The coupling of the biosynthesis and export of the DNA gyrase inhibitor simocyclinone in *Streptomyces antibioticus*

By

Tung Ba Khanh Le

2011

A thesis submitted to the University of East Anglia, Norwich, for the Degree of Doctor of Philosophy

Abstract

Because most antibiotics are potentially lethal to the producing organism, there must be mechanisms to ensure that the machinery responsible for export of the mature antibiotic is in place at the time of biosynthesis. Simocyclinone D8 is a potent DNA gyrase inhibitor produced by *Streptomyces antibioticus* Tü 6040. Within the simocyclinone biosynthetic cluster are two divergently transcribed genes, *simR* and *simX*, encoding proteins that resemble the TetR/TetA repressor-efflux pump pair that cause widespread resistance to clinically important tetracyclines. In this thesis, I show that engineered expression of *simX* from a strong, heterologous promoter confers high level simocyclinone D8 resistance on *Streptomyces lividans*, showing that *simX* encodes a simocyclinone efflux pump. Transcription of *simX* is controlled by SimR, which directly represses the *simX* and *simR* promoters by binding to two operator sites in the *simX-simR* intergenic region. Simocyclinone D8 abolishes DNA binding by SimR, providing a mechanism that couples the biosynthesis of simocyclinone to its export. In addition, an intermediate in the biosynthetic pathway, simocyclinone C4, which is essentially inactive as a DNA gyrase inhibitor, also induces simX expression *in vivo* and relieves *simX* repression by SimR *in vitro.*

I also report the crystal structures of SimR alone and in complex with either simocyclinone D8 or simocyclinone C4. The ligand-binding pocket is unusual compared to those of other characterized TetR-family transcriptional regulators: the structures show an extensive ligand-binding pocket spanning both monomers in the functional dimeric unit, with the aminocoumarin moiety of simocyclinone D8 buried in the protein core, while the angucyclic polyketide moiety is partially exposed to bulk solvent. Through comparisons of the structures, I postulate a derepression mechanism for SimR that invokes rigid-body motions of the subunits relative to one another, coupled with a putative locking mechanism to restrict further conformational change.

The DNA-binding domain of SimR has a classical helix-turn-helix motif, but it also carries an arginine-rich N-terminal extension. The structures of SimR alone or in complex with ligands showed that this N-terminal extension is disordered in the

absence of DNA. Here I show that the N-terminal extension is sensitive to protease, but becomes protease-resistant upon binding DNA. I demonstrate by deletion analysis that the extension contributes to DNA binding, and describe the crystal structure of SimR bound to its operator sequence, revealing that the N-terminal extension binds in the minor groove. Bioinformatic analysis shows that an N-terminal extension rich in positively charged residues is a feature of the majority of TetR family members. Comparison of the SimR-DNA and SimR-simocyclinone complexes reveals that the conformational changes associated with ligand-mediated derepression indeed result primarily from rigid-body rotation of the subunits about the dimer interface.

Abbreviations

apr	apramycin resistance gene
AC	aminocoumarin
bla	carbenicillin resistance gene
kan	kanamycin resistance gene
hyg	hygromycin resistance gene
bp	base pair
CFU	colony forming unit
DNA	deoxyribonucleic acid
DBD	DNA-binding domain
DTT	dithiothreitol
dH_2O	distilled water
EDTA	ethylenediaminetetraacetic acid
EMSA	electrophoretic mobility shift assay
EtBr	ethidium bromide
GyrA	DNA gyrase subunit A
GyrB	DNA gyrase subunit B
kDa	kilodaltons
kb	kilobase pairs
LBD	Ligand-binding domain
LPS	Lipopolysaccharides
MIC	minimum inhibitory concentration
HTH	Helix-turn-helix
IC_{50}	half maximal inhibitory concentration
min	minute
sec	second
S. antibioticus	*Streptomyces antibioticus* Tü 6040
OD	optical density

ORF	open reading frame
PAGE	polyacrylamide
PDB	Protein Data Bank
PK	polyketide
rpm	revolutions per min
RBS	ribosome binding site
RNA	ribonucleic acid
SDS	sodium dodecyl sulphate
SC4	Simocyclinone C4
SD8	Simocyclinone D8
SFM	soya flour and mannitol medium
TEMED	N, N, N', N'-tetramethylethylethylenediamine
TFR	TetR-family transcription regulator

Acknowledgements

Happiness is a journey, not a destination
Dance as though no one is watching you
Love as though you have never been hurt before
Sing as though no one can hear you
Live as though heaven is on earth
(Anonymous)

This quote from an anonymous author describes my "*journey*" through my PhD just perfectly, although I would want to add in just one more line "*work as though you don't need money*"! It was a truly fantastic journey. I enjoy it to the most and I would like to thank a few people who brought me "*happiness*".

The first person I am deeply thankful to is Mark, my supervisor. Thank you for always pointing me in the right direction, and for listening to and respecting my opinion. Most of all, thank you for spending a great deal of time teaching me how to communicate science effectively. I can feel myself growing as a person as well as a scientist under your guidance. The last four years are probably the most memorable years of my youth. Needless to say, you have played a lot of positive roles in it. Thank you very much. Words cannot describe my deepest gratitude to you.

Secondly, I would like to thank Dave Lawson and Clare Stevenson for introducing me to the world of X-ray crystallography and helping me with all the structures we have solved so far. I would never have thought X-ray crystallography was possible, but you certainly made me believe it so now.

I would like to thank Ray Dixon and Tony Maxwell, who gave a lot of helpful comments and suggestions on experiments in every supervisory committee meetings.

I would like to thank Richard Brennan and Maria Schumacher for hosting me in their laboratories at the M. D. Anderson Cancer Center, Houston, Texas, and for showing me the technique of protein-DNA crystallisation.

I also want to thank my Mol Micro colleagues for being great friends and helping me with many experiments. Thanks to Chris den Hengst for guiding my first steps in working with radioactive isotopes. Thanks to Maureen for helping me with *Streptomyces* genetics. The most frequent question I have been asking her is "Does this look like *Streptomyces*, Maureen?" Thanks to Mahmoud Al-Bassam and Antje Hempel for making lab life exciting every day.

I also want to thank Ngat, my labmate, my girlfriend, then my wife, for always being with me. Thank you for your patience, for your willingness to do all the housework so that I can spend time in front of the computer or at the bench. It could have been a very difficult and boring 4 years without you right from the start.

I would like to thank the John Innes Foundation for funding my PhD work. This is the only place that funds overseas students for a PhD. Without this generous fund, I simply could not have been here at all.

Lastly I would like to thank my parents the most. They never invested in any business in their life but had great courage to invest all their savings in my education. Mum and Dad, it was an incredibly stressful time financially during my A-levels and undergraduate degree, wasn't it? I was so lucky to win a PhD position at the John Innes Centre. It was truly a time when I didn't have to worry financially, even saving a great deal to send back home. In fact, it might be the only time in my life that I am totally engulfed in my science and totally living my dream.

I thank you all for giving me such a wonderful time!

Tung BK Le

Table of contents

Chapter 1
General Introduction

1.1 Self-resistance to antibiotics in the producing organisms

By definition, "antibiosis" means "against-life". Antibiotics are toxins. They can potentially kill the producing hosts. Therefore, in order for antibiotic-producing organisms to survive, they should have a self-resistance mechanism to counteract the potential toxicity of their own products (Cundliffe and Demain, 2010). Academically, it is interesting to discover the self-resistance mechanism.

Outside the context of the producing organism, antibiotic resistance in clinical pathogens poses numerous challenges. The rise of antibiotic resistance and the wide dissemination of resistant determinants is a major threat to public health. It is widely held that the resistance determinants found in clinical pathogens are ultimately derived from the organisms that produce the corresponding antibiotics (Davies, 1994; Wright, 2007). Therefore, there is a need to expand research into resistance in antibiotic-producing organisms. In doing so, it might be possible to predict resistance before it emerges clinically and to develop diagnostic techniques and future therapeutic strategies.

1.2 Mechanisms of antibiotic resistance and antibiotic export

1.2.1 Mechanisms of antibiotic resistance

There are five major mechanisms of antibiotic resistance. They are:

1. Decreased permeability and uptake.
2. Export.
3. Enzymatic inactivation of the antibiotic.
4. Modification or overexpression of the target of the antibiotic.
5. Induced lost of the target of the antibiotic.

The first mechanism is decreased permeability and uptake. For example, *Neisseria gonorrhoea* can acquire mutations in the gene encoding the porin channel and these mutations cause reduced penicillin and tetracycline permeability, thereby conferring resistance to these compounds (Olesky *et al*., 2006). The second mechanism of resistance is efflux. The first determinant of this kind to be discovered was the TetA

tetracycline efflux pump, which is described in more detail in section 1.6 (McMurry *et al*., 1980; Ball *et al*., 1980). The third mechanism is the enzymatic inactivation of the antibiotic such as the inactivation of penicillins by β-lactamases, which cleave the β-lactam ring (Holt and Stewart, 1964). Other examples of this type of resistance mechanism are aminoglycoside modifying enzymes that inactivate the antibiotic by catalysing the addition of an acetyl or phosphoryl group (for example in the case of apramycin or kanamycin resistance) (Skeggs *et al*., 1987). The fourth mechanism involves altering the target of the antibiotic. This is typical for vancomycin resistance. Under susceptible conditions, vancomycin prevents peptidoglycan crosslinking by binding to the D-Ala-D-Ala termini of lipid-attached peptidoglycan precursors on the outside of the cytoplasmic membrane (reviewed in Buttner *et al*., 2008). Bacteria (mostly Gram-positive) acquire resistance by reprogramming cell wall biosynthesis such that their peptidoglycan precursors terminate in D-Ala-D-Lactate rather than in D-Ala-D-Ala, which blocks vancomycin binding (Buttner *et al*., 2008). The last mechanism of antibiotic resistance is relatively newly discovered. In this mechanism, the antibiotic is a pro-drug that requires enzymatic activation to generate the active antibacterial compound. Metronidazole is a prodrug that is activated by the nitroreductase RdxA and null mutations in the *rdxA* gene therefore cause resistance to metronidazole (Jenks and Edwards, 2002).

Other physiological states of bacteria, beyond the scope of this thesis, can also contribute to antibiotic tolerance, such as biofilm formation, or the presence of persisters in the bacteria population (Cos *et al*., 2010; Lewis, 2010).

1.2.2 Antibiotic resistance mediated by efflux pumps

Efflux pump-mediated export, both for specific drugs or for multi-drugs, is an important intrinsic or acquired mechanism of resistance to antimicrobials (Poole, 2005). Multi-drug efflux systems are conserved among bacteria, are usually chromosomally encoded, and in clinical isolates their expression often arises through the inactivation of linked repressor genes (Poole, 2005). In contrast, drug-specific efflux pumps are often encoded on plasmids or other kinds of mobile genetic elements in pathogenic bacteria.

Efflux as mechanism for antibiotic resistance was first described in the early 1980s, when the mechanism of tetracycline resistance was discovered (McMurry *et al*., 1980; Ball *et al*., 1980). Since then, numerous chromosomal and plasmid-encoded efflux mechanisms has been identified and characterised. Bacterial efflux systems that are capable of transporting antimicrobials can be divided into different classes: the major facilitator superfamily (MF), the ATP-binding cassette (ABC) family, the resistance-nodulation-division family (RND), the small multi-drug resistance (SMR) family and the multi-drug and toxic compound extrusion (MATE) family (Poole, 2005) (**Fig. 1.1**). With the exception of the ABC family, all these types of efflux systems function as secondary transporters i.e. they are antiporters using a gradient of ions (H^+ or Na^+), rather than the hydrolysis of ATP, to drive the export of drugs (Poole, 2005).

1.3 DNA topoisomerases

DNA topoisomerases are enzymes that regulate the topological state of DNA.

Regulated DNA topology is essential for the following biological processes: genome compaction, DNA replication, gene expression, transcription, recombination and thermal stability of DNA. The vital role of topoisomerases in bacterial physiology is also reflected in the fact that topoisomerases are an established target for antibacterials and chemotherapeutics.

DNA topoisomerase can catalyse supercoiling/relaxing, knotting/unknotting and catenating/decatenating reactions (Bates and Maxwell, 2005). The majority of DNA topoisomerases can relax negatively and positively supercoiled DNA. However, only DNA gyrase can use ATP to introduce negative supercoils into DNA, since this reaction is unfavourable without the input of energy (Schoeffler and Berger, 2008).

DNA topoisomerases can be divided into 5 different classes: IA/IB/IC/IIA/IIB (Schoeffler and Berger, 2008). Type I topoisomerases work via the introduction of a

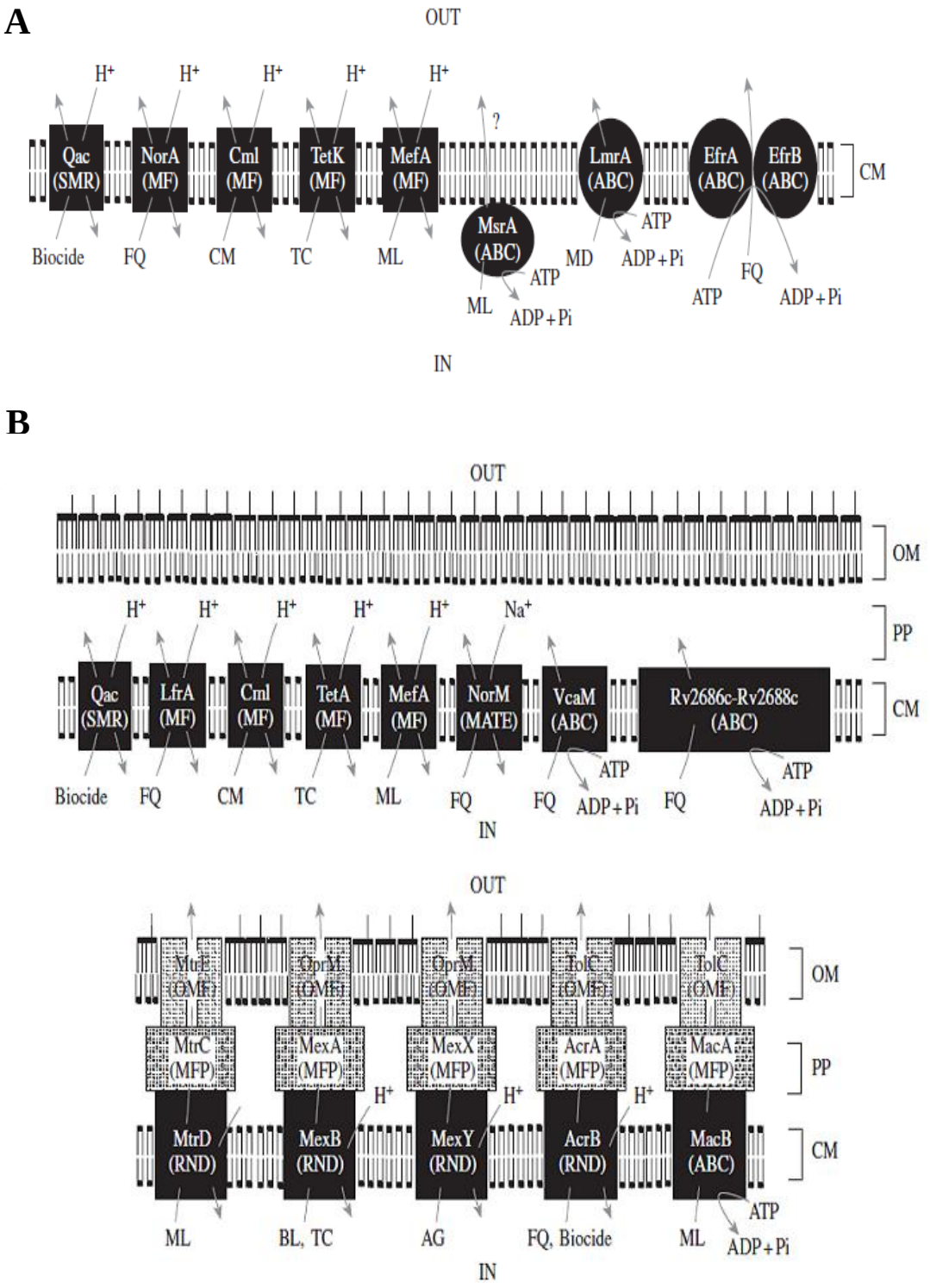


Fig 1.1. Representative drug exporting systems in Gram-positive bacteria (A) and in Gram-negative bacteria (B), showing different families of pumps involved in resistance to antibiotics. FQ: fluoroquinolones; CM: chloramphenicol; TC: tetracycline; ML: macrolides; MD: multi-drugs; BL: beta-lactam. OM: outer membrane; PP: periplasm; CM: cell membrane (Poole, 2005).

single-stranded break into the DNA, whereas Type II topoisomerases work via the introduction of a double-stranded break. Topoisomerases can be further divided into type IA/IB/IC and type IIA/IIB depending on their structures and mechanisms [reviewed in Schoeffler and Berger, 2008).

