J Mol Evol (2002) 55:347–355
DOI: 10.1007/s00239-002-2331-3

JOURNAL OF MOLECULAR EVOLUTION

When Comparative Information Leads Us Astray: The Receptor-Binding Region of Colicin E9

Robert L. Dorit,[1,2] Margaret A. Riley[2]

[1] Department of Biology, Sabin-Reed Hall, Smith College, Northampton, MA 01060, USA
[2] Department of Ecology and Evolutionary Biology, Yale University, 165 Prospect Street, New Haven, CT 06511, USA

Received: 5 December 2001 / Accepted: 4 February 2002

Abstract. In an effort to develop derivatives of the *Escherichia coli* antimicrobial protein colicin E9 that exhibit novel interactions with a target cell, we mutagenized a 10-amino acid region located at the C terminus of the colicin receptor-binding domain. We subsequently selected for those colicin molecules that retain the antimicrobial phenotype and found that, despite a mutagenic strategy that alters every amino acid in the targeted domain, more than 70% of the engineered colicins retained antimicrobial activity. This result is all the more surprising given the extensive phylogenetic conservation of this receptor-binding domain, which originally suggested the operation of strong selective constraints on the amino acid sequence of this region. This apparent contradiction between our experimental results and the comparative data is resolved by exploring the fitness consequences of the experimentally induced amino acid substitutions. In 17 of 52 cases we examined, the fitness of cells harboring the functional engineered colicins was lower than that of our control line (containing wild-type colicin E9), and in 33 of 52 cases, equal to it. Paradoxically, two of the engineered colicins appear to confer a higher fitness to the producer cell lines. While the mechanism linking changes in the amino acid sequence of the colicin receptor-binding domain and the growth rate of the cells remains unclear, these results illustrate the surprising versatility of the colicin/receptor interaction and underscore the importance of distinguishing molecular function from organismal fitness.

Key words: Antibiotic resistance — Colicin — Antimicrobial — Bacteriocin — In vitro design — Protein engineering — ColE9 — Drug design

Introduction

The power of the comparative approach in biology depends in part on our ability to infer function from the observed pattern. Phylogenetically invariant regions of a molecule are generally assumed to result from the operation of a strong selective constraint on the sequence in that region. This constraint, in turn, is seen to reflect the relatively tight sequence requirements underlying a specific molecular function.

While there are numerous examples that confirm the link between sequence invariance and functional constraint (e.g., Reizer and Saier 1997; Yarus and Welch 2000; Yoshida et al. 2000; Owen and Zelent 2000; Irving et al. 2000; Luciano and Karlin 2000), it is not a logically necessary link. Because selection operates on the overall fitness of the organism and not on the kinetics of individual metabolic reactions, alternative functional protein or nucleic acid sequences may exist yet still not be seen in natural populations because they result in lower organismal fitness. Phylogenetic invariance may also reflect the

Correspondence to: Robert L. Dorit; *email:* rdorit@smith.edu

operation of historical constraints that render certain parts of sequence space inaccessible, despite the existence of alternative functional sequences. Finally, invariance can arise from the action of specific genetic events, including gene conversion and horizontal transfer, which can result in the elimination of alternative, fully functional sequences.

The various mechanisms underlying sequence conservation can be distinguished only by experimentally testing for the presence of functional alternatives in sequence space. We report here on an explicit test of the link between low levels of naturally occurring sequence variation and molecular function, using colicin E9, an antimicrobial protein produced by *Escherichia coli*, as a model system. This investigation is part of a larger study undertaken to explore the interaction between colicin E9 and the cell surface receptor (BtuB) required for this toxin to gain entry into the target cell.

We use methods of experimental evolution to probe the limits of the colicin–receptor interaction and examine some of the selective constraints guiding the evolution of the colicin molecule. Colicin E9 is a member of a large family of bacteriocins produced by *E. coli* (James et al. 1991; Riley 1998; Chak and James 1986). Colicin proteins play a crucial role in bacterial invasion and community dynamics, mediating both inter- and intraspecific competition (Riley 1998; Riley and Gordon 1995). They are synthesized in response to cues from the SOS system; when induced, producer cells rapidly release colicin into the environment. The colicin proteins collide with target cells, gain entry through specific cell surface receptors, and kill the cell by one of several distinct mechanisms (James et al. 1996). Colicin E9, the focus of this study, is an endonuclease that kills target cells by randomly cleaving their chromosomal DNA (Pugsley and Oudega 1987). The colicin protein enters the target cell by attaching to the BtuB cell surface receptor; it is subsequently translocated into the cytoplasm via the *tol* translocation system (James et al. 1996). The induction of colicin production also results in the coordinated expression of two additional genes located immediately downstream: an immunity gene (*ceiI*), encoding a protein that binds to the colicin protein and confers specific immunity to that toxin; and a lysis gene (*lys*), encoding a protein involved in the release of colicin from the cell (James et al. 1991, 1996).

The primary functions of colicin E9—receptor binding (R), translocation (T), and DNase degradation (C)—map in a surprisingly modular fashion onto both the sequence and the three-dimensional structure of the protein (Fig. 1) (Eaton and James 1989). This modularity allows for the independent exploration and manipulation of discrete aspects of colicin function. Previous work using site-directed mutagenesis and truncation studies clearly defined the R domain, involved in the interaction between colicin E9 and the BtuB cell surface receptor (James et al. 1996; Penfold et al. 2000; Chak and James 1986). This region has been tentatively defined as spanning amino acids (AA) 160 to 448. Within the R region, a subdomain of approximately 80 AA (363–448) has been directly implicated by truncation studies in receptor interaction: constructs lacking this subdomain do not enter the host cell (Penfold et al. 2000, personal communication).

