b).

The 'super-killer' colicin will have an immediate and large selective advantage. Positive selection will drive cells encoding the 'super-killer' colicin rapidly into the population, until some new 'super-killer' phenotype evolves from it in yet another round of diversification.

At the molecular level, repeated rounds of this form of diversification of immunity function would result in the accumulation of nonsynonymous substitutions in the immunity region of diverging colicin gene clusters. Sequence comparisons of two closely related pairs of nuclease colicin gene clusters, E2/E9 DNase colicins and E3/E6 RNase colicins, show exactly this pattern of divergence (Fig. 2).

Between rounds of immunity region diversification, neutral substitutions will accumulate randomly across the

  0169-5347/97/$17.00 PII: S0169-5347(97)01127-0

plasmid replicon. Recombination between the evolved and ancestral colicins will release the neutral polymorphisms from their linkage with the selected sites and thus homogenize the evolved and ancestral colicin plasmids. However, recombination between the selected sites in the evolved colicin is lethal if the evolved colicin gene is paired with the ancestral immunity gene. Thus, a mutational trap is produced that would result in the accumulation of neutral mutations in the immunity region[13,14]. Again, such a pattern is observed between closely related nuclease colicins (Fig. 2).

One line of evidence supporting the action of positive selection in colicin diversification comes from experimental studies. The diversifying selection hypothesis posits the two following steps: (1) coupled mutations that provide the host cell with an expanded immunity and 'super-killer' function, and (2) rapid invasion of this novel strain.

Studies on immunity specificity determinants of both RNase- and DNase-type colicins suggest a mechanism for the proposed expanded immunity. Several studies have shown that the specificity of colicin–immunity interactions is determined by a very few residues[15,16]. Furthermore, one can generate a broadened immunity function via a single amino acid substitution[17].

The process of invasion of the novel colicin into the ancestral population has also been examined experimentally[18]. Several colicin-encoding plasmids naturally possess an additional immunity gene, which expands the hosts' immunity and killing function in a fashion analogous to that described above (Fig. 3a). Invasion experiments reveal that when these 'super-killer' strains compete against their ancestors, they rapidly displace the ancestral strain, even when initial frequencies of the invader are quite low (Fig. 3b)[18].

The process of sequence homogenization between evolved and ancestral forms (between rounds of diversification) is more difficult to address experimentally, owing to very low rates of recombination. However, the finding of a new type of colicin E2 plasmid may shed some light on this process[19].

A survey of colicin E2 polymorphism revealed two types of E2 plasmids[19]. One type was the classical E2 plasmid. The second type was a mosaic plasmid, apparently repeatedly derived by recombination between colicins E2 and E7. The recombinants had E2 immunity regions in otherwise E7 colicin plasmids. The existence of intact E2 immunity regions suggest that the specific interaction between the immunity protein and the immunity binding domain of a colicin may select against recombination events within the immunity region in order to maintain immunity function[11].

A further line of evidence for positive selection acting in colicin diversification comes from DNA sequence comparisons. Patterns of DNA polymorphism within species contain information about the evolutionary process that are not revealed from sequence divergence patterns (between species) alone. The combination of inter- and intraspecific comparisons provides a powerful tool for investigating the evolutionary histories of DNA sequences[20–22]. In particular, if sequences are evolving in a neutral fashion, the levels of inter- versus intraspecific divergence should be correlated. Regions evolving rapidly between species should accumulate polymorphisms rapidly within species[21,23].

Patterns of DNA sequence polymorphism were examined for 14 colicin E2 gene clusters obtained from natural isolates of *E. coli*[19]. These data were compared to the pattern of DNA sequence divergence between colicin E2 and its close relative, E9 (Fig. 2). The patterns of polymorphism and divergence were not positively correlated. Statistical tests, including the HKA and MK tests[21,23], which assess the correlation between levels of polymorphism and divergence, confirm a significant departure from neutral predictions[19].