1.3.1 DNA gyrase

DNA gyrase belongs to the type IIA class of topoisomerases and is capable of using energy derived from ATP hydrolysis to drive the introduction of negative supercoils into DNA (Gellert *et al*., 1976). The other topoisomerases of this class are topoisomerase II and topoisomerase IV. Topoisomerases are widely distributed in all kingdoms of life.

DNA gyrase is a hetero-tetramer, having two gyrase A (GyrA) subunits and two gyrase B (GyrB) subunits (**Fig. 1.2**). The crystal structure of the complete hetero-tetramer has not been determined, however crystal structures of several subunits and constituent domains of the *Escherichia coli* enzyme are available (Brino *et al*., 2000; Fu *et al*., 2009; Morais Cabral *et al*., 1997; Ruthenburg *et al*., 2005). The N-terminal domain of the GyrA subunit in its dimeric state forms the DNA-binding saddle. This is known as the "DNA breakage-reunion domain". Within this domain, there is a 30 Å cavity which is sufficient in volume to accommodate a DNA duplex (Berger, 1998). This is located in between the primary and secondary dimerisation interface. The C-terminal domain of GyrA forms a β-pinwheel fold (similar to the β-propeller fold) (Corbett *et al*., 2004) and shows a basic surface on the outside of the structure that is probably involved in DNA wrapping (Corbettt *et al*., 2004). The N-terminal domain of GyrB forms a dimer that has a central cavity of 20 Å in diameter, which is also large enough to accommodate a DNA duplex (Wigley *et al*., 1991). The N-terminal domain of GyrB contains the GHKL ATPase motif. The structure of the C-terminal domain of GyrB of *Mycobacterium tuberculosis* DNA gyrase has also been solved (Fu *et al*., 2009). *E. coli* GyrB has a 170-amino acid insertion in this domain in comparison to other type IIA topoisomerases and this insertion has been shown to be important for DNA binding in *E. coli* DNA gyrase (Chatterji *et al*., 2000).

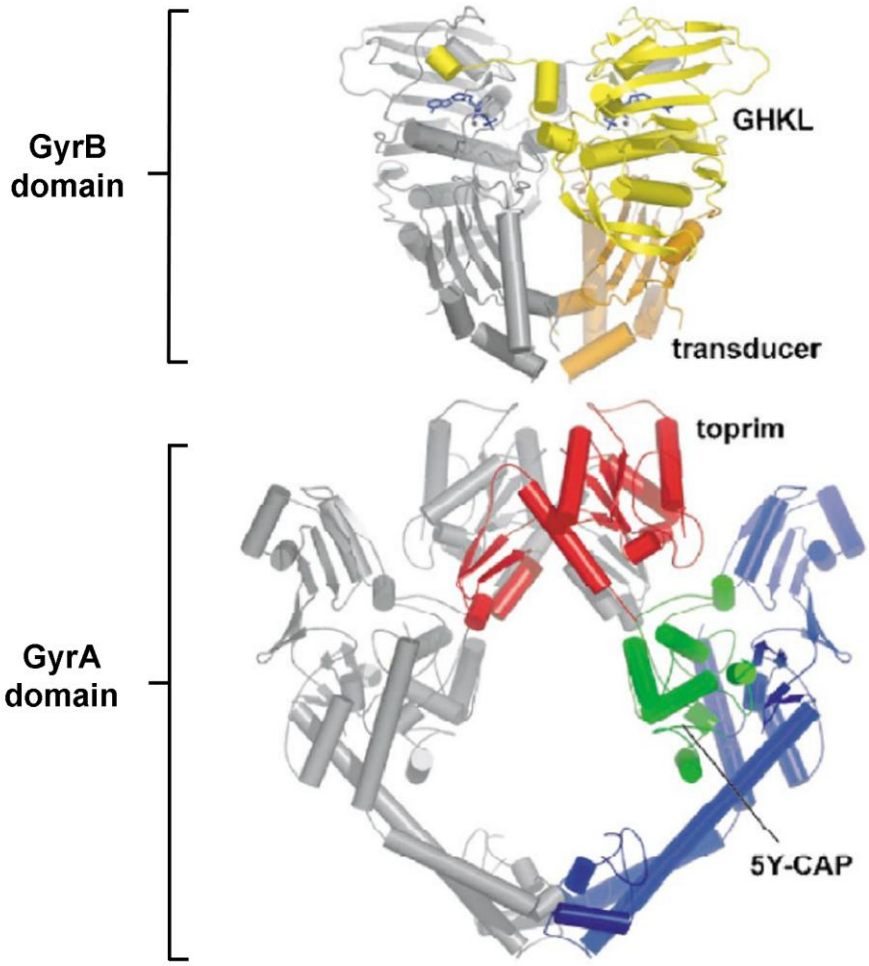


Fig. 1.2. General domain organisation of type IIA topoisomerases, of which DNA gyrase is a member. Conserved domains are highlighted in colour.

1.3.2 The mechanism of action of DNA gyrase

The first step in the mechanism of DNA gyrase is the introduction of a double-strand break, through which the second strand of a DNA duplex is passed (**Fig. 1.3**). This action introduces negative supercoiling into the circular DNA. The first segment of DNA where the double-strand break is introduced is termed the gate segment or G-segment. The double strand break is introduced by nucleophilic attack on the DNA phosphate backbone and the subsequent formation of a covalent bond between the active site Tyr in each monomer with the 5'-phosphate backbone (Roca and Wang, 1992; Roca, 1995). The next step is the separation of the broken G-segment and the passage of the intact DNA segment through the gap (Roca and Wang, 1992). The segment to be transferred is termed the transport segment or T-segment. The passage of the T-segment is achieved through the conformational changes induced by ATP binding. These conformational changes are transmitted throughout gyrase, enabling both the capture of the T-segment through the closing of the N-terminal gate, the separation of the broken ends of the G-segment, and the closing of the C-terminal gate (Corbett and Berger, 2004; Dong and Berger, 2007). Once the T-segment is transported through the G-segment, the broken ends of the G-segment are re-ligated (Corbett and Berger, 2003). The N-terminal gate then opens to allow the T segment to exit from the enzyme (Corbett and Berger, 2004). The T and G segments can be on the same DNA molecule, resulting in the introduction of negative supercoils, or they can be on different DNA molecules, in which case DNA gyrase action leads to catenation or decantenation of these molecules (Roca, 1995).

1.4 Aminocoumarins

The aminocoumarin family of antibiotics comprises three classical members: novobiocin, chlorobiocin and coumermycin A1 (**Fig. 1.4**) (Berger and Batcho, 1978; Li and Heide, 2006), each produced by different *Streptomyces* species. Novobiocin is produced by *Streptomyces spheroides*, chlorobiocin is produced by *Streptomyces roseochromogenes* and coumermycin by *Streptomyces rishiriensis* (Berger and Batcho, 1978; Li and Heide, 2006). These antibiotics are characterised by their 3-amino-4,7-dihydroxycoumarin moiety (**Fig. 1.4**). The appearance of this moiety in non-streptomycete natural products has

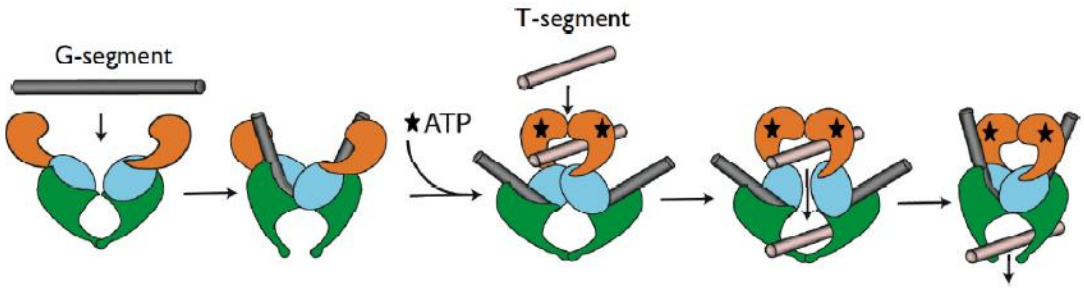


Fig. 1.3. The two gate mechanism of the type IIA class of topoisomerases, of which DNA gyrase is a member (Dong and Berger, 2007). The ATPase domain in shown in orange, the Toprim domain in blue and the DNA-binding domain in green.

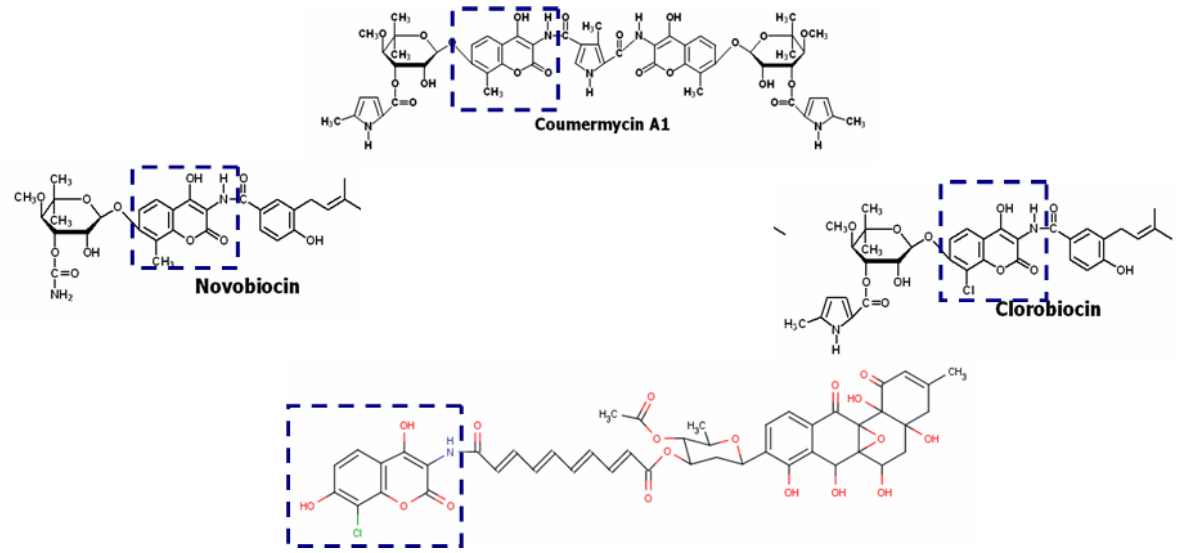


Fig. 1.4. Structures of members of the aminocoumarin class of antibiotics (coumermycin A1, chlorobiocin, novobiocin and simocyclinone D8) with the characteristic aminocoumarin ring boxed (Berger and Batcho, 1978; Schimana *et al.*, 2001).

not been reported. Aminocoumarins are known as potent DNA gyrase inhibitors (Maxwell and Lawson, 2003). Their equilibrium dissociation constants are in the nM range, which is considerably lower than that of clinically important fluoroquinolone antibiotics (Gormley *et al*., 1996). Aminocoumarins are not used clinically, although novobiocin has been licensed for treatment of infections caused by multi-resistant Gram-positive bacteria. Novobiocin, chlorobiocin and coumermycin A1 are competitive inhibitors that compete for the ATP-binding site in GyrB (Lamour *et al*., 2002).

The biosynthetic gene clusters for all three classical aminocoumarins have been cloned and sequenced (Pojer *et al*., 2002; Steffensky *et al*., 2000; Wang *et al*., 2000). The biosynthesis of novobiocin and chlorobiocin are among the best-understood secondary metabolic pathways, and nearly all the steps within these two biosynthetic pathways have been investigated biochemically and genetically (Pojer *et al*., 2002; Steffensky *et al*., 2000; Wang *et al*., 2000).

1.4.1 Self-resistance to aminocoumarin antibiotics

Novobiocin, chlorobiocin and coumermycin A1 are potent inhibitors of DNA gyrase and can potentially kill the producer cells. Thus the producing organisms must have a self-resistance mechanism to protect their own gyrases from the inhibitory effect of these compounds.