Our approach, broadly defined, explores the sequence space surrounding the existing sequence of the receptor-binding domain of the colicin E9 protein. In so doing, we probe the nature of the protein–protein interaction between the receptor–binding domain of colicin E9 and the BtuB cell surface receptor in *E. coli*. Specifically, we are interested in probing for the existence of alternative sequences—and, by extension, of alternative structures—for the R domain of colicin E9. We report here on the outcome of experiments focused on AA 423 to 432, a region located at the C-terminal end of the R domain. This region represents the entry point for an ongoing investigation of sequence versatility in the entire colicin E9 receptor-binding domain. For this initial region, we provide molecular and phenotypic characterization of a large family of engineered colicins generated by introducing significant changes in the AA sequence of the otherwise highly conserved receptor-binding domain of the E9 protein. The existence of such a large variety of fully functional alternative sequences for this domain underscores the need for caution when drawing functional inferences from observed intra- and interspecific sequence conservation.

Materials and Methods

Strains. The host producer strain, *E. coli* JM83 [F–Δ(*lac–pro*AB) phi80 Δ(*lac*Z)M15 *ara rps*L *thi* λ–], was obtained from the American Type Culture Collection. The strain used for plasmid construction, *E. coli* DH5α [*sup*E44 *lac*U169 (*lac*ZM15) *hsd* R17 *rec*A1 *end*A1 *gyr*A96 *thi*-1 *re1*A1], was obtained from Gibco BRL. *E. coli* BZB1011, used in all assays as the colicin-sensitive indicator strain (''target cell''), was obtained from Dr. A.P. Pugsley and has been described previously (Pugsley and Oudega 1987). All cultures were routinely grown under standard conditions in Luria–Bertani (LB) broth or on LB agar plates supplemented when required with ampicillin (50 μg/ml).

Design of the Mutagenic Cassette. In this study we have focused on a 10-AA cassette (AA 423 to 432) located at the C-terminal region of the R domain of colicin E9 (Fig. 1). Biased mutagenesis of the 10-AA cassette allowed for a wide range of AA substitutions (Table 1). In particular, the bias avoided stop codons, limited the reappearance of the wild-type residues, and often resulted in substitutions altering the charge or polarity. The synthetic oligomer containing the degenerate cassette also included conserved 5′[44-nucleotide (nt)] and 3′ (33-nt) flanking regions that allowed for the amplification, restriction digestion, and cloning of the mutagenized cassette (Fig. 1D).

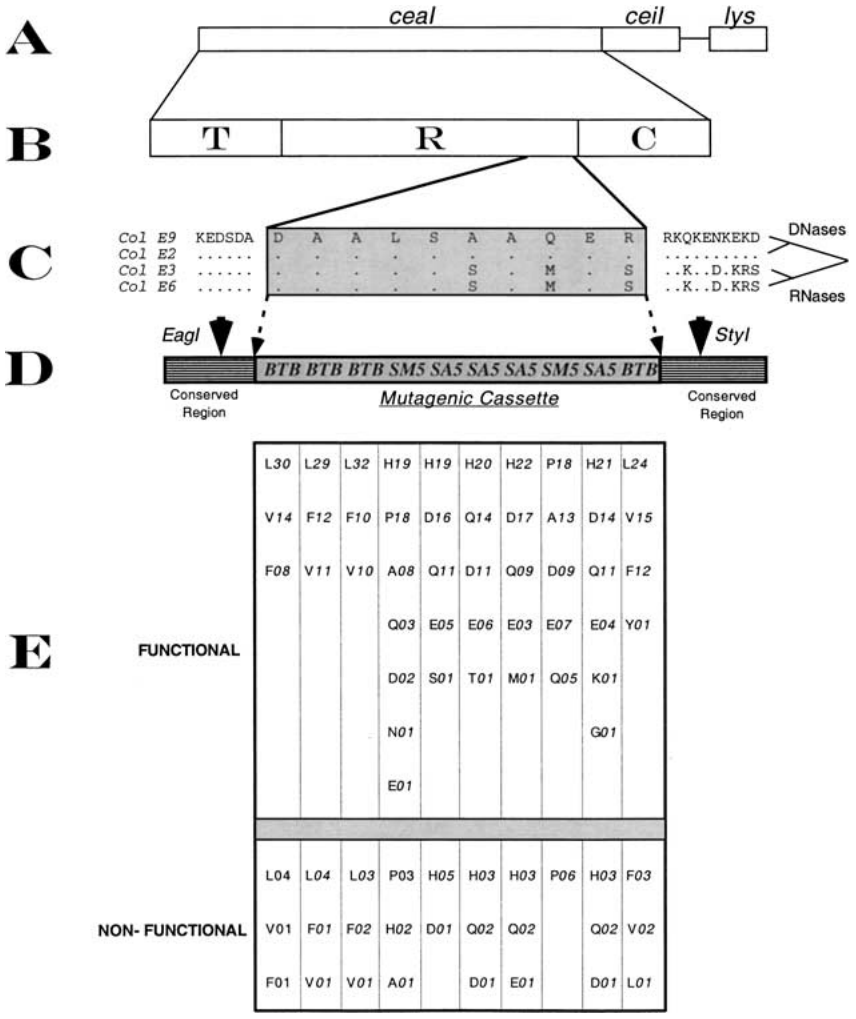

Fig. 1. **(A)** The colicin E9 gene cluster (*ceaI*, colicin gene; *cei*I, immunity gene; *lys*, lysis gene). **(B)** Schematic of the colicin E9 protein, with the translocation (T), receptor recognition (R), and killing (C) domains indicated. **(C)** The targeted 10-AA R-domain (shaded) and its flanking regions, showing the degree of sequence divergence in this region and the relationships among closely related colicins. **(D)** The overall design of the mutagenic cassette and the flanking conserved regions. The approximate positions of the restriction sites used for directional cloning are indicated. **(E)** The identity and frequency of amino acid replacements observed at each position in 52 functional and 6 nonfunctional engineered colicins.

Table 1. Allowable amino acid substitutions encoded by the mutagenic cassette[a]

Position									
1	2	3	4	5	6	7	8	9	10
Coding sequence of mutagenic cassette									
BTB	*BTB*	*BTB*	*SM5*	*SA5*	*SA5*	*SA5*	*SM5*	*SA5*	*BTB*
Expected amino acid replacements									
F	F	F	A	D	D	D	A	D	F
L	L	L	D	E	E	E	D	E	L
V	V	V	E	H	H	H	E	H	V
			H	Q	Q	Q	P	Q	
			P				Q		
			Q				H		

[a] *B*, T:G:C at a 1:1:1 proportion; *M*, A:C at a 1:1 proportion; *S*, G:C at a 1:1 proportion; *5*, G:C:T:A at a 3:3:3:1 proportion.