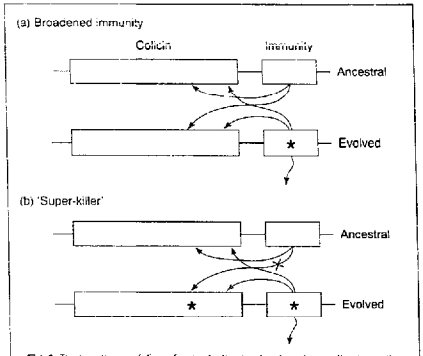

Fig. 1. The hypothesis of diversifying selection invokes two steps in the generation of a novel immunity function. (a) A point mutation in the immunity gene generates a broadened immunity function (noted with an asterisk). The strain with this colicin gene cluster is immune to self, to its ancestor, and to other closely related colicins. The ancestral colicin is immune to self and to the evolved colicin. Immunity is indicated with arrows. (b) A paired mutation occurs in the immunity binding portion of the evolved colicin gene which generates a 'super-killer' (noted with a second asterisk). The evolved strain is still immune to self, its ancestor and other colicins. However, the ancestral strain is now no longer immune to the evolved strain.

Multiple recombination events across the immunity region (with unrelated colicins serving as the donor DNA) could result in the observed patterns of substitution. However, during the time in which multiple recombination events take place, neutral substitutions in the flanking regions should accumulate at an even faster rate, unless conjugation and recombination rates are many orders of magnitude higher in natural populations than suggested[24]. Furthermore, the

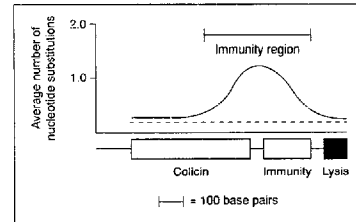

Fig. 2. DNA sequence comparisons between closely related colicin types (or species) reveal a highly clustered pattern of divergence that differs dramatically from the pattern of polymorphism observed within colicin types. In this example, a comparison between colicins E2 and E9, most of the divergence between colicin types (solid line) occurs in the immunity region (comprised of the immunity gene and the immunity binding region of the colicin gene). The same pattern is obtained for total (shown), nonsynonymous and synonymous substitutions. Polymorphism segregating within the colicin E2 population (dashed line) is distributed evenly along the gene cluster. The same pattern is obtained for total (shown), nonsynonymous and synonymous polymorphism.

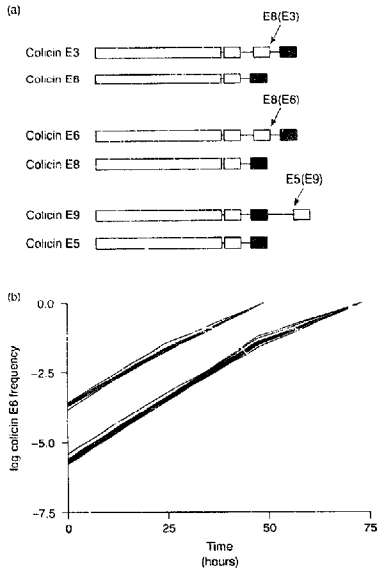

Fig. 3. Methods of experimental evolution were used to determine the invasion dynamics of colicin strains that produce novel immunity functions. (a) Three colicins naturally have extra immunity genes (colicins E3, E6 and E9). In each case, in addition to encoding a colicin (shaded), an immunity (unshaded) and a lysis (black) gene, a second immunity gene (unshaded with superscript label) is found. These dual immunity strains ('super-killers') were competed against their single immunity counterparts (termed their ancestors); E8 in the case of colicins E3 and E6; E5 in the case of colicin E9. (b) The invasion dynamics of one pair (colicins E6 and E8) illustrates the general point that, even with very low initial frequencies, 'super-killer' strains rapidly invade and displace the ancestral strain. The initial frequencies of 'super-killer' strains are $\sim 10^{-4}$ and $\sim 10^{-6}$ with 6 replicates each. The *y*-axis represents the common log of frequency for the dual immunity colicinogenic strain. The *x*-axis represents the number of transfers (24 h = 1 transfer).