The principal self-resistance mechanism in the novobiocin producer, *S. spheroides*, is the *de novo* biosynthesis of a resistant form of DNA gyrase B subunit (Thiara and Cundliffe, 1988, 1989, 1993; Schmutz *et al.*, 2003). The gene encoding the resistant form of DNA gyrase subunit B ($gyrB^R$) is associated with the biosynthetic cluster for novobiocin (at the right border of the cluster), while the gene encoding the sensitive "host" gyrase subunit B ($gyrB^S$) is located elsewhere on the chromosome (**Fig. 1.5**). Thus, the novobiocin-producing organism has a constitutively expressed sensitive $gyrB^S$ gene, and a resistant $gyrB^R$ gene that is only expressed when novobiocin biosynthesis is activated. The promoter of $gyrB^R$ seems to be regulated by the superhelical density of the DNA (Thiara and Cundliffe, 1988, 1989, 1993; Schmutz

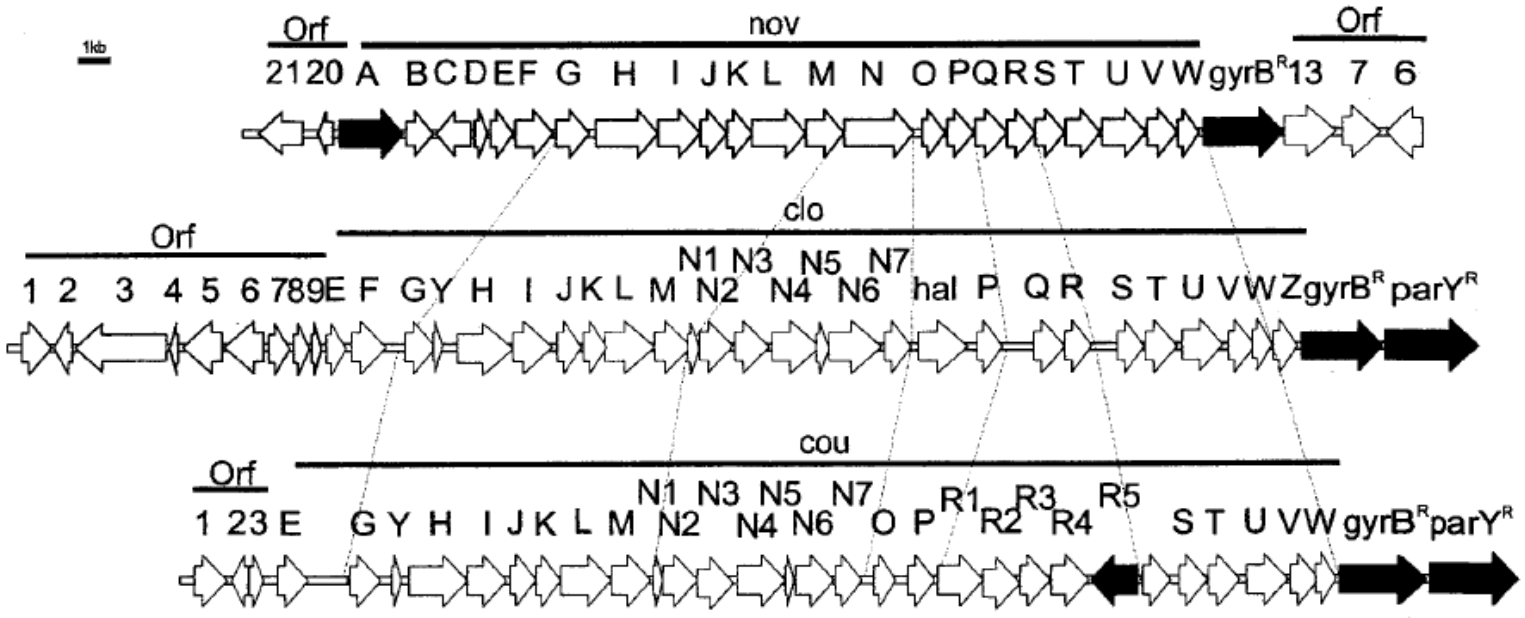


Fig. 1.5. Biosynthetic clusters for novobiocin (top), clorobiocin (middle) and coumermycin A1 (bottom) with genes involved in self-resistance coloured black (Schmutz *et al.*, 2003).

et al., 2003). This might constitute a feed-back mechanism to turn on self-resistance when the organism in exposed to novobiocin. The resistant $GyrB^R$ protein replaces the sensitive $GyrR^S$ subunit in the active GyrA-GyrB heterotetramer giving an enzyme that is no longer sensitive to novobiocin.

On the left border of the novobiocin cluster there is *novA*, encoding an ABC transporter suggested to be involved in transport of, and possibly resistance to, novobiocin (**Fig. 1.5**) (Schmutz *et al.*, 2003). After the biosynthetic clusters for chlorobiocin and coumermycin A1 were sequenced, the self-resistance mechanisms for these aminocoumarin compounds were also investigated by Schmutz *et al.* (2003). As for novobiocin, these two biosynthetic clusters were found to contain a $gyrB^R$ resistance gene, encoding a resistant $GyrB^R$ subunit (**Fig. 1.5**).

An aminocoumarin-resistant topoisomerase IV subunit, $ParY^R$, encoded within the clorobiocin and coumermycin A_1 biosynthetic clusters (but not the novobiocin biosynthetic cluster) (**Fig. 1.5**), also confers resistance to these antibiotics when introduced into a naïve host, presumably by an analogous mechanism (Schmutz *et al.*, 2003). A putative transporter, CouR5, is also encoded in the biosynthetic cluster for coumermycin A1 (**Fig. 1.5**). Expression of this putative transporter in *S. lividans* resulted in moderate levels of resistance against coumermycin A1, suggesting that it might be involved in antibiotic export and possibly in antibiotic resistance (Schmutz *et al.*, 2003).

1.5 Simocyclinones

Simocyclinones represent a new class of topoisomerase inhibitors with a novel mode of action. Simocyclinones are naturally produced by *Streptomyces antibioticus* Tü 6040 (Schimana *et al*., 2000). The main simocyclinone produced by *S. antibioticus* is simocyclinone D8 (SD8) (**Fig. 1.6**) (Schimana *et al*., 2000).

In addition to SD8, there are ~13 other simocyclinone-related compounds that have been isolated from *S. antibioticus*. They represent either variants of the end product

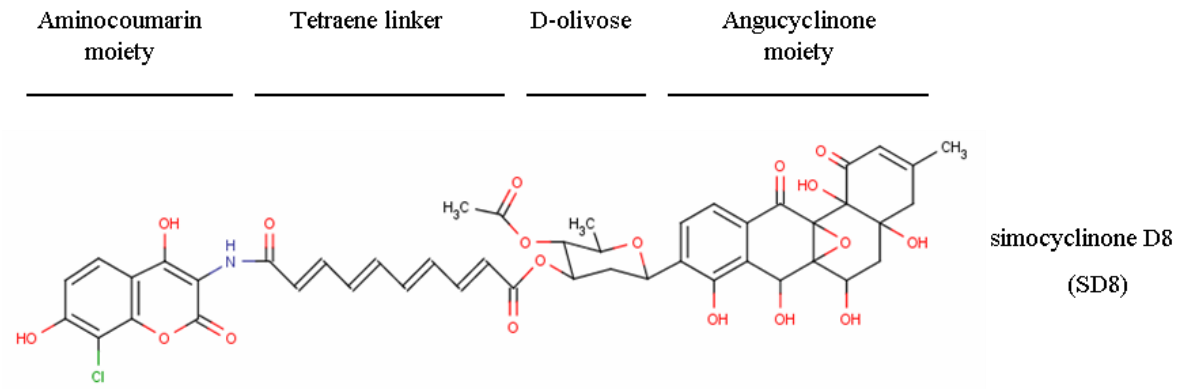
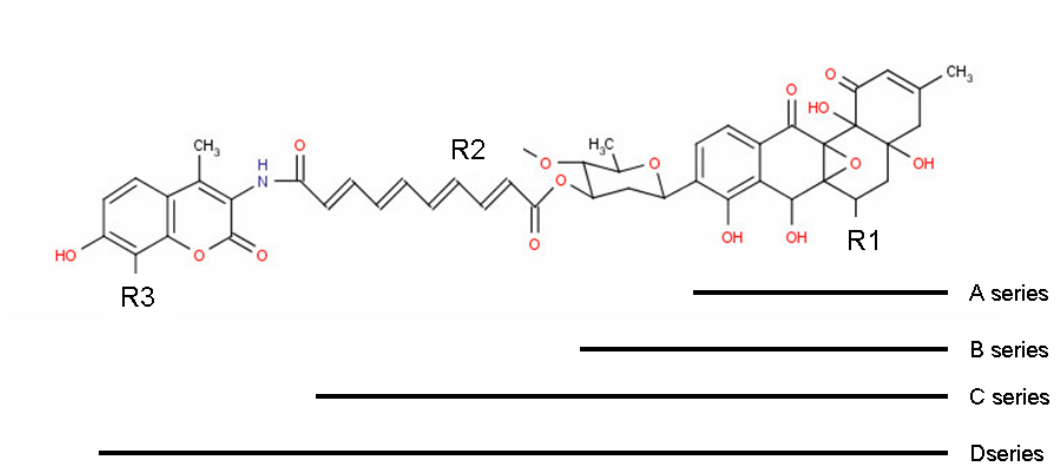


Fig. 1.6. The chemical structure of simocyclinone D8 (SD8) (Holzenkämpfer *et al.*, 2002).



	A1	B2	C2	C4	D6	D4	D8	D7
R1	H	OH	OH	OH	OH	OH	OH	H
R2	-	H	H	CO-CH3	H	CO-CH3	CO-CH3	CO-CH3
R3	-	-	-	-	H	H	Cl	Cl

Fig. 1.7. Simocyclinone-related compounds that have been isolated from *S. antibioticus* Tü 6040. These compounds represent either variants of the end product or naturally occurring biosynthetic intermediates of SD8 (Schimana *et al.*, 2001).

or naturally-occurring biosynthetic intermediates of SD8. These 13 compounds can be classified into four different groups based on their chemical structure (**Fig. 1.7**). The proportion of each simocyclinone produced can be dramatically affected by the media composition and fermentation conditions (Schimana *et al*., 2001).

SD8 has a linear and modular chemical structure. The core structure consists of an aminocoumarin (AC) moiety and an angucyclinone polyketide (PK) moiety linked together by the tetraene dicarboxylic linker and a D-olivose sugar (Holzenkämpfer *et al*., 2002) (**Fig. 1.6**).

The simocyclinones appear to be 'natural hybrid' antibiotics, with moieties related to antibiotics produced by other *Streptomyces* species. The AC ring of SD8 resembles the same moiety in novobiocin, produced by *S. spheroides* (Steffensky *et al*., 2000), the PK moiety is similar to the angucyclic antibiotic aquayamycin, produced by *Streptomyces misawanensis* (Rohr and Thiericke, 1992), and the tetraene linker resembles the conjugated double bond segment of the polyene antibiotic amphotericin, produced by *Streptomyces nodosus* (Caffrey *et al*., 2001).

The hybrid nature of simocyclinone is also reflected in the biosynthetic cluster (**Fig. 1.8**), which appears to result from the integration of four smaller clusters, each responsible for the biosynthesis of one chemical moiety (Galm *et al*., 2002; Trefzer *et al*., 2002). These observations raise interesting questions about the evolution of simocyclinone biosynthesis.

1.5.1 The simocyclinone biosynthetic pathway in *S. antibioticus*

The biosynthetic cluster for simocyclinones (the *sim* cluster) has been cloned and sequenced (**Fig. 1.8**) (Galm *et al*., 2002; Trefzer *et al*., 2002). Analysis of the cluster revealed ~50 open reading frames. They are responsible for the biosynthesis of the four different chemical moieties of SD8, for the export of the simocyclinones and also for the regulation of the biosynthetic pathway (Galm *et al*., 2002; Trefzer *et al*., 2002). Genes putatively responsible for the biosynthesis of each chemical moiety are largely grouped together to form sub-clusters within the total ~84 kb *sim* cluster. In

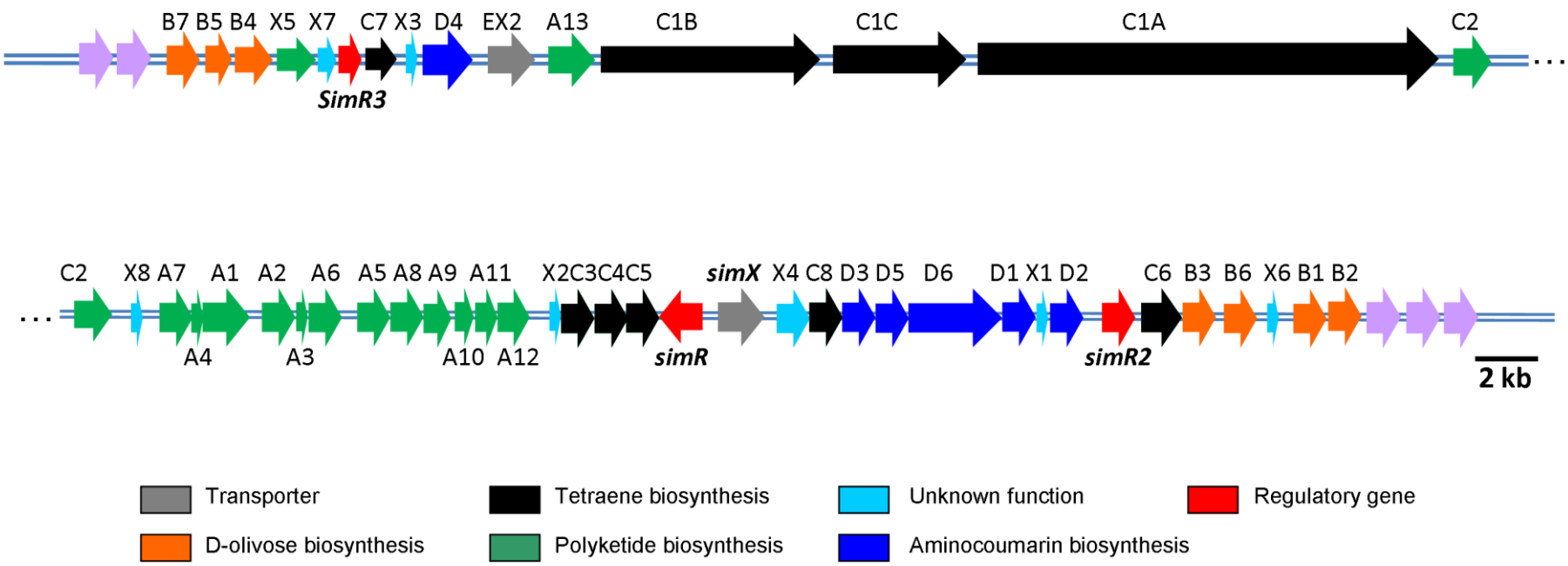


Fig. 1.8. The organisation of the biosynthetic cluster for SD8 (Galm *et al.*, 2002; Trefzer *et al.*, 2002). Analysis of the cluster revealed ~50 open reading frames. The gene products are responsible for the biosynthesis of the four different chemical moieties of SD8, for the export of the simocyclinones and also for the regulation of the *sim* cluster. Genes proposed to be involved in the biosynthesis of each of the four chemical moiety are colour-coded separately, as are genes encoding transporters and regulatory proteins. Genes of unknown function are coloured turquoise.

addition there are several genes with unknown functions. Three regulatory genes are distributed in the *sim* cluster. Based on sequence similarity and knowledge of natural product biosynthesis, a biosynthetic pathway for SD8 has been proposed (**Fig. 1.9**) (Galm *et al*., 2002; Trefzer *et al*., 2002).

Since the *sim* biosynthetic cluster has been cloned and sequenced, it should be possible to generate novel simocyclinones through pathway engineering and by feeding the producer with chemically modified intermediates. These approaches have been successful for generating novel novobiocin-like compounds (Heide, 2009).

1.5.2 The mode of action of simocyclinone

Simocyclinones display inhibitory activity against the growth of Gram-positive bacteria but not against Gram-negative ones. Work described in Chapter 3 suggests that Gram-negative bacteria like *E. coli* are resistant to SD8 because the outer membrane is impermeable to the antibiotic. Oppegard *et al.* (2009) suggested that efflux pumps such as the AcrB multidrug efflux pump may also contribute to the ineffectiveness of SD8 against Gram-negative bacteria. SD8 does not inhibit the growth of yeast or filamentous fungi up to the concentration of 100 µg/ml (Schimana *et al*., 2000).