The mutagenic oligodeoxynucleotide cassette MC1 was chemically synthesized (1-μmol scale) at the Keck Oligosynthesis Center at Yale University. The oligodeoxynucleotide, AAT TTA T*CC TTG G*CG TCC TTT TCT TTA TTT TCC TTC TGT TTG CG VAV 5TS 5KS 5TS 5TS 5TS 5KS VAV VAV VAV AGC ATC TGA CTT CTC TTT TG*C GGC CG*C ATC AAA, contained partially randomized positions 2014 to 2043 of the 5523-nt-long pMc27 sequence (see below and Fig. 1D) (5 = 3:3:3:1 proportion for A:C:G:T, V = 1:1:1 proportion for A:C:G, S = 1:1 proportion for C:G, K = 1:1 proportion for T:G; constant primer-binding sequences are underlined; *Sty*I and *Eag*I restriction sites are italicized). This mutagenic oligonucleotide was subsequently made into a double-stranded fragment with PCR amplification using flanking oligomers complementary to the conserved regions. This PCR was performed in six reaction mixtures (100 μl each) containing 4 nmol of each primer

(MCA.fwd, 5′-AAT TTA TCC TTG GCG TCC-3′; and MCB.rev, 5′-TTT GAT GCG GCC GCA AAA G-3′), 3.8 μg of ssDNA template, PCR buffer [10 × buffer contains Tris–HCl, KCl, $(NH_4)_2SO_4$, 15 m*M* $MgCl_2$, pH 8.7], and 50 U of HotStart Taq DNA polymerase (QIAGEN, Germany) using a GeneAmp thermocycler (Model 9700; Perkin Elmer), under the following amplification conditions: 10 cycles of amplification, initial denaturation at 94°C for 15 min, annealing at 60°C for 1 min, primer extension at 72°C for 1 min, denaturation at 94°C for 1 min, and final extension at 72°C for 7 min. The resulting amplified DNA was purified using the PCR Purification Kit according to the manufacturer's instructions (QIAGEN) and double-digested with the restriction endonucleases *Eag*I and *Sty*I (1.8 μg of DNA and 30 U of each enzyme from NEB; 2 h at 37°C). The coding sequence obtained is thus the reverse complement of the MC1 oligomer described above. This reverse complementation is reflected in Fig. 1D.

Plasmid Construction. The initial plasmid used in this study, pMC27, was kindly provided by Dr. C. Penfold and has been described previously (Curtis and James 1991). This plasmid consists of a pUC18 backbone into which the colicin gene cluster containing the entire colicin E9 coding sequence (*cea*I), as well as the E9 immunity gene (*cei*I) and lysis gene (*lys*), has been inserted. To permit the rapid cloning of the mutagenized cassette, the pMC27 plasmid was modified by the addition of two new restriction sites. This was accomplished by identifying sites that differed at a single nucleotide from the canonical sequence of unique restriction sites. The modifications that introduce the restriction sites were incorporated as follows: plasmid pMC27 was isolated from *E. coli* JM83 using the QIAprep Miniprep Kit (QIAGEN). Fifty nanograms of plasmid DNA was amplified by PCR using 130 ng of each mutagenic primer (*Sty*I.fwd, 5′-AGA AAA GGA CGC CAA GGA TAA ATT-3′; and *Sty*I.rev, 5′-AAT TTA TCC TTG GCG TCC TTT TCT-3′) (*Sty*I restriction sites are underlined), synthesis buffer [10X buffer contains 200 m*M* Tris–HCl, 100 m*M* KCl, 100 m*M* $(NH_4)_2SO_4$, 20 m*M* $MgSO_4$, 1.0% Triton X-100, and 1 mg/ml BSA, pH 8.8 at 25°C], and 4.5 U of Pfu DNA polymerase (Promega) with a GeneAmp thermocycler (Model 9700; Perkin Elmer). Fifteen cycles of amplification were employed, with initial denaturation of the DNA at 94°C for 30 s, denaturation at 94°C for 30 s, annealing at 55°C for 1 min, and primer extension at 68°C for 12 min. The parental methylated DNA template was digested with *Dpn*I endonuclease (10 U of enzyme from NEB; 1 h at 37°C). The nicked vector DNA incorporating the desired mutation was then transformed into *E. coli* DH5 supercompetent cells (Life Technologies) following the protocol recommended by the manufacturer. Next, plasmid DNA from the transformed cells was isolated using the QIAprep Miniprep Kit (QIAGEN) and digested with the restriction endonuclease *Sty*I to confirm the desired mutation. The DNA was then used as the template in a second site-directed mutagenesis performed as described, above with the following modification: mutagenic primers *Eag*I.fwd, 5′-GCA TTT GAT GCG GCC GCA AAA GAG AAG-3′; and *Eag*I.rev, 5′-CTT CTC TTT TGC GGC CGC ATC AAA TGC-3′ (*Eag*I restriction sites are underlined). PCR was again performed as above. The desired mutation was confirmed by digestion of the plasmid vector with the restriction endonuclease *Eag*I. This resulted in the creation of unique sites in the plasmid that did not alter the coding sequence of the colicin E9 *cea*I gene. The presence in the modified (MpMC27) plasmid of the introduced restriction sites was subsequently verified by sequencing.

MpMC27 was then restriction digested with *Eag*I and *Sty*I, creating a directional cloning orientation and compatible ends for the insertion of the mutagenic cassette. The plasmid was dephosphorylated using calf intestinal alkaline phosphatase according to the manufacturer's instructions (New England BioLabs). Dephosphorylated plasmid was purified from agarose gel using the Gel Purification Kit (QIAGEN). The mutagenic cassette was subsequently ligated using T4 DNA ligase according to the manufacturer's instructions (New England BioLabs).