availability of an appropriate, unrelated, recombination template is problematic given that the frequency of a specific colicin type is usually less than 1% in natural *E. coli* populations[4,5]. Finally, the specific interaction required between the immunity binding domain of the colicin and immunity protein would prevent recombination events within the immunity region that break up proper protein interactions. Thus, although the diversifying recombination hypothesis cannot be rejected, features of colicin biology make it a less likely explanation.

Recombination drives the diversification of pore-forming colicins

In sharp contrast with the nuclease colicins, recombination is likely the dominant force in the diversification of pore-forming colicins. The most remarkable finding in the evolution of the pore-forming colicins is that regions of colicin sequence, encoding rather precise functional domains, have recombined to give rise to entirely new classes of colicins. Such recombination events occur both between and within the two major groups of pore-forming colicins.

Colicin B is frequently cited as an example of the power of multiple recombination events in colicin diversification[25]. Colicin B is composed of three segments from three origins: the colicin E1-like upstream region; the D-like N-terminal and central region; and the A-like C-terminal region.

Further support for recombinational diversification of pore-forming colicins comes from the protein sequence of colicin 10, which has a mosaic structure (Fig. 4a)[26,27]. It is similar to E1 in the two domains that determine uptake and killing activity but differs entirely from E1 in the remaining two domains, which determine the Tcn/Tol dependence and receptor binding.

Sequence comparison among colicins K, 5 and 10 reveal that these colicins were also assembled via recombination (Fig. 4b). The N-terminal half of colicin K is derived from a source different to that of colicins 5 and 10 while the C-terminal half of the colicin and the entire immunity and lysis proteins of colicin 10 must have an origin other than that of the colicin K and 5 gene clusters.

Recombination seems to have occurred between chromosomally encoded bacteriocins and colicins, as well. *Serratia marcescens* produces a chromosomally encoded bacteriocin, N28b, which shares a high degree of similarity in the first 45 residues with colicin A, with 35 identical residues[28]. Moreover, alignment of the C-terminal amino acid sequence of N28b to the pore-forming domains of several colicins also demonstrates a considerable degree of similarity distributed throughout the sequence[28]. The hydrophobicity profile of the C-terminal region of bacteriocin N28b shows a high degree of similarity to the corresponding domain of colicin A, and is characterized by a major hydrophobic segment of 44 residues. Based upon levels of sequence similarity, 28b appears to be an evolutionary intermediate between the A and E1 groups of pore-forming colicins[28].

Recombination between functional domains may also explain an unusual pattern of sequence divergence between colicins Ia and Ib. These are closely related colicins which show a high degree of substitution clustered in the immunity region, similar to that seen in several nuclease colicins (see Fig. 2)[13,14]. Studies of DNA sequence polymorphism for six colicin Ia plasmids revealed that patterns of divergence between colicins Ia and Ib and patterns of polymorphism within Ia were significantly different, with an elevated level of divergence in the immunity region and an even distribution of DNA polymorphism across the entire colicin gene cluster (again, just like the nuclease colicins)[29]. This result contrasts with neutral predictions, which suggest a positive correlation between patterns of divergence and polymorphism[30]. Unlike the case of nuclease colicins, in which multiple recombination events are required to account for the observed patterns of clustered divergence, a single recombination event between either Ia or Ib and some highly divergent colicin could produce the observed divergence pattern.