When the structure of SD8 was first determined, the similarity between the simocyclinones and other aminocoumarin antibiotics was immediately apparent (**Fig. 1.4**). Based on this similarity, SD8 was assayed against *E. coli* DNA gyrase, the target of the previously known aminocoumarin antibiotics, and was shown to be a potent inhibitor (Flatman *et al*., 2005). It was initially thought that SD8 would have the same mode of inhibition as the aminocoumarins, inhibiting the ATPase activity of GyrB. However, Flatman *et al*. (2005) demonstrated that SD8, unlike novobiocin, does not inhibit the ATPase activity of GyrB, which was surprising given that SD8 and novobiocin are both aminocoumarins. However, close inspection of the structure of the GyrB-novobiocin complex showed that the majority of interactions between novobiocin and GyrB involve the noviose sugar of the drug rather that the aminocoumarin ring that is common between novobiocin and SD8 (Lamour *et al*., 2002).

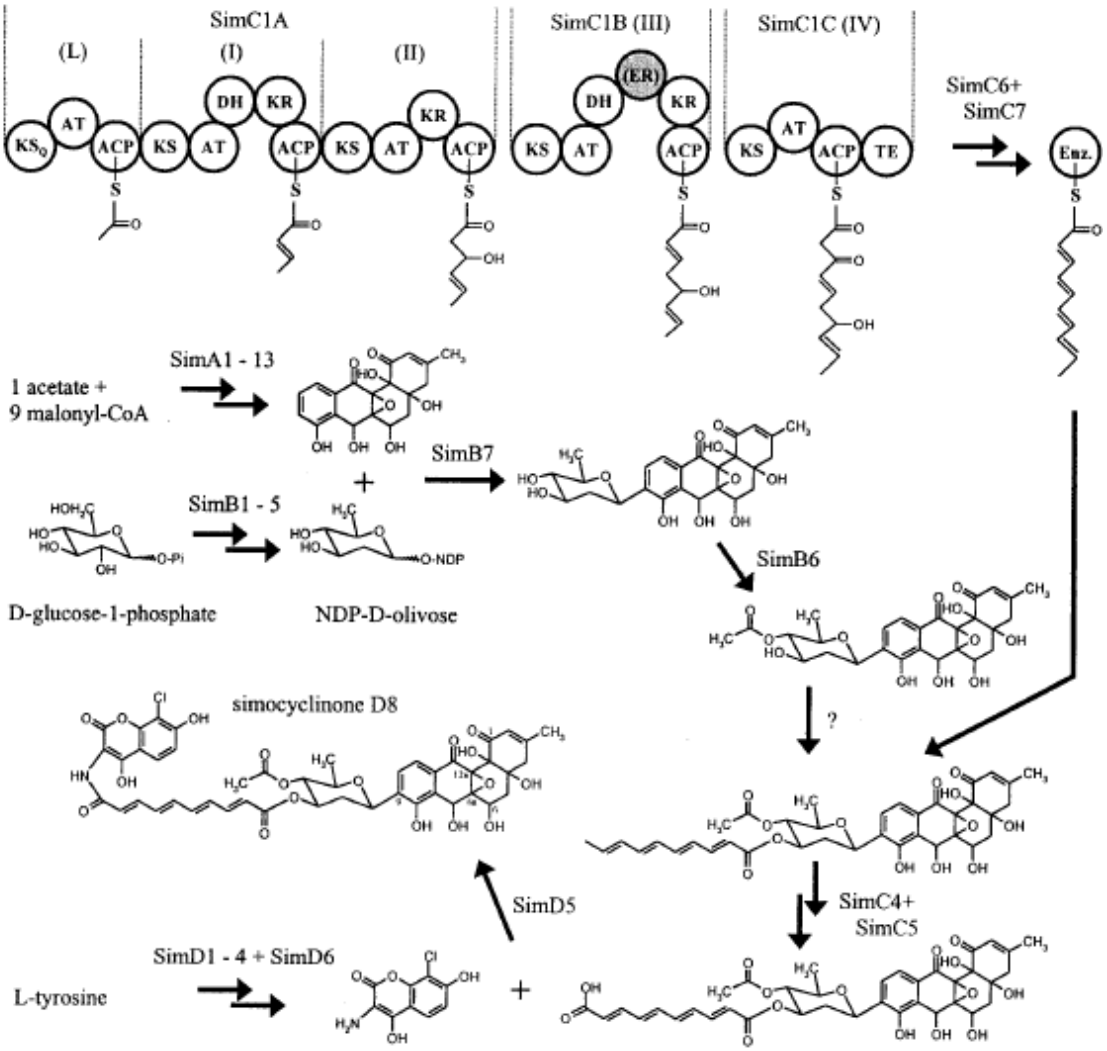


Fig. 1.9. The proposed biosynthetic pathway for SD8 (Galm *et al*., 2002; Trefzer *et al*., 2002). Domain designations: KS: ketosynthase; AT: acyltransferase; ACP: acyl carrier protein; DH: dehydratase; KR: ketoreductase; ER: enoyl reductase; TE: thioesterase. Single enzymatic steps are indicated by a single arrow, and multiple reactions steps are indicated by two arrows.

Flatman *et al*. (2005) subsequently showed by surface plasmon resonance and isothermal titration calorimetry that SD8 has an entirely novel mode of action, binding to the N-terminal domain of the GyrA subunit of DNA gyrase and preventing DNA binding.

Following the work by Flatman *et al*. (2005), a full biochemical dissection of the interaction between the N-terminal of the GyrA and SD8 was reported, including the crystal structure of the complex (Edwards *et al*., 2009). The structure of the complex revealed two distinct binding pockets that separately accommodate the AC and PK moieties of the antibiotic (**Fig. 1.10**). Biochemical studies showed that the individual AC and PK moieties of SD8 are very weak inhibitors of gyrase but, covalently linked, they generate a much more potent inhibitor (Edwards *et al*., 2009).

1.6 The TetR family of transcriptional regulators

The TetR-family transcriptional regulators (TFRs) are named after the founding member of the group, the TetR repressor, which has been thoroughly characterised genetically, biochemically and structurally (Helbl *et al*., 1995; Hillen and Berens, 1994; Hinrichs *et al*., 1994; Kamionka *et al*., 2004; Kisker *et al*., 1995; Orth *et al*., 2000). This transcriptional regulator controls the expression of the TetA efflux pump, which confers resistance to the antibiotic tetracycline (Hillen and Berens, 1994). TFRs are known principally because well-characterized members control genes whose products are involved in antibiotic resistance. Other TFRs regulate genes involved in diverse processes, including the biosynthesis of secondary metabolites (e.g. *Streptomyces coelicolor* CprB) and the pathogenicity of both Gram-negative (e.g. *Vibrio cholerae* HapR) and Gram-positive bacteria (e.g. *Bacillus cereus* HlyIIR) (Ramos *et al*., 2005). TFRs are present in Gram-positive bacteria, α-, β- and γ-proteobacteria, cyanobacteria and archaea, suggesting that TFRs have an ancient origin (Ramos *et al*., 2005). Their numerical abundance and wide taxonomic distribution make them an important family of regulators.

TFRs are found to be particularly abundant in microbes that are exposed to changing environments or to multi-stress insults. This is particularly true for soil-dwelling

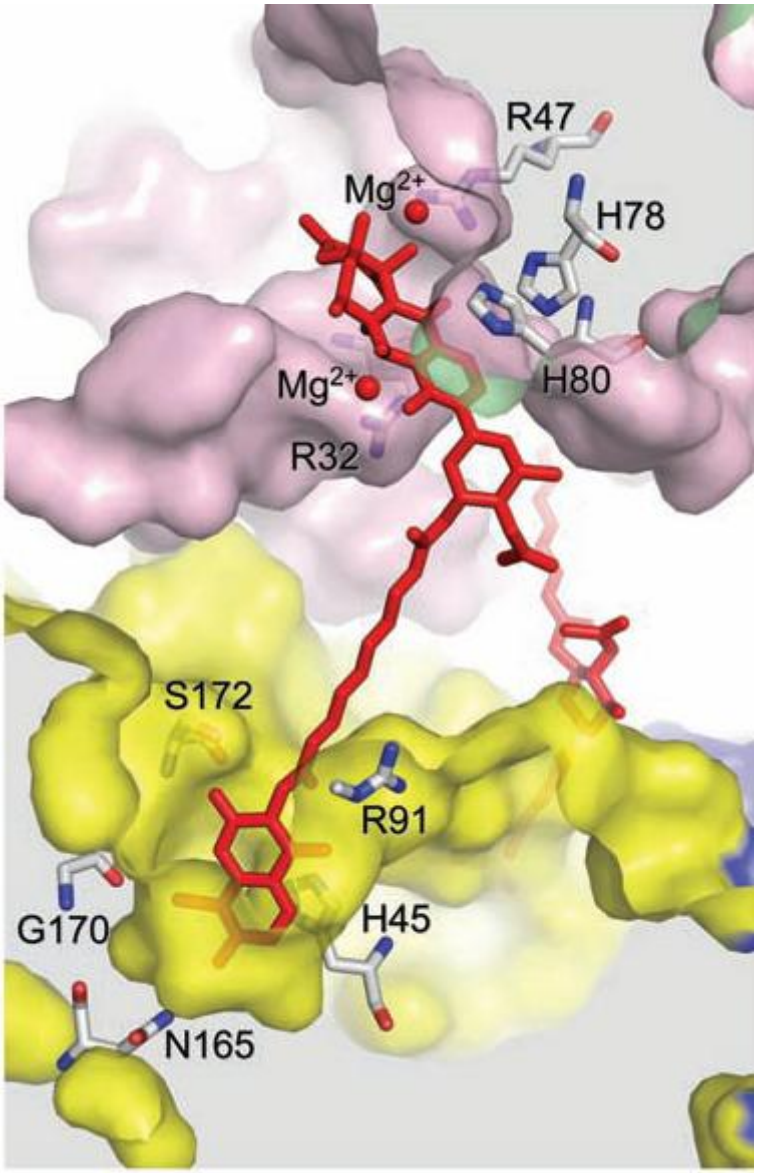


Fig. 1.10. A cross-section through the GyrA-SD8 complex showing the two separate binding pockets for the polyketide (PK) end of SD8 (in purple) and for the aminocoumarin (AC) end (in yellow). The SD8 molecule is in red stick representation. Key residues that lie close to the SD8 molecule are also displayed (Edwards *et al*., 2009).

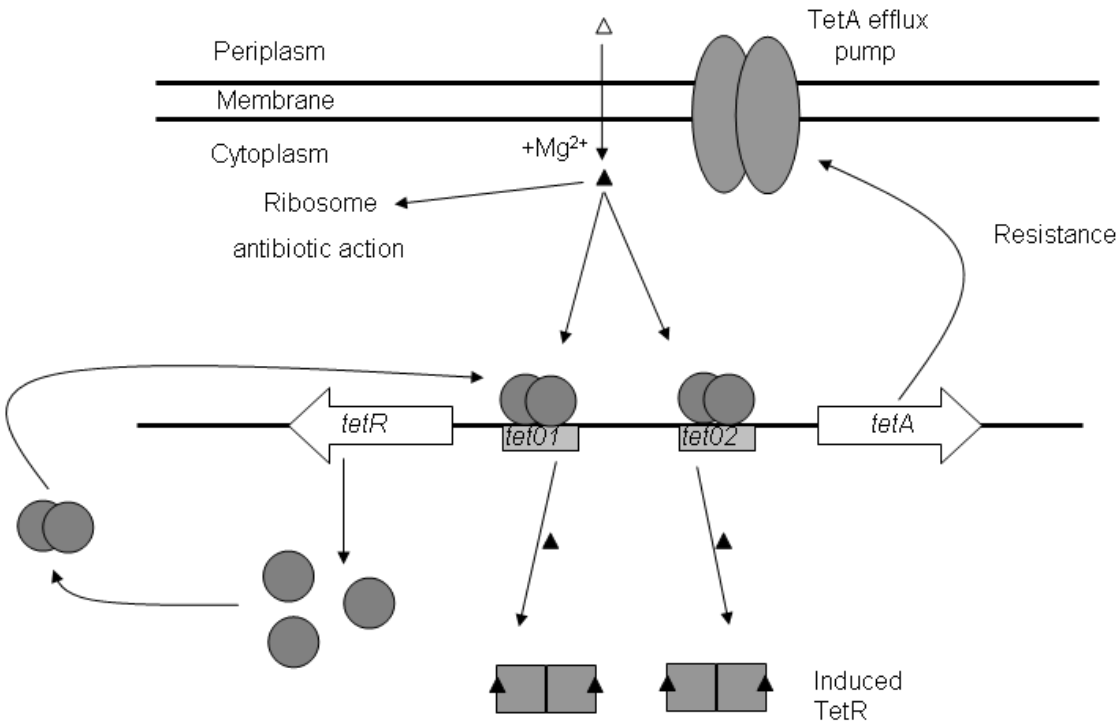
bacteria such as *Streptomyces*, *Bradyrhizhobium* and *Norcadia* (Ramos *et al*., 2005). TFRs are also relatively abundant in plant and animal pathogens (such as *Pseudomonas, Salmonella spp.*) and in extremophiles (such as *Deinococcus*) (Ramos *et al*., 2005). However, TFRs do not appear to be widely used in intracellular pathogens such as *Chlamydia* or *Mycoplasma* (Ramos *et al*., 2005), possibly reflecting the relatively constant life style of those bacteria. It is interesting to note that bacteria which have abundant extracytoplasmic function sigma factors also possess a large number of TFRs (Cases *et al*., 2003; Mahren and Braun, 2003; Martinez-Bueno *et al*., 2002). TFR genes are found not only on the chromosome but also on plasmids and mobile genetic elements, suggesting that horizontal transfer also contributes to the wide-spread distribution of this family (Heuer *et al*., 2004; Szczepanowski *et al*., 2004).

TFRs consist of an N-terminal DNA-binding domain (DBD) and a C-terminal ligand-binding domain (LBD) (Ramos *et al*., 2005). TFRs exhibit a high sequence similarity in their DBDs. In contrast to the conservation of the DBD, the LBDs show little or no sequence similarity (Ramos *et al*., 2005), reflecting the diversity of ligands these domains sense. Despite this, the overall three-dimensional structures of TFRs are strikingly conserved (Ramos *et al*., 2005; Yu *et al*., 2010).

1.6.1 Tetracycline resistance and TetR

Tetracyclines are among the most widely used broad-spectrum bacteriostatic antibiotics (Levy, 1984, 1988). They act by binding to the small subunit of the ribosome, thereby inhibiting polypeptide chain elongation, although the exact details of this mechanism are still unknown. Many Gram-negative bacteria have developed resistance to tetracyclines. The most common mechanism of resistance is the use of an efflux pump protein (TetA) to export the antibiotic out of the cell (**Fig. 1.11A)** (Kaneko *et al*., 1985; Tauch *et al*., 2000; Zhao and Aoki, 1992). Divergently transcribed from *tetA* is *tetR*, whose product functions to control the expression of both genes (George and Levy, 1983; Hillen and Berens, 1994; Hinrichs *et al*., 1994). The intergenic region in between *tetR* and *tetA* houses two TetR operator sequences, separated by 11 bp (Hillen and Berens, 1994) (**Fig. 1.11B**).