Transformation and Screening. MpMC27 plasmids containing variant colicins were transformed via electroporation into JM83 cells using the Gene Pulser II (Bio-Rad) under the following conditions: 50-μl volume of cells, 0.1-cm cuvette gap, 1.8-kV voltage, 18-kV/cm field strength, 25-μF capacitor, 200-Ω resistance, and 4.2- to 5.0-ms time constant. This resulted in transformation efficiencies of approximately 6×10^7 cfu/μg DNA (1.5×10^9 cells/ml). Our original protocol called for the simultaneous screening of lysates derived from 10^7 cells, each likely to contain a different variant of colicin E9 (for the mutagenized 10-AA region of the R domain). It soon became clear, however, that this could be replaced by a simpler assay involving replica-plating of transformed JM83 cells containing the variant MpMC27 plasmids, first onto LB agar containing ampicillin and subsequently onto LB plates preseeded with a lawn of sensitive *E. coli* BZB1011. The lawns were prepared using the standardized protocol for colicin induction and screening, which involves adding 10^6 sensitive cells and 0.5 μg/ml mitomycin C to 7 ml of top agar (20 g/L LB, 7 g/L bacto-agar), and poured over LB plates. After overnight incubation at 37°C, the plates were inspected for the presence of a clearing diameter on the BZB1011 lawn, indicating an active variant. Alignment of the LB/Amp and BZB1011 lawn plates allowed individual colonies to be selected, retested for active colicin production, and further characterized.

Sequence Analysis. Colony PCR was performed using reaction mixtures (50 μl each) containing 10 pmol of each primer (PMC27A.rev, 5′-GCT CCT GAA TCT TTA CCT GC-3′; and PMC27B.fwd, 5′-GGT CAC AGA ATG TGG CAA ATG G-3′), PCR buffer [10X buffer contains Tris–HCl, KCl, $(NH_4)_2SO_4$, 15 m*M* $MgCl_2$, pH 8.7], and 1.25 U of HotStart Taq DNA polymerase (QIAGEN). Each colony was picked with a sterile toothpick and transferred to the PCR master mix. Twenty-five cycles of amplification were employed, with initial denaturation of the DNA at 94°C for 15 min, annealing at 60°C for 1 min, primer extension at 72°C for 1 min, denaturation at 94°C for 1 min, and final extension at 72°C for 7 min. Amplified DNA was purified using the PCR Purification Kit (QIAGEN).

The first screening of mutagenic clones was performed by digestion of an amplified product aliquot with the restriction endonuclease *Mse*I (NEB; 2 h at 37°C), directed at a site present in the wild-type sequence but unlikely to be conserved in any of the engineered variants. The products of digestion were analyzed on a 3% agarose gel, and clones exhibiting a band pattern different from the wild-type sequence were selected for sequencing. DNA sequencing was performed using the BigDye Terminator Kit (Perkin Elmer) according to the manufacturer's instructions, and products were visualized on the automated ABI 377 sequencer.

Scoring and Further Phenotypic Characterization. One hundred microliters of a fresh overnight growth of sensitive *E. coli* BZB1011 was added to 3 ml of top agar (20 g/L LB, 7 g/L bacto-agar) and poured onto an LB plate. Two microliters of lysate of each colicin variant and of the wild-type colicin (derived from the control plasmid MpMC27, containing the new restriction sites and the wild-type E9 coding sequence) was then spotted twice onto this BZB1011 lawn. After overnight incubation at 37°C, the phenotype was scored as + + + when the colicin produced a clear zone of growth inhibition indistinguishable from the wild-type colicin, as + + when the colicin produced a visible translucent clearing zone of reduced diameter (relative to the wild type), and as + when it produced a visible but faint clearing in the lawn with an opaque plaque. When no zone of inhibition was present, the phenotype was scored as −.

Lysate Preparation. One hundred microliters of a fresh overnight growth of each active colicin variant and wild-type colicin producer cells was transferred into 5 ml of fresh LB medium in 50-ml Erlenmeyer flasks. The cultures were grown at 37°C with agitation. When the A_{600} reached 0.2, mitomycin C (0.5 μg/ml) was added to induce colicin production, and growth was continued for 2.5 h. Aliquots of each culture were transferred to an Eppendorf tube. The cells were lysed by adding 100 μl of chloroform and vortexing. After centrifugation at 10,000*g* for 10 min, the supernatant containing the colicin molecules was transferred to a clean Eppendorf tube and stored at −70°C.

Growth Rate Assays. Cells containing the variant and wild-type MpMC27 plasmids were taken from storage at −70°C and grown overnight at 37°C in 5 ml of LB, using 50-ml Erlenmeyer flasks under constant rotation at 250 rpm. Three replicate flasks (per strain) of a 10^{-3} dilution of the overnight growth were prepared in 14 ml of fresh LB. The cultures were incubated at 37°C with agitation at 250 rpm. Cell growth was monitored after 2 h by measuring the absorbance at 600 nm for each culture, using a Benchmark Microplate Reader (Bio-Rad) at 20-min intervals for up to 6 h. Measurements of OD_{600} at each time point were log-transformed and regressed against time; the growth rate was calculated using the slope of the regression. Whenever necessary, the growth rate was translated into the mean doubling time for the cell line. The growth rate for each engineered strain was measured three times as described above. The mean growth rate for each strain was calculated, along with its standard deviation, and compared with the growth rate of the strain harboring the wild-type colicin. The doubling times for our engineered strains were compared with the wild-type growth rates employing a *t* test corrected for multiple comparisons using Dunnett's method, as implemented in the software package JMP (SAS Institute 1999).

Results

Phenotypic characterization of 5280 colonies harboring variant MpMC27 plasmids revealed that 3720 of them (70.45%) produced a visible clearing diameter when plated onto a sensitive BZB1011 lawn. One hundred such colicin-producing transformants, as well as 85 inactive transformants, were selected at random and individually isolated for more detailed phenotypic and genotypic characterization.