After observing so many examples of recombination among pore-forming colicins, one may ask why this mechanism seems more prominent in the diversification of pore-forming than in nuclease colicins. One clue may be that in most cases, pore-forming colicins are the most frequently encountered colicins in natural populations of *E. coli*. Colicin surveys reveal that 73–84% of colicin producing strains produce pore-forming colicins[4,5,12]. It is tempting to speculate that the high frequency at which pore-forming colicins are encountered in nature facilitates frequent recombination events. Furthermore, in the one case in which nuclease

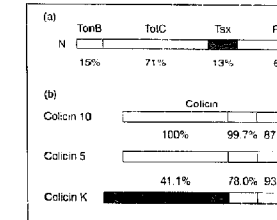

Fig. 4. Recombination between functional domains of pore-forming colicins gives rise to new colicins. (a) A map of the percentage identity between amino acid sequences of colicin 10 and E1 reveals regions of high levels of similarity interspersed with short stretches of reduced similarity. These data suggest a recombination origin of colicin 10 that includes several functional domains from colicin E1. *Reproduced, with permission, from Ref. 30.* (b) Similarities in the nucleotide sequences of colicin, immunity and lysis genes of colicins 5, 10 and K gene clusters suggest shared ancestry. The fragmented nature of the patterns of sequence similarity suggests that multiple recombination events have occurred in the evolution of these colicins. The percentages indicate degree of similarity between colicins 10 or K relative to 5. *Reproduced, with permission, from Ref. 31.*

colicins are encountered in high numbers, an instance of recombination between two nuclease colicins is observed[19].

Conclusions

Colicins represent a diverse group of antimicrobials. Patterns of colicin diversification suggest that two evolutionary forces, recombination and positive selection, are likely to play dominant roles in the origin and evolution of new colicins. In the case of the more abundant, and more heterogeneous, pore-forming colicins, recombination between functional domains of unrelated, or highly divergent, colicins can result in the production of entirely new colicin classes. In contrast, in the case of the rare, and more closely related, nuclease colicins, selection for novel immunity functions appears to account for the observed patterns of diversification. Furthermore, a very limited amount of data suggest that as the nuclease colicins increase in frequency, they will also begin to diversify via recombination.

Producing antimicrobial compounds seems to be a generic phenomenon for most if not all bacteria[1]. Although this review focuses on the molecular mechanisms underlying the diversification of colicins, which is only one class of bacteriocins, colicins may provide a useful framework for investigating similar questions regarding other classes of bacteriocins, and even more generally, other classes of antimicrobial compounds.

Acknowledgements

The preparation of this review was supported by an NIH FIRST award, an NSF Young Investigator Award and by a grant from the General Reinsurance Corporation to MAR. YT was affiliated to Yale University during the preparation of this article.