A



B

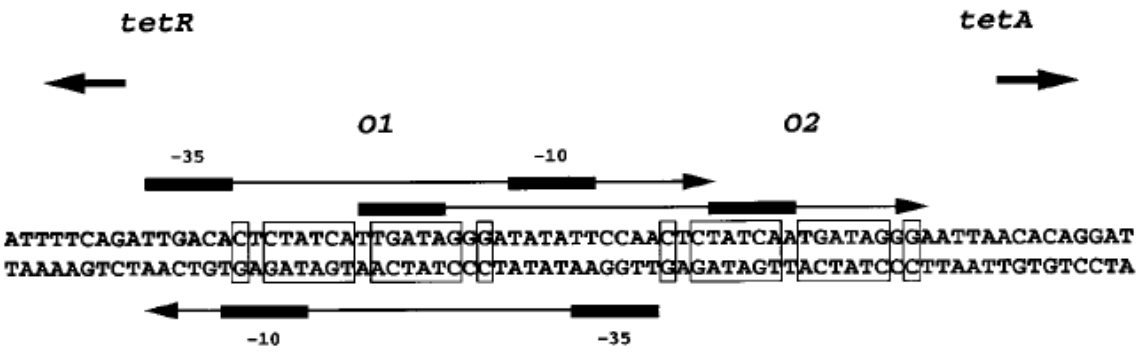


Fig. 1.11. (A) A schematic representation of the tetracycline resistance system specified by *tetAR*. (B) Nucleotide sequence of the *tetAR* intergenic region showing the sequences of the tet_{O1} and tet_{O2} operators (boxed) and the -10 and -35 regions of the *tetA* and *tetR* promoters. The direction of transcription is marked with arrows (Adapted from Saenger *et al*., 2000).

In each operator, there is a core inverted repeat 15 bp long (Hillen and Berens, 1994), with the two half-sites separated by a central base pair (**Fig. 1.11B**). These operators overlap with the core promoters of *tetA* and *tetR* (**Fig. 1.11B**). When the cell is not exposed to tetracyclines, TetR functions as a repressor, binding to both operators and blocking the access of RNA polymerase to the *tetA* and *tetR* promoters (Hillen and Berens, 1994). TetR functions as a homodimer. When the cell is exposed to drug, tetracycline in complex with Mg^{2+} binds to TetR, capturing a conformation of the protein that is unable to bind DNA, thereby allowing RNA polymerase access to the *tetR* and *tetA* promoters (**Fig. 1.11A)** (Hillen and Berens, 1994; Kisker *et al*., 1995; Orth *et al*., 2000; Ramos *et al*., 2005). As a consequence, the TetA efflux pump is produced and exports tetracycline from the cell (**Fig. 1.11A)**. The *tetA* promoter is stronger than the *tetR* promoter, resulting in higher expression of TetA than TetR. As a result of TetA action, the intracellular concentration of tetracycline is reset to a sub-lethal concentration, also resulting in the re-repression of the *tetA* and *tetR* promoters (Hillen and Berens, 1994).

Each monomer within the TetR homodimer has 10 α-helices connected together via turns and loops (**Fig. 1.12**) (Hinrichs *et al*., 1994; Orth *et al*., 2000). TetR is all α-helical, and this is a shared feature among TFRs. The three-dimensional structure of TetR is mainly stabilised by the hydrophobic packing between helices. The DBD consists of helices α1 to α3 with helices α2 and α3 and a connecting loop forming the HTH motif. Helix α1 packs against α2 and α3 in order to stabilise the DBD. Helix α4 connects the DBD to the LBD. The LBD spans from helix α5 to α10. The LBD is also responsible for dimerisation. The diagonal helices α8 and α10 form the dimerisation interface together with their counterparts in the opposing subunit, α8' and α10' (**Fig. 1.12**). Dimerisation is further enhanced by the wrapping arm α9 which embraces the opposing subunit (**Fig. 1.12**) (Hinrichs *et al*., 1994; Orth *et al*., 2000). The LBD houses the tetracycline-binding pocket (**Fig. 1.12**). In the original model for tetracycline-mediated derepression, tetracycline was interpreted to induce (rather

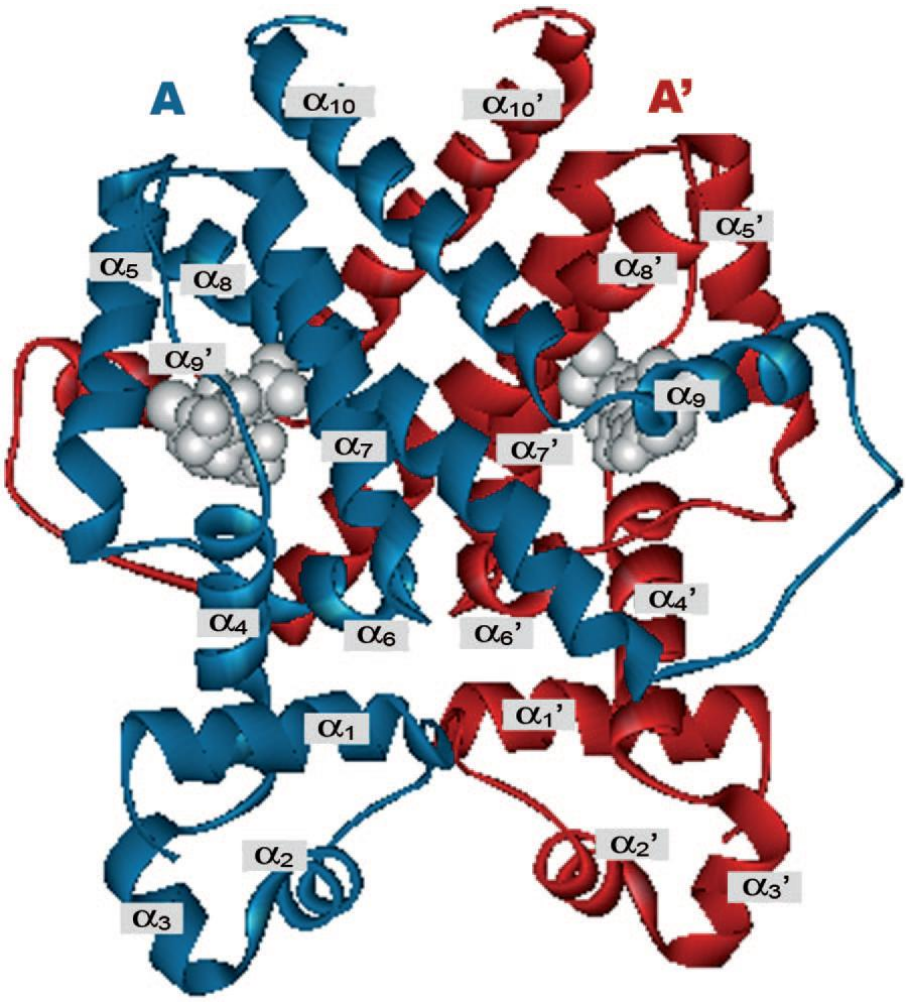


Fig. 1.12. The structure of the TetR homodimer with one subunit shown in blue and one in red (Hinrichs *et al*., 1994). The DBD consists of helices α1 to α3 and the LBD consists of helices α5 to α10, with helix α4 connecting the two domains together. Two tetracycline molecules, one bound to each TetR subunit, are shown in grey (adapted from Ramos *et al*., 2005).

than capture) the observed conformational changes (Hinrichs *et al*., 1994; Orth *et al*., 2000), as follows. The opening of this pocket is controlled by helix α9', by the C-terminal end of helix α8' and by the loop that connects the two helices together (**Fig. 1.12**). The exit of the tetracycline binding pocket is closed by the loop that connects helices α4 and α5. When the complex of tetracycline and Mg^{2+} enters the binding pocket, the distal ring of tetracycline contacts the α4-α5 loop and thereby triggers conformational changes. The magnesium atom contacts His100 and Thr103 on α6 (**Fig. 1.13**). This causes a displacement in α6. This causes the C-terminus of α6 to form a β-turn (**Fig. 1.13**). The α6-α7 loop is also drawn closer to the tetracycline molecule so that Arg104 and Pro105 can establish contact with it (**Fig. 1.13**). The translation of α6 causes helix α4 to move in the same direction due to the van der Waals contacts. Residue His64 of α4, which contacts α5 and tetracycline, acts as a pivot point through which α4 moves as a pendulum (**Fig. 1.14**). Since the DBD is connected to the LBD via α4, the pendulum movement of α4 causes the DBD to rotate with it. As a consequence, the recognition helices α3 and α3' of the HTH motif move further apart, disrupting their interaction with the DNA operator (**Fig. 1.14**) (Hinrichs *et al*., 1994). Upon binding to tetracycline, the binding pocket closes and prevents the drug from escaping the binding cavity. This ligand-bound TetR is no longer able to bind DNA (Hinrichs *et al*., 1994; Orth *et al*., 2000).

However, certain aspects of TetR function were not explained by the originally proposed mechanism of tetracycline-induced derepression. For example, if tetracycline binding was responsible for inducing conformational changes in TetR, thus driving the recognition helices of the HTH motif apart, then one would expect the apo-TetR structure to resemble the structure of TetR in the TetR-DNA complex, but it did not. Moreover, a number of amino acids residues were identified that, when mutated, affected the binding affinity for tetracycline or affected the allosteric response, but were found to reside outside the tetracycline binding pocket or were not involved in conformational changes seen upon tetracycline binding (Hecht *et al*., 1993; Muller *et al*., 1995). Further, there is a group of non-inducible mutant proteins that are not subjected to tetracycline-induced derepression, although they still bind to tetracycline with wild-type affinity (Hecht *et al*., 1993; Muller *et al*., 1995). Finally, there is also a group of TetR mutant proteins that bind to DNA with a higher affinity in the presence of tetracycline than in its absence (Scholz *et al*., 2004). Addressing

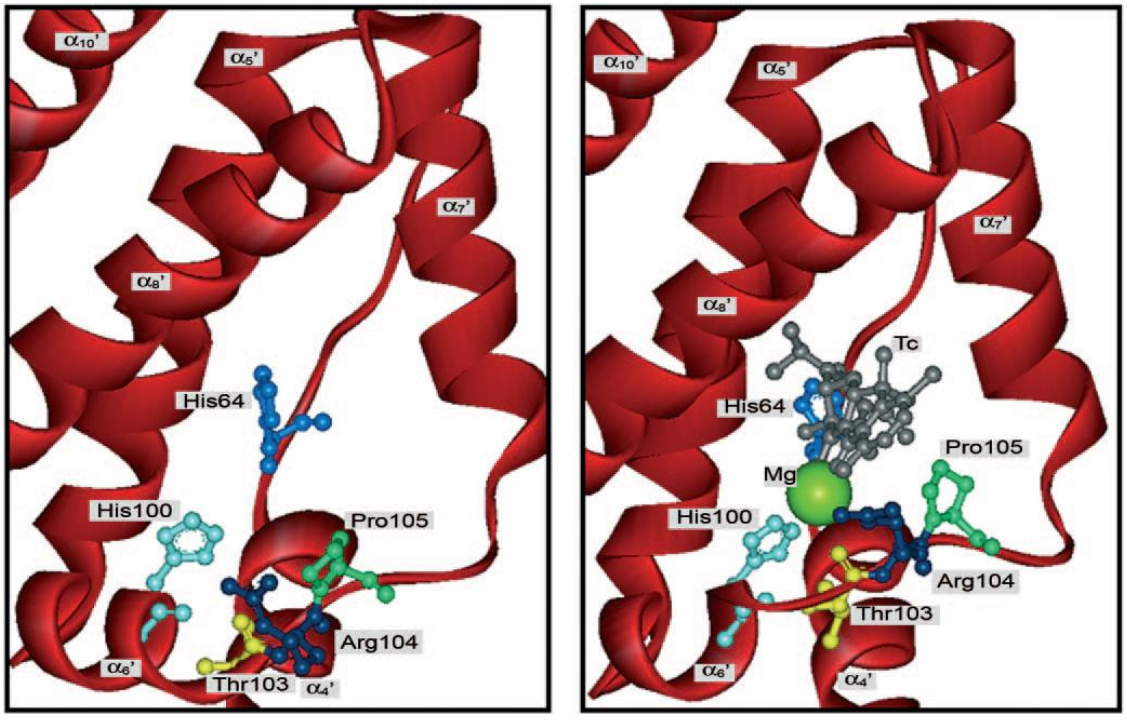


Fig. 1.13. The ligand-binding cavity of TetR in the absence (left panel) and presence of tetracycline-Mg^{2+} (right panel). Relevant residues involved in interactions with Tetracycline-Mg^{2+} are shown in stick representation (adapted from Ramos *et al.*, 2005).