Further characterization involved a four-step process: (1) isolation of the engineered MpMC plasmid containing the variant colicin, (2) retransformation of the native JM83 host cell with the isolated plasmid, (3) confirmation of the bactericidal activity of the new transformant, and (4) sequencing of a region of the colicin E9 region containing the mutagenic cassette. The result of this second-stage screening confirmed the initial screen: all of the putative positives show a colicin-like ability to clear a visible diameter on a sensitive BZB1011 lawn. Conversely, the loss-of-function transformants that did not show bactericidal activity in the first screen continued to appear negative in subsequent assays. Following this confirmation, 52 active ("functional") colicin-producing transformants were chosen for subsequent characterization (Table 2).

Sequencing revealed that 79 of 85 (93%) inactive constructs showed a deletion or insertion resulting in a frame shift of the open reading frame. Only six of the negative constructs retained their reading frame and therefore represented full-length AA sequences that did not confer the bactericidal phenotype. These six sequences were used in subsequent analyses.

Sequencing of the isolated positive variants revealed the enormous sequence flexibility of this portion of the receptor-binding region (Table 2 and Fig. 1E), reflected in the diversity of AA sequences compatible with a bactericidal phenotype. The sequences of both active (functional) and inactive ("nonfunctional") variants were subjected to a clustering analysis, shown in Fig. 2. As this figure suggests, there is little hierarchical structure evident in this clustering analysis: the sequences are equidistant from one another, differing on average at 7.2 AA (of a possible 10 AA in the targeted cassette) and roughly equidistant from the original sequence (mean distance, 9.6 AA). We note that the design of the mutagenic cassette allows for only a subset of all 20 AA at each of the targeted positions (Table 1). However, certain additional AA, not encoded in the original mutagenic cassette, do appear, likely as a result of PCR error (cassette positions 4:N, 5:S, 6:T, 7:M, 9:G,K, 10:Y). We find at least one representative (save for cassette position 8:H) of each possible AA encoded by our cassette at each of the 10 positions. Subtle biases are evident in the AA present at each of the 10 positions (Fig. 1E and Table 2). Thus, for instance, the cassette design for the first codon (G/C/T T G/C/T) allows for three AA (F, L, V), with the following expected proportions: L (∼44.4%), V (∼33.3%), and F (∼22.2%). We obtain quite different proportions in our pool of active variants: L (∼57.7%), V (∼26.9%), and F (∼15.4%).

The active constructs display a range of antimicrobial activities, scored qualitatively in relation to the phenotype of wild-type E9 colicin (Table 2). Constructs with a + + + designation show clearing diameters similar to those seen with wild-type Co1E9 lysate; + + and + denote decreasing clearing diameters. There is no obvious correlation between the sequence of the engineered colicin and the extent of antimicrobial activity it displays. Similarly, no single AA change (or suite of changes) appears to underlie the loss of the bactericidal phenotype. Instead, the sequences of inactive variants are distributed throughout the clustering diagram (Fig. 2) and do not form a clearly defined subset in sequence space.

The cells carrying engineered colicin plasmids were further characterized by examining their relative growth rates under ideal laboratory growth conditions. Such a measure was intended to quantify the possible fitness costs (or benefits) conferred by a variant colicin on its JM83 host cell line. In these

Table 2. Molecular and phenotypic characterization of active and inactive colicin E9 engineered variants

No.	AA sequence	Killing activity	Mean relative doubling time ± SD	No.	AA sequence	Killing activity	Mean relative doubling time ± SD
wt[a]	DAALSAAQER	+ + +	1.000 ± 0.062	wt	DAALSAAQER	+ + +	1.000 ± 0.062
178	LFFPQHDAHF	+ + +	0.750 ± 0.053*	2	VLFHDQQ.DL	+ + +	1.188 ± 0.047
5	FLLDEDHAHL	+ + +	0.890 ± 0.094	206	LLLPQDHPDV	+ +	1.227 ± 0.016
23	LLFN.TMPKV	+ + +	0.897 ± 0.020	175	VLVAEQHAHL	+ + +	1.234 ± 0.023
27	LLLPHEQE.V	+ +	0.903 ± 0.028	182	FFLPEHQPHL	+ + +	1.245 ± 0.162
169	VLLHDHEADV	+ + +	0.911 ± 0.060	38	FVLHDEDDHV	+ + +	1.249 ± 0.042
21	LLVPDQHEHY	+ +	0.929 ± 0.113*	1	LVFAHHDEDF	+ + +	1.251 ± 0.034
209	LLVPHHHADV	+	0.935 ± 0.063	25	LFVHEQHPHF	+	1.261 ± 0.065*
11	FLLPQHDP.V	+ + +	0.947 ± 0.070	29	LVVHQDDADL	+ + +	1.263 ± 0.029
14	VLLAHHH.HL	+ + +	0.947 ± 0.039	6	LFLAQHQPHL	+ + +	1.269 ± 0.022*
212	FFLQHDHDQF	+ +	0.986 ± 0.112	168	LLVHQHQ..F	+ + +	1.279 ± 0.154
22	LLFPQHHEQF	+ +	0.988 ± 0.058	42	LLLHHEHDDF	+ + +	1.291 ± 0.119*
176	VVLPHDQDQV	+ +	0.996 ± 0.025	37	VVLQDQHAHL	+ + +	1.305 ± 0.073*
34	FLLAHQDDHL	+ + +	1.009 ± 0.032	10	LVFPDQQDQL	+ + +	1.306 ± 0.143*
36	LFVHDDD.QL	+ + +	1.009 ± 0.044	17	LFLPDDEDHL	+ + +	1.367 ± 0.092*
202	LVLHHDHPQF	+	1.012 ± 0.035	7	VVFADHH.HL	+ + +	1.376 ± 0.063*
28	LLLPHDDEHL	+ + +	1.064 ± 0.122	200	LLLHDQHPQL	+ +	1.386 ± 0.016*
35	LFLAQQHPHL	+ + +	1.066 ± 0.076	173	LLLHHEHDDL	+ + +	1.419 ± 0.078*
20	LFLADQDPGF	+ + +	1.071 ± 0.108	170	FVLHHEHPDL	+ + +	1.488 ± 0.180*
32	LLLHDDDPDL	+ + +	1.072 ± 0.026	26	LLLDHQHADF	+ + +	1.500 ± 0.092*
213	LFLHDHDADL	+ + +	1.082 ± 0.057	181	LLFHEHDEHV	+ + +	1.503 ± 0.094*
177	VLLPHHHPHF	+ + +	1.087 ± 0.035	31	LVLHHHHEQV	+ + +	1.548 ± 0.130*
33	VFLEHHQAHL	+ + +	1.108 ± 0.107	172	VLFPDHDD.L	+ + +	1.556 ± 0.096*
40	VLFQHQDPHV	+ + +	1.116 ± 0.114	179	LLLHHHHADL	+ + +	1.610 ± 0.160*
4	VVVPQHQPDV	+	1.116 ± 0.131	*57*	*FLLPHQHPHL*	–	*0.905 ± 0.022*
9	LFVPDEDPQV	+ + +	1.131 ± 0.017*	*41*	*LLFPHHQPHV*	–	*0.968 ± 0.065*
180	FLLHHHDPQV	+ + +	1.159 ± 0.149	*19*	*VLLHHQQPQF*	–	*0.998 ± 0.036*
15	VLLPQQDAHL	+ + +	1.164 ± 0.198	*146*	*LLFPHHEPQV*	–	*1.040 ± 0.089*
16	VLLHDQHAQV	+ +	1.178 ± 0.050*	*207*	*LVVHHDHPHF*	–	*1.115 ± 0.075*
8	LLVPQDEPHF	+ + +	1.188 ± 0.045	*18*	*LFLADHHPDF*	–	*1.132 ± 0.044**