References

1 Tagg, J.R., Dajani, A.S. and Wannamaker, L.W. (1976) **Bacteriocins of gram-positive bacteria,** *Bacteriol. Rev.* 40, 722–756
2 Dykes, G.A. (1995) **Bacteriocins: ecological and evolutionary significance,** *Trends Ecol. Evol.* 10, 186–189
3 Torreblanca, M., Meseguer, I. and Ventosa, A. (1994) **Production of halocin is a practically universal feature of archeal halophilic rods,** *Appl. Microbiol. Lett.* 19, 201–205
4 Achtman, M. *et al.* (1983) **Six widespread bacterial clones among *Escherichia coli* K1 isolates,** *Infect. Immun.* 39, 315–335
5 Riley, M.A. and Gordon, D.M. (1992) **A survey of col plasmids in natural isolates of *Escherichia coli* and an investigation into the stability of col-plasmid lineages,** *J. Gen. Microbiol.* 138, 1345–1352
6 Pugsley, A.P. (1984) **The ins and outs of colicins,** *Microbiol. Sci.* 1, 168–175, 203–205
7 Riley, M.A. and Gordon, D.M. (1996) **The ecology and evolution of bacteriocins,** *J. Indust. Microbiol.* 17, 151–158
8 Benedetti, H. *et al.* (1989) **Comparison of the uptake systems for the entry of various BtuB group colicins into *Escherichia coli*,** *J. Gen. Microbiol.* 135, 3412–3420
9 Lazdunski, C. (1995) **Colicin import and pore formation: a system for studying protein transport across membrane?** *Mol. Microbiol.* 16, 1059–1066
10 Braun, V. (1995) **Energy-coupled transportation and signal transduction through the gram-negative outer membrane via TonB-ExbB-ExbD-dependent receptor proteins,** *FEMS Microbiol. Rev.* 16, 295–307
11 Konisky, J. (1982) **Colicins and other bacteriocins with established modes of action,** *Annu. Rev. Microbiol.* 36, 125–144
12 Pugsley, A.P. and Oudega, B. (1987) **Methods for studying colicins and their plasmids in** *Plasmids: a Practical Approach* (Hardy, K.G., ed.), pp. 105–161, IRL Press
13 Riley, M.A. (1993) **Molecular Mechanisms of colicin evolution,** *Mol. Biol. Evol.* 10, 1380–1395
14 Riley, M.A. (1993) **Positive selection for colicin diversity in bacteria,** *Mol. Biol. Evol.* 10, 1048–1059
15 Akutsu, A., Masaki, H. and Ohta, T. (1989) **Molecular structure and immunity specificity of colicin E6, an evolutionary intermediate between E-group colicins and cloacin DF13,** *J. Bacteriol.* 171, 6430–6436
16 Curtis, M.D. and 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
17 Masaki, H. *et al.* (1991) **Identification of a unique specificity determinant of the colicin E3 immunity protein,** *Gene* 107, 133–138
18 Tan, Y. and Riley, M.A. (1996) **Rapid invasion of colicinogenic bacteria with novel immunity functions,** *Microbiol.* 142, 2175–2180
19 Tan, Y. and Riley, M.A. **Nucleotide polymorphism in colicin E2 gene clusters: evidence for non neutral evolution,** *Mol. Biol. Evol.* (in press)
20 Kreitman, M. (1983) **Nucleotide polymorphism at the alcohol dehydrogenase locus of *Drosophila melanogaster*,** *Nature* 304, 412–417
21 Hudson, R., Kreitman, M. and Aguade, M. (1987) **A test of neutral molecular evolution based upon nucleotide data,** *Genetics* 116, 153–159
22 Stoltzfus, A., Leslie, J.F. and Milkman, R. (1988) **Molecular evolution of the *Escherichia coli* chromosome. Analysis of structure and natural variation in a previously uncharacterized region between *trp* and *tonB*,** *Genetics* 120, 345–358
23 McDonald, J.H. and Kreitman, M. (1991) **Adaptive protein evolution at the *adh* locus in *Drosophila*,** *Nature* 351, 652–654
24 Gordon, D.M. (1992) **The rate of plasmid transfer among *Escherichia coli* strains,** *J. Gen. Microbiol.* 138, 17–21
25 Roos, U., Harkness, R.E. and Braun, V. (1989) **Assembly of colicin genes from a few DNA fragments: Nucleotide sequence of colicin D,** *Mol. Microbiol.* 3, 891–902
26 Pilsl, H. and Braun, V. (1995) **Novel colicin 10: assignment of four domains to TonB- and TolC-dependent uptake via the Tsx receptor and to pore formation,** *Mol. Microbiol.* 16, 57–67
27 Pilsl, H. and Braun, V. (1995) **Strong function-related homology between the pore-forming colicins K and 5,** *J. Bacteriol.* 177, 6973–6977
28 Vicjo, M.B. *et al.* (1992) **Cloning and DNA sequence analysis of a bacteriocin gene of *Serratia marcescens*,** *J. Gen. Microbiol.* 138, 1737–1743
29 Riley, M.A., Tan, Y. and Wang, J. (1994) **Nucleotide polymorphism in colicin E1 and Ia plasmids from natural isolates of *E. coli*,** *Proc. Natl. Acad. Sci. U.S.A.* 91, 11276–11280
30 Kimura, M. (1983) *The Neutral Theory of Molecular Evolution*, Cambridge University Press