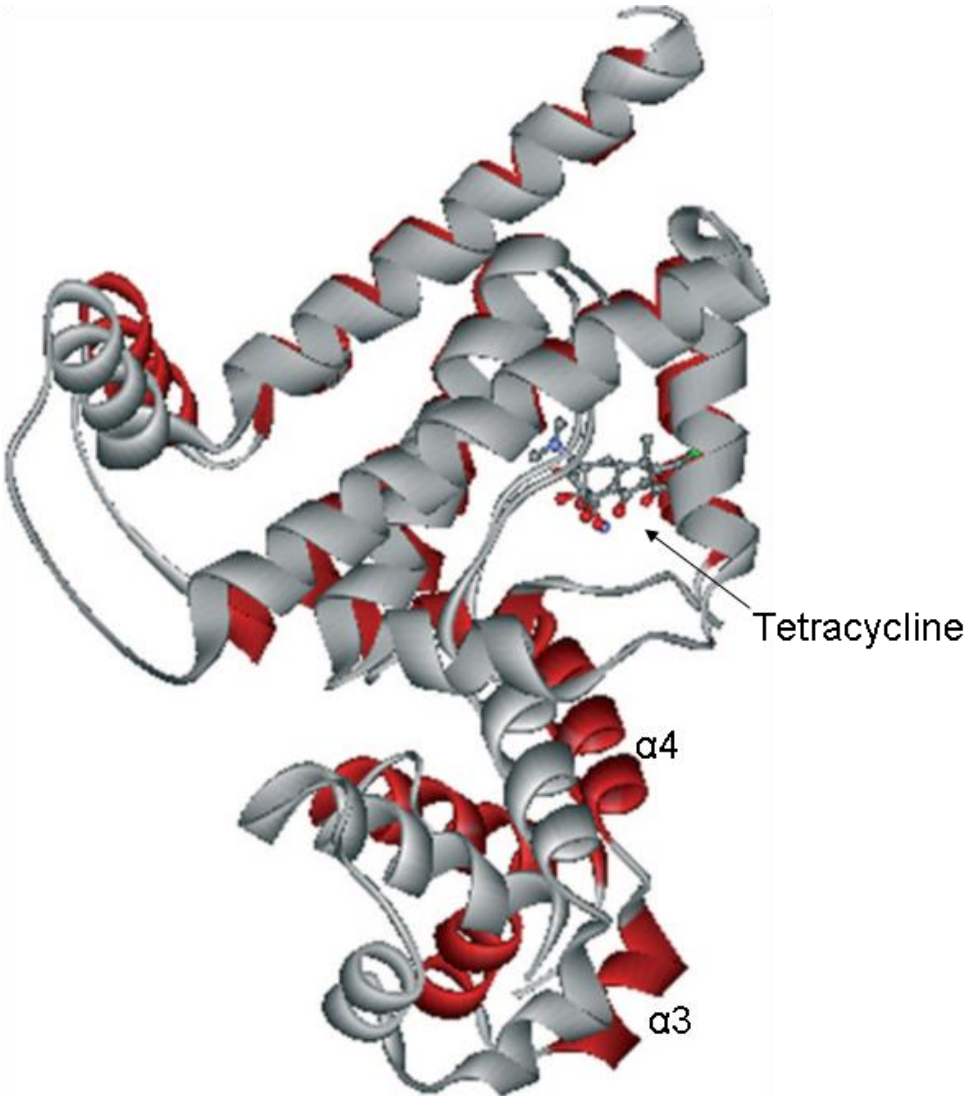


Fig. 1.14. The superimposition of a TetR monomer with (red) and without (grey) tetracycline bound, showing the pendulum-like movement of helix α4, which is translated to the DBD. The increase in distance between the recognition helices α3 and α3' results in the inability of TetR-tetracycline to bind to its DNA operator (adapted from Ramos *et al*., 2005).

these issues, Reichheld *et al.* (2009) performed equilibrium protein unfolding on wild-type TetR and on mutants with altered allosteric behaviours, and then proposed a new model to explain TetR allostery. It was found that the DBD of TetR is relatively flexible and unfolds independently of the LBD in the absence of tetracycline (**Fig. 1.15**) (Reichheld *et al*., 2009). However, in the presence of bound tetracycline, the unfolding of the DBD is coupled to that of the LBD and, as a result, the DBD increases in stability upon ligand-binding (Reichheld *et al*., 2009) (**Fig. 1.15**). On the other hand, the non-inducible TetR mutants displayed less interdomain stabilisation upon tetracycline binding (Reichheld *et al*., 2009). They proposed that the thermodynamic coupling between the LBD and the DBD resulted from tetracycline-induced rigidification of TetR, locking the protein into a conformation which is no longer capable of DNA binding. This rigidification and locking mechanism by bound ligand was proposed to be the cause of the allosteric response in TetR, rather than the originally proposed mechanism in which tetracycline induces conformational changes that are transmitted to the DNA-binding domain (Reichheld *et al*., 2009). In this new model, TetR is thought to be in equilibrium between DNA-binding compatible and incompatible states, and the DNA or antibiotic captures TetR in one state or the other (Reichheld *et al*., 2009). In a recent comprehensive analysis of structural and sequence conservation of TFRs, it was found that for all members where structures of the apo-protein and the protein-ligand complex are available, the distances between the recognition helices are not significantly different between the two structures (and thus both the apo structures and the TFR-ligand structures are incompatible with DNA binding) (**Fig. 1.16**) (Yu *et al*., 2010). The authors suggested that all apo-TetR family members probably sample DNA-binding compatible and incompatible conformations in solution, but crystallisation only traps them in the incompatible conformation. This also suggests that the derepression mechanism by interdomain stabilisation, rigidification and ligand-locking may be widespread in members of this family (Yu *et al*., 2010).

1.6.2 DNA binding by TetR and derepression by tetracycline

The crystal structure of TetR in complex with its DNA operator has been solved, allowing a detailed molecular description of the interaction between protein and DNA (Orth *et al*., 2000). TetR binds perpendicularly to the DNA as a homodimer

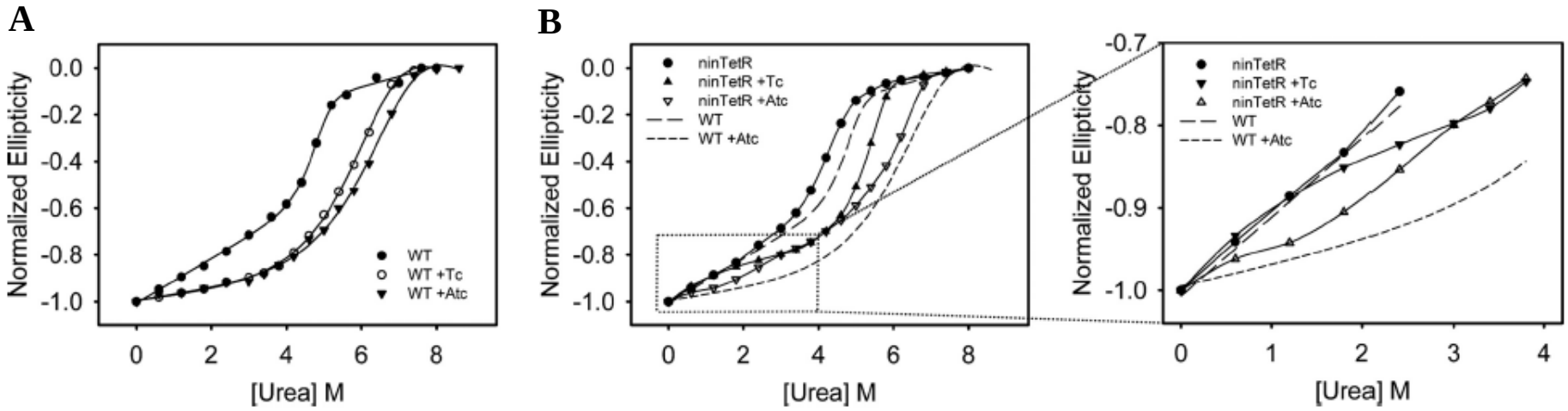


Fig. 1.15. Protein unfolding of wild-type TetR (WT) and of the non-inducible mutant (ninTetR) in the presence or absence of tetracycline (Tc) or anhydrotetracycline (Atc). Urea denaturation profiles of WT (A) and ninTetR (B) in their unliganded form or in the presence of Tc or Atc were monitored by measuring CD ellipticity at increasing concentrations of urea. In (B) a close-up of the pretransition region is shown and the lines denoting the WT profiles from *A* are shown for comparison (Adapted from Reichheld *et al.*, 2009).

Protein name/locus ID	Recognition DNA	Helix α3–helix α3′ (Å)*		
		Apo-	DNA-bound	Ligand-bound
BC5000	AAACTAATnnnATTAGTTT	48.3		
CgmR	GTAACTGTACCGAnnnnTCGTTACAGTTAC	41.7	36.3	41.9
CprB	CGGGAnnnTCCAG	38		
DhaS	GGACACATnnnnnnATTTGTCC	63.4		
DesT	AGTGAACnnnnGTTGACT		36.7	41.9
EF0787	TTTATnAAAAA	38.3		
EthR	TCAACnnnnnGTCGA	41.3		48.1
HlyIIR	TTTAAAnnnnnnnTTTAAA	33.3		
PA3133	AGGGCCATTCCnnnnnGGAATGGCCCT	40.4		
PksA	GCGCACnnnnnnnnGTGCGC	37.9		
PsrA	CAAACAnnTGTTTG	44.1		
QacR	TTATAGACCGATCGATCGGTCTATAA		36.9	44.5
Rha04620	AATCGAAnnnnnTTCGATT	45.1		
Rha06780	TCCTACAnnCGTAGAG	37.7		
RHA1_ro03468	TTGTTTGTnnnACAAACAA	37		
RHA1_ro04212	GGTGAAnnnnnnTTTACC	44.5		
Sco7222	TGGAACGnCGTTCCA	54.2		
Sco7704	CGAACnnGTTCG	37.2		
TetR(D)	TCTATCAnTGATAGA	38.8	34.7	37.4
YbiH	TTAATCAAnnnnTTGATTAA	36		
Ycdc	AACGGTnnnnACAGTT	45.7		
YsiA/Yer0	AATGAATnnnnATTCATT			41.5

* Distances are measured between Cα atoms on each monomer of the positions equivalent to Tyr42 in TetR.

Fig. 1.16 Operator sequences and distances between recognition helices (helix α3 to helix α3′) for TetR-family members where structures of the apo-form and ligand-bound form are available (Adapted from Yu *et al*., 2010).

(**Fig. 1.17A**) and the two HTH motifs engage on two consecutive major grooves of the 15-bp DNA operator (Orth *et al*., 2000). The detailed interactions between the HTH and the DNA are summarised in **Fig. 1.17B**. One TetR subunit binds to the leading DNA strand from deoxynucleotide number -4 to -7 and also interacts with those on the lagging strand from position +4 to +2. Since the TetR is a homodimer and the DNA operator is itself palindromic, the opposing subunit makes symmetrical contacts to DNA, namely with residues +2 to +4 on the leading strand and -4 to -7 on the lagging strand.

From the crystal structure of the TetR-DNA complex, it is clear that helix α3 is the recognition helix, which is responsible for sequence-specific interactions (Orth *et al*., 2000). In the case of TetR, all the residues in helix α3 except Leu41 interact with DNA. Nevertheless, Leu41 is an important residue, contributing to the hydrophobic core that stabilises the α1, α2 and α3 helix bundle. Among the α3 residues, Thr40 directly contacts the base pairs T(-7) and C(-6) on the leading strand and Tyr43 interacts with T(-7) on the leading strand. Pro39 interacts with both strands, at bases T(-5) and A(-4) of the leading strand and at T (+4) on the lagging strand. Tyr42 makes contact with T(+4) and Gln38 with A(+3) on the lagging strand. Residues outside the recognition helix also contribute to the protein-DNA interaction. Arg28 in helix α2 interacts with G(+2) and Lys48, which lies outside the HTH motif at the N-terminus of α4, also interacts with the DNA operator. This lysine residue is relatively well-conserved between among TFRs, which suggest that the equivalent residue in different TFRs might play the same role in DNA interaction.

The interaction between DNA and the DBD of TetR also leads to conformational changes in the HTH motif and in the DNA (Orth *et al*., 2000). The most apparent change in the DBD is the conversion of α3 into a 3_{10} helical conformation. The overall bending of the DNA in the complex is ~17^{o}, with the DNA kinked away from TetR at position +2 in both operator strands. This is compensated by the bending toward TetR in the area corresponding to positions +3 to +6.

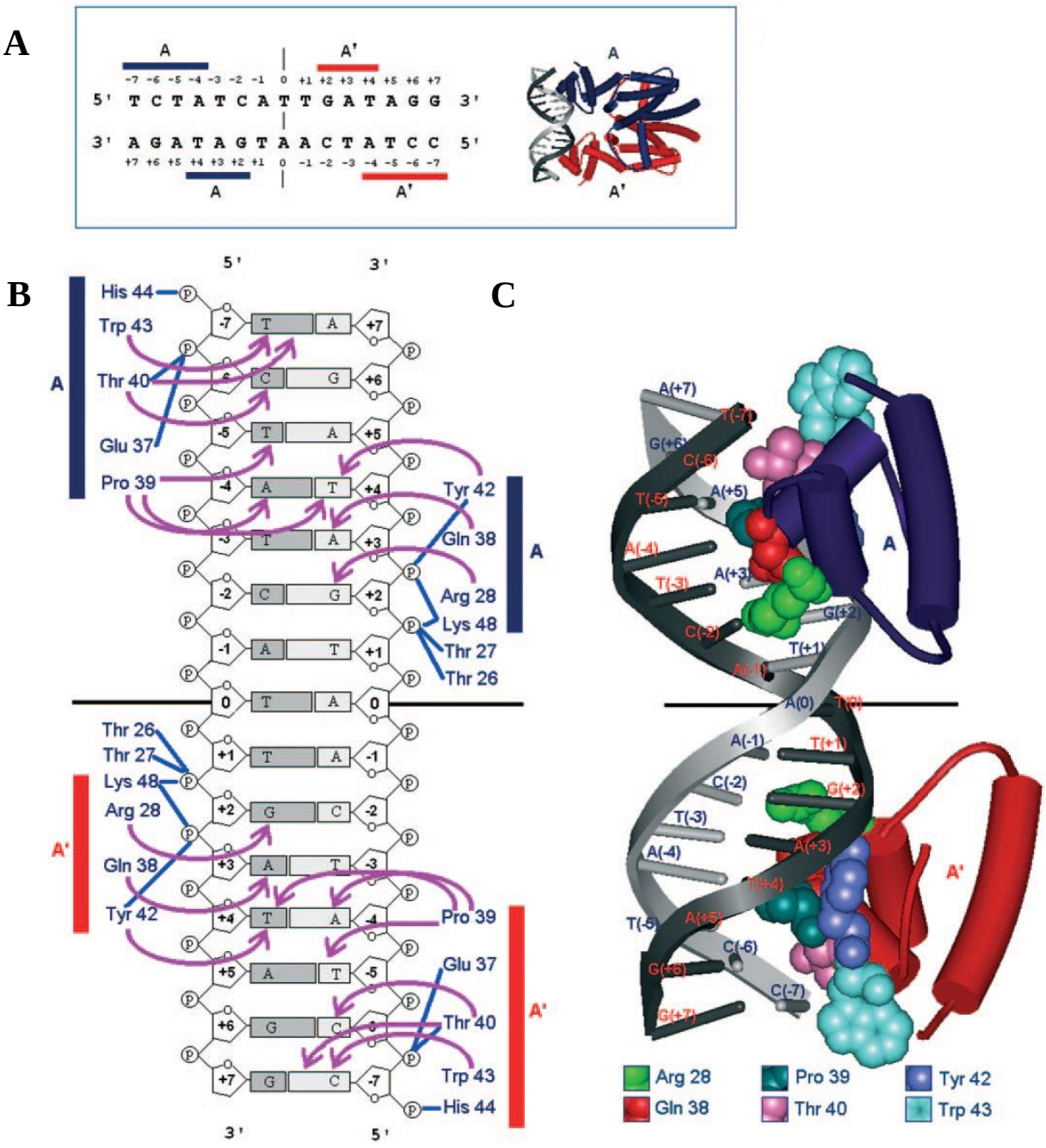


Fig. 1.17. Binding of TetR to its operator site (Orth *et al*., 2000). (A) The *tetR* operator and contacts made by TetR. The *tetR* operator is palindromic. Horizontal bars show nucleotides which are contacted by monomers of the TetR dimer. (B) Interaction of TetR residues with specific nucleotides (arrows) and the phosphate backbone (blue lines) in the operator region. The amino acids involved in DNA binding extend from residues 27 to 48. (C) Representation of each homodimer bound to the *tet* operator in a double-helical representation. The LBD of TetR dimer is omitted for clarity (Orth *et al*., 2000).