[a] Wild type.

assays, all JM83 cell lines were identical in all respects except for the 30-nt region within the R domain of the colicin E9 coding region. Consequently, any difference in growth rate could be traced back to that 30-nt region. The results of this analysis are summarized in Table 2. They reveal that, with two exceptions, none of the variant colicins results in a significant decrease in the mean doubling time of the cell line (relative to a JM83 line harboring the wild-type MpMC27 plasmid). The exceptions are the JM83 cell lines containing variants 178 and 21, whose relative doubling times are 33 and 8% shorter, respectively, than that of the control line. In contrast, 17 of the engineered variants significantly increased the doubling times for their respective cell lines.

Finally, the engineered colicins did not show a weakened killing phenotype or killing activity relative to wild-type E9. The qualitative comparison, summarized in Table 2, reveals that 40 of the 52 engineered variants exhibit clearing activity on a sensitive BZB1011 lawn comparable to that of wild-type colicin E9. In addition, when the antimicrobial activity of lysates that had been repeatedly diluted 10-fold were tested on a sensitive BZB1011 lawn, the limiting dilution for 43 of the 52 constructs tested was comparable (10^{-2} dilution) to that of wild-type Co1E9 (data not shown).

Discussion

The E-series colicins utilize the BtuB cell surface receptor for entry into the target cell (James et al. 1996). This series includes the pore former colicin E1; the DNase colicins E2, E7, E8, and E9; and the RNase colicins E3, E5, and E6. Among the DNase and RNase colicins the receptor recognition domain is highly conserved, with 94% of the residues identical (Lau et al. 1992) (Fig. 1C). In addition, DNA sequence polymorphism surveys of colicins E1 and E2 revealed no sites segregating in the R domain (Tan and Riley 1997a; Riley and Gordon 1992). This lack of interspecific divergence and of intraspecific polymorphism in the R domain led us to predict that this region would be under strong selective constraint and that functional AA replacements would be rare. We assumed a direct connection between functional constraint and sequence conservation. Thus, we initially designed our activity screen to enable the recovery of active constructs, which we estimated would occur at