1.7 Other members of the TetR family

1.7.1 QacR

QacR from *Staphylococcus aureus* controls the expression of a divergently transcribed gene, *qacA* (Grkovic *et al*., 1998). QacA is an efflux pump which confers resistance to monovalent and bivalent cationic lipophilic antiseptics and disinfectants such as quaternary ammonium compounds (Brown and Skurray, 2001; Paulsen *et al*., 1996). In the absence of the inducers, the QacR protein represses the expression of *qacR* by binding to the two nested inverted repeats that overlap with the *qacR* transcription start site, downstream from the core promoter (**Fig. 1.18A**) (Grkovic *et al*., 1998). QacR seems to inhibit transcription by hindering the transition of the closed RNA polymerase-promoter complex into a productively transcribing complex rather than by blocking the binding of RNA polymerase to DNA, as in the TetR case. The two nested inverted repeats bound by QacR are different from the two separated operators bound by TetR (Grkovic *et al*., 1998).

Structures of QacR, alone, in complex with its inducers and with its DNA operators have been determined, revealing the same domain organisation as in TetR. (Schumacher *et al*., 2001, 2002). The monomers of each dimer have been referred to as proximal or distal, referring to their positions with respect to the *qacA* gene (**Fig. 1.18A**). The operator to which each QacR dimer binds is symmetrical, and partially overlaps with that bound by the other dimer (Schumacher *et al*., 2002). This nested organisation of two operators is different from the TetR case. However, considering one operator at a time, the operator bound by QacR is equivalent to that bound by TetR except for the spacer between the two inverted repeat half-sites. The spacer for the inverted repeat is 1 bp in the case of TetR, while it is 4 bp in the case of QacR. The related architecture of the operators suggests that the binding of TFRs to DNA is similar, leaving aside the number of dimers involved.

In the DNA-binding domain of QacR, the recognition helix, α3, establishes most of the specific contacts with the operator (**Fig. 1.18**) (Schumacher *et al*., 2002). Tyr41 of the distal monomer A establishes a hydrophobic interaction with the base

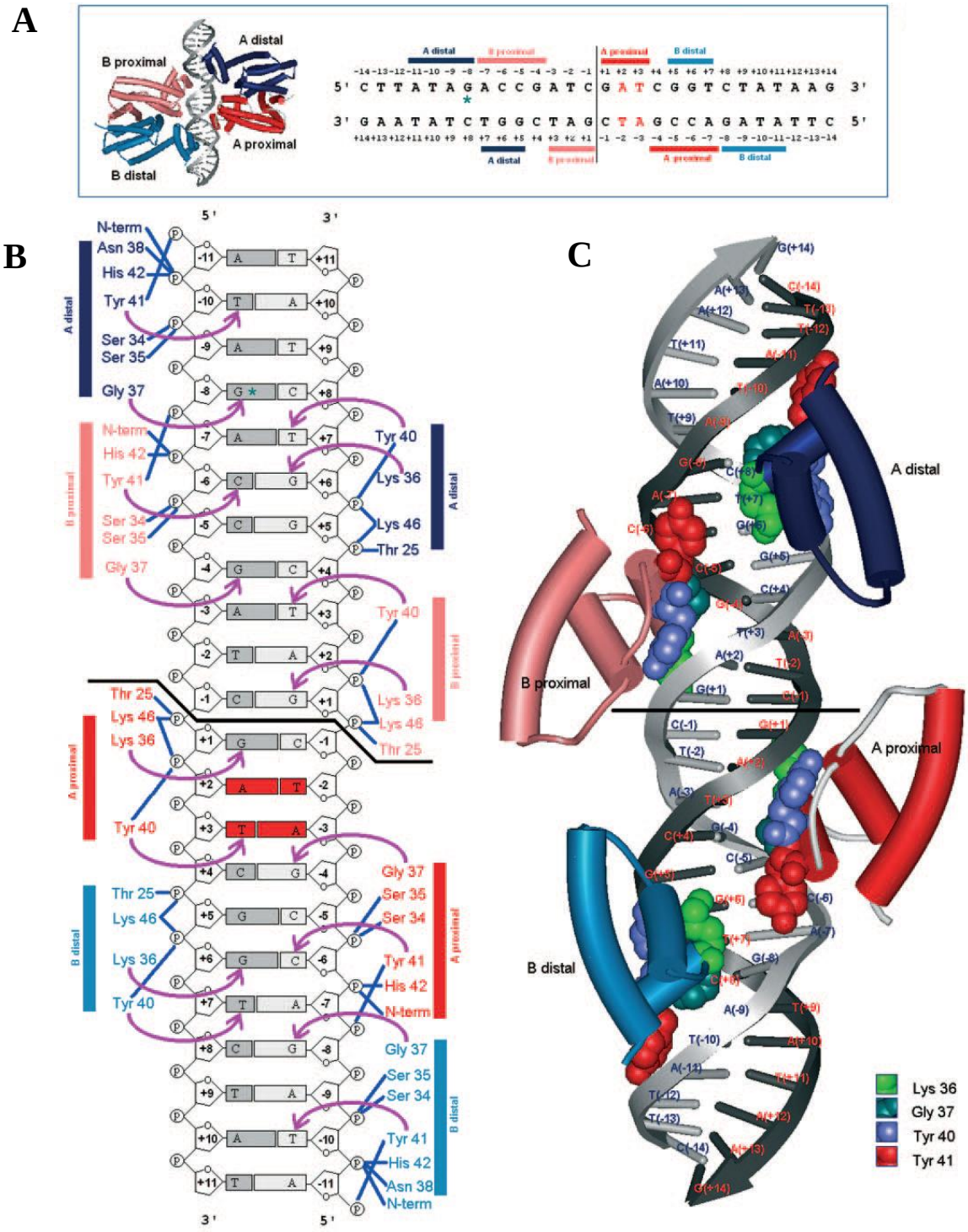


Fig. 1.18. Binding of QacR to its operator site. (A) Interaction of QacR with the *qac* operator. (B) Contacts established by residues of helix α3 of QacR homodimers A and B with specific nucleotides (arrows) and the phosphate backbone (blue lines) in the *qacR* operator. (C) Representation of the two QacR homodimers bound to the *qac* operator in a double-helical representation (adapted from Schumacher *et al*., 2002 and Ramos *et al*., 2005).

T(-10) and with the phosphate group at position (-11) on the leading strand. Tyr40 contacts the T at position +7. Interaction with DNA is facilitated by hydrogen bonds between Lys36 and G(+6) on the lagging strand and between Gly37 and the base of G(-8) on the leading strand. G(-8) is the transcription start site for *qacA* and interaction between Gly37 and this base is considered to be important. The proximal monomer A and B also establish important contacts with DNA. Tyr41 of the proximal B monomer contacts the C(-6) base, while Tyr40 contacts base T(+3) and the phosphate group at position +2. Gly37 in the proximal A monomer contacts G(-4) in the lagging strand, while Lys36 contact G(+1) in the leading strand. A number of residues outside the recognition helix α3 also enhance DNA-binding affinity through interactions with the phosphate groups (Schumacher *et al*., 2002).

Since the two operators are nested in the case of QacR, the two QacR dimers bind on almost opposing faces of the DNA, with the angle between the two dimers only slightly less than 180 degrees (Schumacher *et al*., 2002). QacR binds to DNA cooperatively. The structure of the QacR-DNA complex suggests that this cooperativity does not arise from protein-protein interaction, since the closest distance between the two dimers is 5 Å. Rather, the cooperativity arises from the modification of the bound DNA from a B-DNA conformation to a high-affinity under-twisted conformation, as observed in the crystal of the QacR-DNA complex. This unwinding of the DNA is necessary to allow optimal QacR binding since the distance between the recognition helices in the QacR dimer is 37 Å, greater than the 34 Å between successive major grooves of canonical B-DNA. The untwisting leads to the increase in the distance between successive major grooves required. It is possible that the binding of the first QacR dimer induces the energetically-unfavourable DNA conformational changes that permit the second QacR dimer to bind subsequently (Schumacher *et al*., 2002). While TetR bends the DNA by 17^{o} through two local kinks to optimise for the distance between their recognition helices, QacR bends the bound DNA smoothly and by a total of just 3^{o} (Orth *et al*., 2000; Schumacher *et al*., 2002). Overall, although QacR and TetR are structural similar, detailed analysis shows that they bind DNA in quite different ways.

Unlike TetR, which is extremely specific for tetracyclines, QacR responds to a wide range of inducers (Grkovic *et al*, 1998, Schumacher *et al*., 2001). QacR can be

derepressed by a number of cationic lipophilic compounds such as rhodamine 6G, crystal violet and ethidium, and by bivalent cationic dyes and plant alkaloids. QacR can therefore be considered as a model for multidrug recognition. Also, whereas one tetracycline molecule is bound in each monomer of the TetR dimer, only one inducer molecule is bound per QacR dimer. This was seen in the crystal structure of the QacR-rhodamine 6G complex and was confirmed by isothermal titration calorimetry and equilibrium dialysis (Schumacher *et al*., 2001). The inducer-binding pockets in QacR are extensive, with a volume of ~1,100 A^3, and can be divided into several sub-pockets. Each sub-pocket can accommodate a different inducer. The binding pocket does not require a sophisticated recognition mechanism to bind inducers. The entry to the binding pocket is through a small opening formed by helices α6, α7, α8 and α8'. Since only one monomer in the QacR dimer binds the drug, the structural changes are asymmetrical, occurring only in the drug-bound monomer. Superimposition of the QacR-ligand and QacR-DNA structures reveals the coil-to-helix transition of residues 89 to 93 in helix α5, which extends this helix by another turn. This conformational change forces the drug-surrogate Tyr92 and Tyr93 from the hydrophobic core of the protein. The removal of those residues leads to the relocation of the nearby helix α6 which becomes tethered to the DBD. As a result, the DBD translates by 9 Å and rotates by 37^o, thereby rendering the structure of QacR incompatible with DNA binding (Schumacher *et al*., 2001, 2002).

1.7.2 EthR

Ethionamide is a second-line drug, used in the treatment of tuberculosis patients who have already developed resistance to first-line drugs such as isoniazid or rifampin. Ethionamide is a pro-drug that has to be converted into an active compound in a reaction catalysed by EthA, a flavin-containing monooxygenase. EthR controls the expression of *ethA*, which is divergently transcribed from *ethR* (Baulard *et al*., 2000; Engohang-Ndong *et al*., 2004). In the absence of ethionamide, EthR binds to the intergenic region in between *ethR* and *ethA* and represses the transcription of *ethA*. The intergenic EthR operator sequence is 55 bp long, significantly longer than the 15-bp binding site for TetR. This suggests that multiple dimers of EthR bind to the overlapping and tandemly-repeated palindromic sequences. Surface plasmon

resonance suggested that four EthR dimers bind to the operator (Engohang-Ndong *et al*., 2004).

The structures of EthR-apo and EthR-drug bound are available (Dover *et al*., 2004, Frenois *et al*., 2004). EthR shares the same domain organisation with TetR (**Fig. 1.19**). The LBD is most similar to QacR since EthR and QacR do not possess the wrapping arm that embraces the opposite monomer. The dimerisation interface is principally formed through the diagonal pairs of helices α8, α9 and α8', α9' (**Fig. 1.19**). The most striking feature of EthR is its ligand-binding pocket (**Fig. 1.19)**. The helices α4, α5, α7 and α8 in each monomer form a narrow tunnel-like passage that opens from the bottom to the top of each EthR monomer (Frenois *et al*., 2004). The tunnel is 20 Å long and is lined predominantly by aromatic residues. When the crystal structure of EthR was solved, a compound was fortuitously found to be bound in this long tunnel-like cavity, which was identified as hexadecyl octanoate (Frenois *et al*., 2004). The binding of hexadecyl octanoate induces a conformational state which is incompatible with DNA binding. This leads to the derepression of the *ethA*, and consequently increases sensitivity of the cell to ethionamide or other thioamides through activation of pro-drug processing.

1.7.3 ActR

Actinorhodin is a blue-pigmented antibiotic produced by *S. coelicolor* that is exported out of the cell by the ActA efflux pump (Hopwood, 2007). Analogous to *tetAR*, expression of *actA* is under control of a repressor, ActR, *actR* and *actA* are divergently transcribed, and ActR binds to the intergenic region and represses transcription of *actA* (Tahlan *et al*., 2007). Actinorhodin itself, or the biosynthetic intermediate (S)-DNPA [4-dihydro-9-hydroxy-1-methyl-10-oxo-3-H-naptho-[2,3-c]-pyran-3-(S)-acetic acid], can cause derepression of ActR, which leads to expression of the efflux pump ActA (**Fig. 1.20A**) (Tahlan *et al*., 2007).

The antibiotic actinorhodin is a hexacyclic polyketide (**Fig. 1.20A**). The biosynthesis of actinorhodin is accomplished through a type II polyketide synthase, resulting first in the tricyclic intermediate S-(DNPA), which does not have antibiotic activity (Keatinge-Clay *et al*., 2004). The tricyclic intermediate is then "dimerised" to form

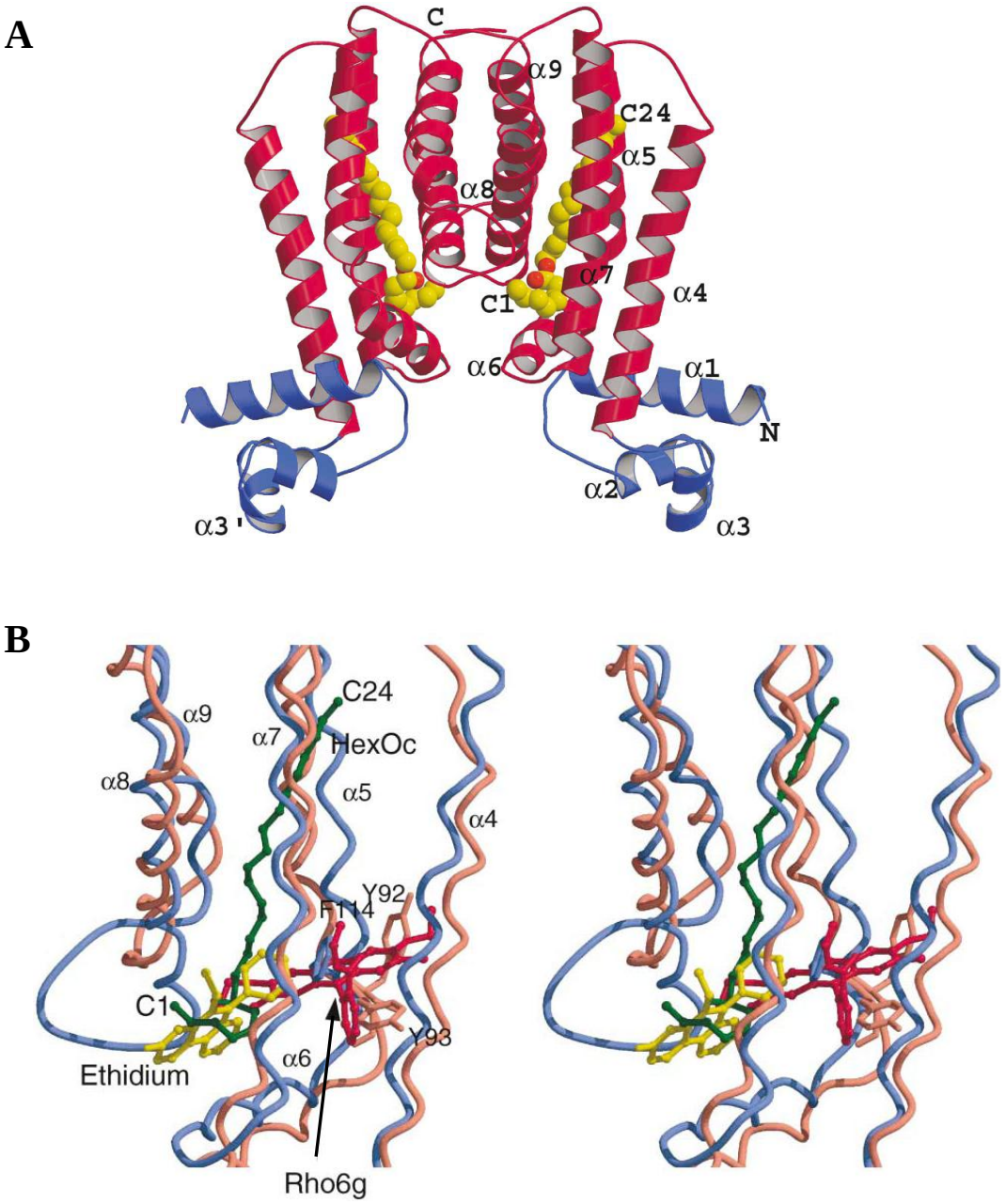


Fig. 1.19. (A) The structure of EthR in complex with a fortuitous ligand, hexadecyl octanoate. The EthR dimer is shown with ribbon representation, with the DBD coloured in blue and the LBD in red. The hexadecyl octanoate ligand, shown in yellow, runs parallel to the 2-fold dimer axis. (B) Superimposition of the binding pocket of EthR (in blue) and that of QacR (in red). Although the two proteins superpose well with each other, they have a divergent binding pocket. The ligand for EthR (green) resides deep in the hydrophobic core of the dimer and runs parallel to the axis of helices α4 and α7. In contrast, the pocket of QacR is closer to the DBD and is parallel to helix α6 instead (adapted from Frenois *et al*., 2004).