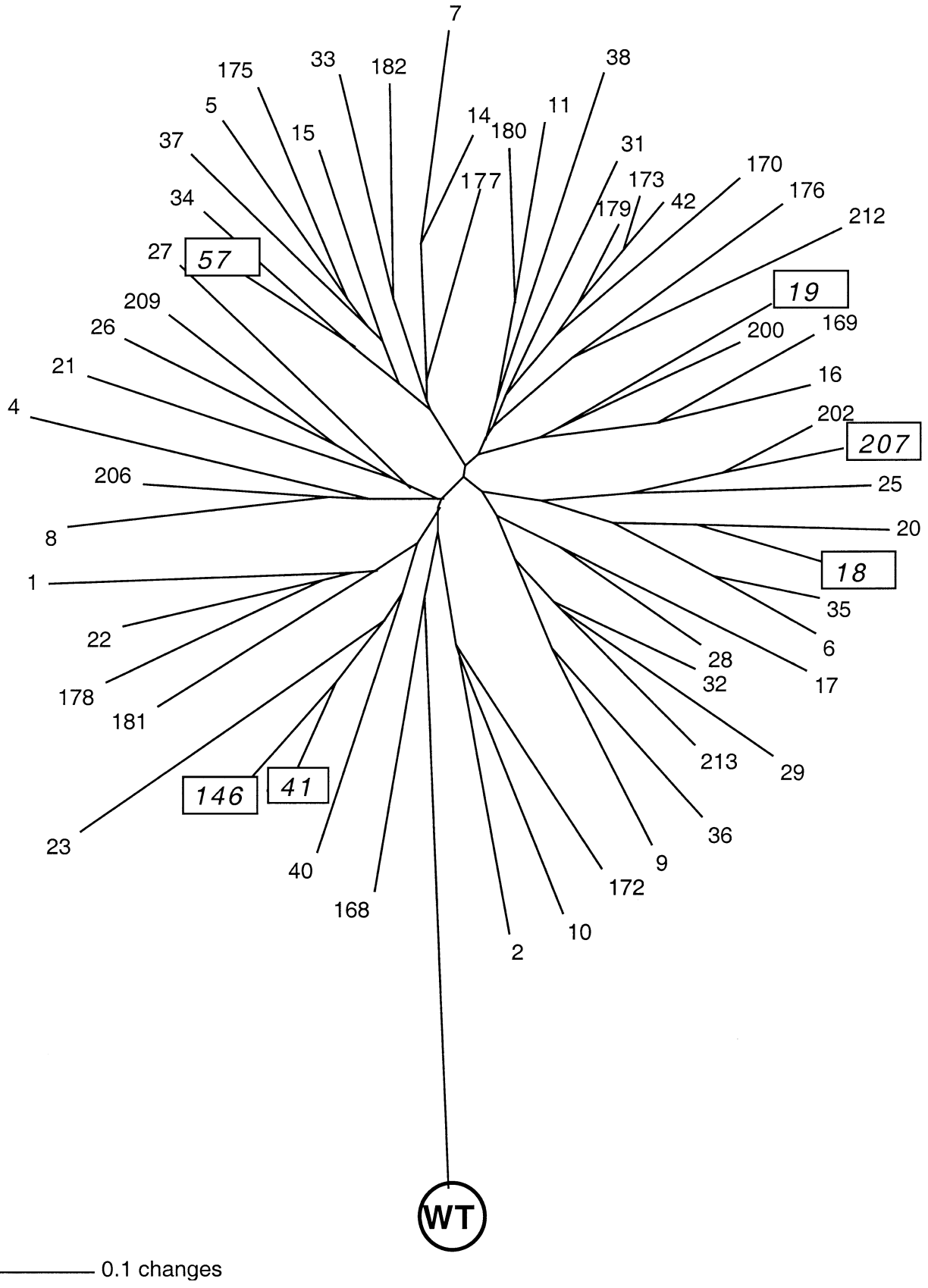

Fig. 2. Phylogram based upon neighbor-joining methods (Saitou and Nei 1987) for the wild type (WT) and 52 active and 6 inactive (boxed) colicin E9 engineered variants.

or below a frequency of 10^{-7} in the mutagenized pool. The results we obtained strongly contradict this expectation. In effect, with a frequency of active variants in excess of 70% of the transformants surveyed, we had little difficulty in retrieving functional colicins that diverged significantly from the wild-type colicin E9 in the sequence of the target cassette.

No less puzzling is the enormous diversity of AA sequences that are compatible with colicin function. As shown in Fig. 1E and Table 2, the number of sequences we recovered from active variants is surprisingly large, and no sequence appeared more than once in our sample. The region we are addressing is at the boundary of the interaction between colicin E9 and the BtuB receptor, an interaction that may be largely structurally mediated (Penfold et al. 2000). Nevertheless, we would have expected at least some sequence-dependent pattern to emerge.

A number of studies have shown that certain central features of protein structure are surprisingly robust to changes in AA sequence (Luciano and Karlin 2000; Osvaldo et al. 1999; Skirgaila et al. 1998; Huang and Blundell 1999; Aronson et al. 1994; MacKinnon et al. 1998). The sequence flexibility of the colicin E9 R domain suggests that the structural interaction between the colicin and the BtuB cell surface receptor is not disrupted by the AA substitutions we recovered. Given that the interactions both with the receptor and with the translocation machinery depend primarily on the three-dimensional structure, we predict that similar sequence diversity can be recovered from in vitro engineering of the colicin E9 translocation domain.

How can the low levels of divergence seen in the comparative analyses be reconciled with the large numbers of diverse functional molecules emerging from these in vitro protein engineering studies? Three broad explanations emerge. The first, function-based, explanation would argue that the engineered colicins we have generated, while still retaining the ability to kill sensitive BZB cells, do so with a lower efficiency than the wild-type colicin E9 molecule. If so, this implies that the version of colicin E9 actually seen in natural populations is, in some sense, the best around. Other sequences may result in functioning colicin E9, but they produce an evolutionarily inferior product. We examined this possibility by investigating the limiting dilution for each of the characterized engineered colicins relative to that for wild-type colicin E9. The results certainly suggest a lower activity for some of the synthetic colicins: in some cases, a 10-fold dilution of the engineered colicin lysate is enough to make the antibacterial phenotype disappear (versus 10^{-2} dilution for wild-type colicin E9). Most of the engineered colicins, however, behave as wild-type colicin E9 does, retaining visible activity up to a 10^{-2} dilution. The simple explanation—that all engineered colicins are phenotypically less active (at least in this qualitative assay)—is not borne out. What our assays cannot rule out, however, is the possibility that our engineered colicins would exhibit impaired killing activity (relative to wild-type E9) when presented with a broader diversity of target cells, akin to what might be encountered in natural environments.

The second explanation that reconciles the diversity of engineered colicins with the paucity of naturally occurring colicin E9 variants centers on possible fitness costs to the host JM83 cell of carrying an alternative version of the R domain. A variety of possible mechanisms, such as codon bias, could in principle impose a fitness cost associated with the engineered plasmids. We explored this scenario by measuring the growth rates of cells containing alternative versions of the colicin coding sequence and comparing them to the doubling time of a JM83 host carrying wild-type colicin E9 on the plasmid. We indeed found that many of the engineered colicins impose a fitness cost on the host, likely reflecting pleiotropic effects of the altered sequence. However, this is not always the case. Many of the strains carrying engineered colicins grow as fast as, and in two cases faster than, the wild-type control (Table 2), at least under our experimental conditions.

The third explanation centers on the population genetic mechanisms shaping colicin diversity. Prior studies have emphasized the potential role of diversifying selection in eliminating variation across most of the colicin gene cluster (Riley 1993a, b, 1998; Tan and Riley 1997b). The central homogenizing force is the interaction between the colicin immunity protein and that portion of the colicin protein that serves to bind the immunity protein. Novel immunity proteins capable of binding both ancestral and related colicin proteins enjoy a strong selective advantage and are quickly driven to fixation. The relatively low level of recombination characteristic of these plasmids results in incidental fixation of the entire colicin plasmid replicon. Because selective sweeps are frequent and recurrent, a single plasmid sequence dominates the population at any given time. From sampling considerations alone, this is the sequence most likely linked to the advantageous immunity mutation. As a result, this selected sequence will be swept to fixation, resulting in the elimination of all functional alternative sequences. Thus, the fact that none of our engineered colicins has been found in natural populations may reflect neither on their functionality nor on the relative fitness they confer. Instead, their absence may simply result from the operation of population-level phenomena, including these recurrent selective sweeps. Such linkage-based explanations, however, cannot readily account for the low level of divergence seen in this region among the E-series colicins.

This study echoes the findings of a recent report documenting the sequence tolerance seen at the active site of DNA polymerase (Patel and Loeb 2000). Such studies serve an important cautionary purpose and remind us of the important distinction between molecular function—a boundary condition that must be met for organisms to survive and reproduce—and the selective and populational forces that govern the fate of variants in natural populations. This distinction is not merely semantic: the advent of genomics has meant an explosion of functional inferences drawn from sequence comparisons. It behooves us to remember that invariance in natural populations has many causes and may not simply come about because alternative sequences are not capable of carrying out the needed function.

Acknowledgments. This study was supported by grants from the NIH (R21GM60044-01-02), the Rockefeller Brothers Fund, and the Charles E. Culpeper Foundation. We are grateful to Cynthia Hunt, Carla Goldstone, and Livia Tomasini for their superb technical support and to Richard Lawler for statistical assistance. The manuscript was greatly improved by the comments of Magnus Norberg and of Siobian Duffy and by the suggestions and persistence of two anonymous reviewers.

References

Aronson HEG, Royer WE Jr , Hendrickson WA (1994) Quantification of tertiary structural conservation despite primary sequence drift in the globin fold. Protein Sci 3:1706–1171

Chak K, James R (1986) Characterization of the ColE9-J plasmid and analysis of its genetic organization. J Gen Microbiol 132:61–71

Curtis M, James R (1991) Investigation of the specificity of the interaction between colicin E9 and its immunity protein by site-directed mutagenesis. Mol Microbiol 5:2727–2734

Eaton T, James R (1989) Complete nucleotide sequence of the colicin E9 (cei) gene. Nucleic Acids Res 17:1761

Huang JF, Blundell TLI (1999) The relations of protein sequence and structural conservation with function. Zool Res 20:21–25

Irving JA, Pike RN, Lesk AM, Whisstock JC (2000) Phylogeny of the serpin superfamily: Implications of patterns of amino acid conservation for structure and function. Genome Res 10:1845–1864

James R, Lazdunski C, Pattus F (eds) (1991) Bacteriocins, microcins and lantibiotics. NATO ASI series. Springer-Verlag, New York, Vol 65

James R, Kleanthous C, Moore GR (1996) The biology of E colicins: Paradigms and paradoxes. Microbiology 142:1569–1580

Lau PCK, Parsons M, Uchimura T (1992) Molecular evolution of E colicin plasmids with emphasis on the endonuclease types. In: James R, Lazdunski C, Pattus F (eds) Bacteriocins, microcins and lantibiotics. Springer-Verlag, Berlin/Heidelberg, Vol H 65

Luciano B, Karlin S (2000) Conservation among HSP60 sequences in relation to structure, function, and evolution. Protein Sci 9:476–486

MacKinnon R, Cohen S, Lee A, Kuo A, Chait BT (1998) Structural conservation in prokaryotic and eukaryotic potassium channels. Science 280:106–109

Osvaldo O, Burkhard R, Alfonso V (1999) Effective use of sequence correlation and conservation in fold recognition. J Mol Biol 293:1221–1239

Owen G, Zelent A (2000) Origins and evolutionary diversification of the nuclear receptor superfamily. Cell Mol Life Sci 57:809–827

Patel PH, Loeb LA (2000) DNA Polymerase active site is highly mutable: Evolutionary consequences. Proc Natl Acad Sci USA 97:5095–5100

Penfold C, Garinot-Schneider C, Hemmings A, Moore G, Kleanthous C, James R (2000) A 76 residue polypeptide of colicin E9 confers receptor specificity and inhibits the growth of vitamin B12-dependent *E. coli* 113/3 cells. Mol Microbiol 38:639–649

Pugsley AP, Oudega B (1987) Methods for studying colicins and their plasmids. In: Hardy KG (ed) Plasmids, a practical approach. IRL Press, Oxford, pp 105–161

Reizer J, Saier MJ (1997) Modular multidomain phosphoryl transfer proteins of bacteria. Curr Opin Struct Biol 7:407–415

Riley MA (1993a) Positive selection for colicin diversity in bacteria. Mol Biol Evol 10:1048–1059

Riley MA (1993b) Molecular mechanisms of colicin evolution. Mol Biol Evol 10:1380–1395

Riley MA (1998) Molecular mechanisms of bacteriocin evolution. Annu Rev Genet 32:255–278

Riley MA, Gordon DM (1992) A survey of Col plasmids in natural isolates of *Escherichia coli* and an investigation into the stability of Col-plasmid lineages . J Microbiol 138:1345–1352

Riley MA, Gordon D (1995) Ecology and evolution of bacteriocins. J Industr Microbiol 17:155–158

Saitou N, Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

SAS Institute (1999) JMP version 3. SAS Institute, Cary, NC

Skirgaila R, Grazulis S, Bozic D, Huber R, Siksnys V (1998) Structure-based redesign of the catalytic/metal binding site of Cfr101 restriction endonuclease reveals importance of spatial rather than sequence conservation of active centre residues. J Mol Biol 279:473–481

Tan Y, Riley MA (1997a) Nucleotide polymorphism in colicin E2 gene clusters: Evidence for nonneutral evolution. Mol Biol Evol 14:666–673

Tan Y, Riley MA (1997b) Positive selection and recombination: Major molecular mechanisms in colicin diversification. Trends Ecol Evol 12:348–351

Yarus M, Welch M (2000) Peptidyl transferase: Ancient and exiguous. Chem Biol 7:187–190

Yoshida Y, Aoyama Y, Noshiro M, Gotoh OI (2000) Sterol 14-demethylase P450 (CYP51) provides a breakthrough for the discussion on the evolution of cytochrome P450 gene superfamily. Biochem Biophys Res Commun 273:799–804