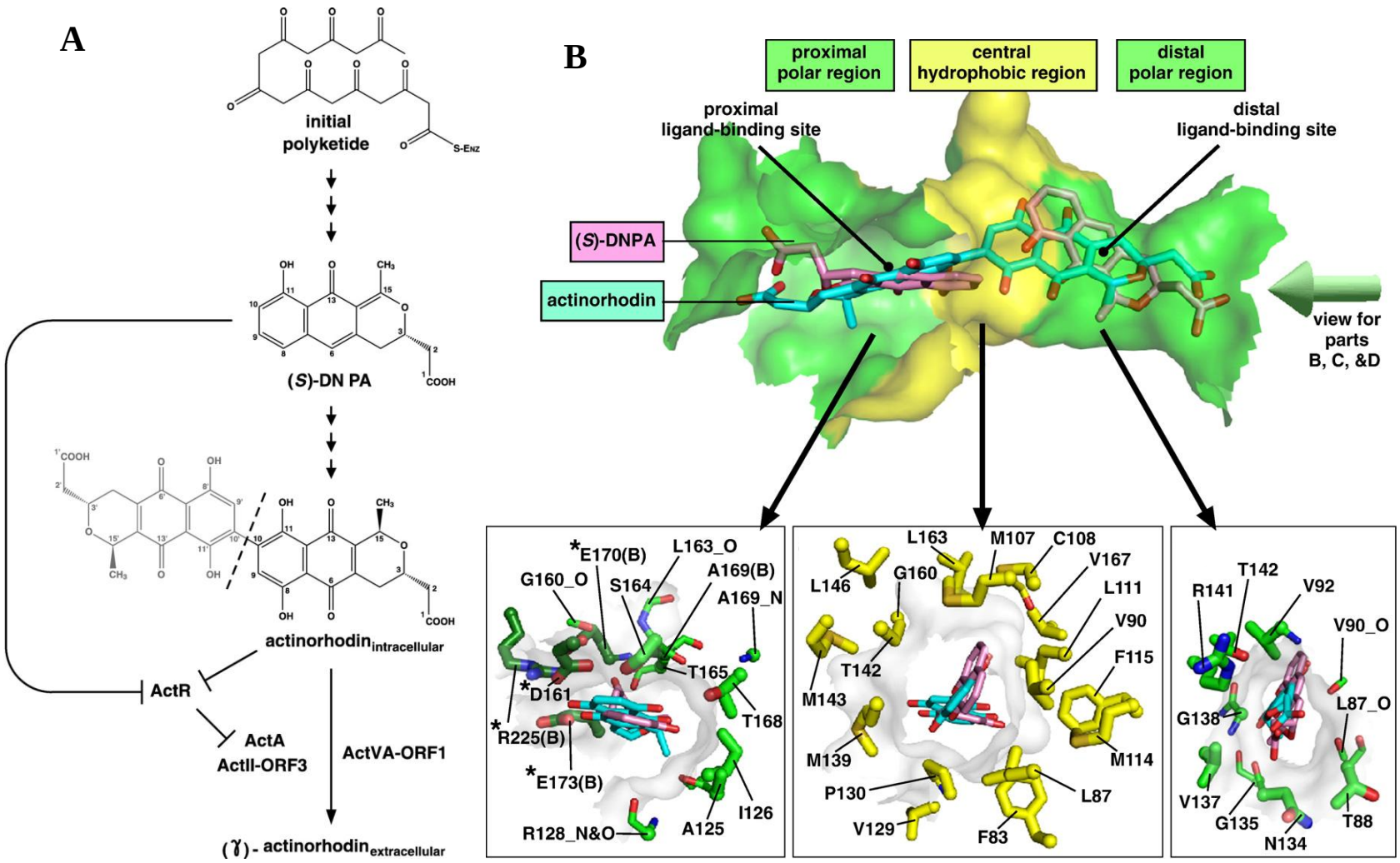


Fig. 1.20 (cont next page)

Fig 1.20. (A) Schematic representation of actinorhodin biosynthesis and export. An initial polyketide is enzymatically transformed into the 3-ring intermediate (S)-DNPA, and two molecules of (S)-DNPA "dimerise" to form the 6-ring final product, actinorhodin, which is then exported from the cell by the ActA efflux pump. Both actinorhodin and the intermediate (S)-DNPA can efficiently derepress ActR. (B) The binding pocket of ActR, showing the alternating, polar-hydrophobic-polar, pattern of regions. The actinorhodin and (S)-DNPA ligands are shown in blue and magenta, respectively. Residues within the polar (green) and hydrophobic (yellow) regions are shown with stick representation (Willems *et al*., 2008).

the mature hexacyclic product, actinorhodin (**Fig. 1.20A**) (Taguchi *et al*., 2000). Recent studies by Tahlan *et al.* (2007) have established that the tricyclic intermediate is more effective in derepressing ActR than the mature antibiotic actinorhodin. The fact that ActR responds to an inactive biosynthetic intermediate potentially allows the producing cells to activate the downstream process (i.e. assembly of the actinorhodin export machinery) before the upstream process (i.e. biosynthesis of mature actinorhodin) has been completed, in a 'feed-forward' mechanism (Tahlan *et al*., 2007). The structures of ActR in complex with actinorhodin or with its tricyclic intermediate explained why the intermediate was active in depression (Willems *et al*., 2008).

The overall structure of ActR is very similar to the classical TetR protein. Hence, only the binding pocket is described in detail here. The binding pocket of ActR has an alternated hydrophilic/hydrophobic/hydrophilic pattern of lining residues (Willems *et al*., 2008). This pocket can accommodate one hexacyclic actinorhodin molecule or two back-to-back tricyclic (S)-DNPA molecules (**Fig. 1.20B**). The bilateral symmetry of the pocket environment correlates well with the symmetry seen in actinorhodin. The central hydrophobic region of actinorhodin (C9, C10, C9' and C10') mirrors the hydrophobic midpoint of the ligand-binding tunnel. In contrast, the more polar oxygen-containing side groups of actinorhodin reside in the more hydrophilic proximal and distal half-sites of the binding tunnel (**Fig. 1.20B**). This structure provides a potential explanation for the potency of the intermediate S-DNPA for derepression of ActR. In the proximal site (close to the tunnel opening), actinorhodin and S-DNPA are reasonably well superimposed. However, at the distal binding site, actinorhodin and S-DNPA occupy distinct positions. It is not physically possible for the distal-half of actinorhodin to adopt the same position as S-DNPA, due to the constraint imposed by the covalent bond that links the two halves of actinorhodin together. It is therefore possible that S-DNPA can interact more favourably with ActR and thereby induce derepression more efficiently.

1.7.4 MphR(A)

Macrolide antibiotics are among the most widely used drugs to combat bacterial infections. Examples of clinically important macrolides include the natural product

erythromycin, produced by the actinomycete *Saccharopolyspora erythraea*, and semisynthetic derivatives of erythromycin such as clarithromycin or azithromycin. Resistance to this family of antibiotics can be achieved through various mechanisms including efflux pump-mediated export and target modification (Leclercq and Courvalin, 1991; Ross *et al*., 1990; Weisblum, 1995). Another resistance strategy is the enzymatic inactivation of the antibiotic by a macrolide 2'-phosphotransferase, encoded by *mphA* (Noguchi *et al*., 1995). MphA phosphorylates the 2'-hydroxyl position on the desosamine sugar of erythromycin (**Fig. 1.21A**), which renders the antibiotic inactive because the 2'-hydroxyl position mediates the key interaction with the ribosome. The expression of *mphA* is under the control of a TetR family member MphR(A). MphR(A) normally binds to the 35-bp operator region upstream of the *mphA* start codon and represses the expression of *mphA* in the absence of erythromycin (Noguchi *et al*., 2000). In the presence of erythromycin, MphR(A) dissociates from the DNA and allows the expression of the MphA phosphotransferase. In contrast to the classical *tetRA* system where the two genes are transcribed divergently, *mphR* lies downstream of *mphA* and they are co-transcribed in an operon (**Fig. 1.21B**). This gene cassette has been isolated in several different clinical and environment sources, appearing to be part of a mobile element that contributes to antibiotic resistance (Szczepanowski *et al*., 2004, 2007).

Again the overall structure of MphR(A) is highly similar to TetR so only the erythromycin-binding pocket of MphR(A) is discussed here in detail. MphR(A) binds one molecule of erythromycin per monomer in a large hydrophobic binding pocket (Zheng *et al*., 2009). The binding pocket is lined by residues from helices α4-α8 and by residues from the dimerisation interface of the opposite monomer (**Fig. 1.21C** and **D**). The binding pocket encloses most of the ligand including the majority of the desosamine and the cladinose sugars. The entrance to the pocket is lined with 17 residues, most of which are hydrophobic. These residues interact extensively with the erythromycin ligand. The interactions are summarised in **Fig. 1.21.D**. Zheng *et al*. (2009) also noted the presence of the ε-nitrogen of Lys21 4.6 Å away from the 2'-hydroxyl of the desosamine sugar, the site of modification by the MphA phosphotransferase. It is possible that the product of MphA action - phosphorylated erythromycin - would be a better ligand for MphR than erythromycin itself, due to a favourable interaction between the 2'-O-phosphate and Lys21. If this were true, it

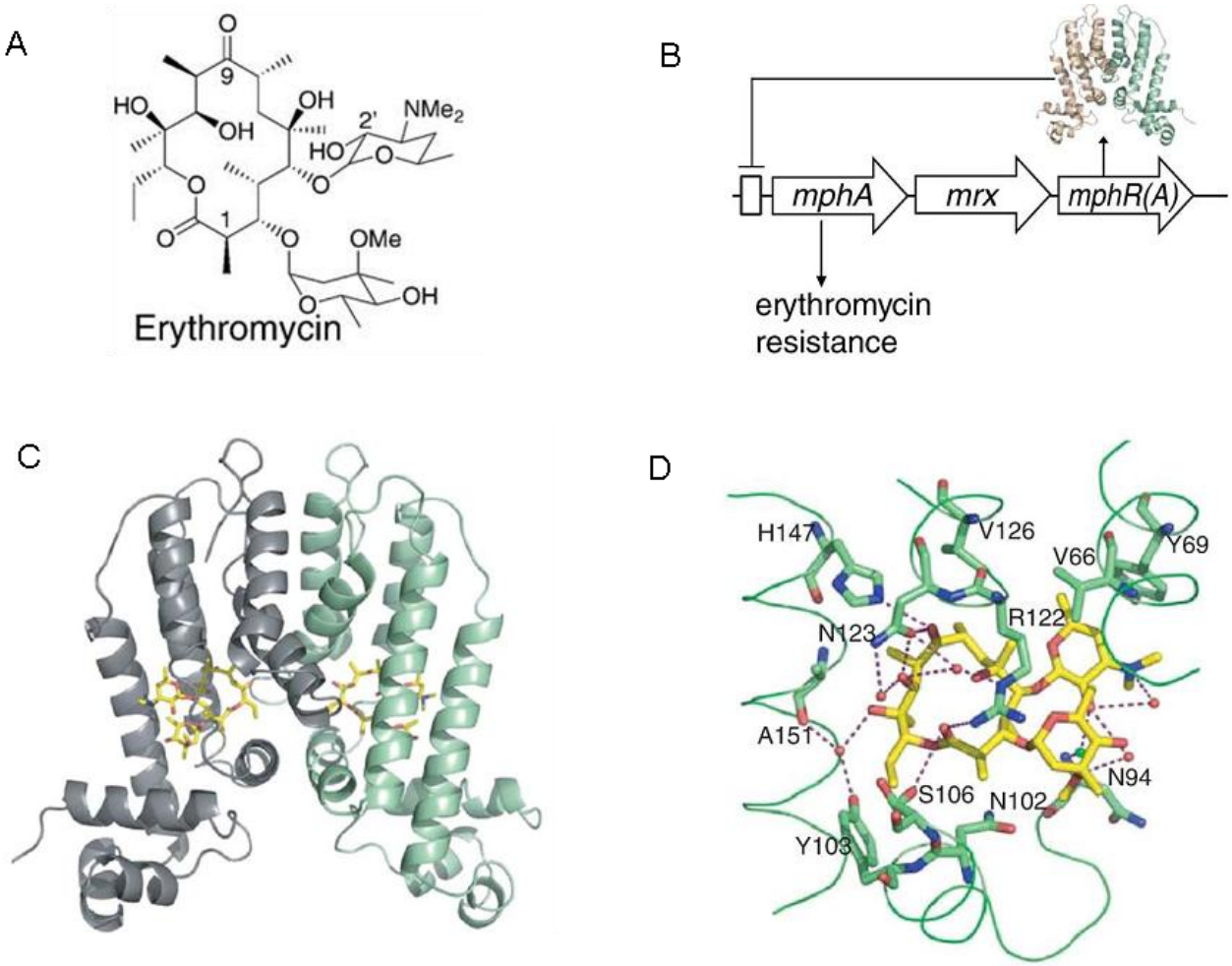


Fig. 1.21 (cont next page)

Fig. 1.21. (A) Chemical structure of erythromycin. Note the 2'-hydroxyl group on the desosamine sugar, the site that is phophorylated by the MphA phosphotransferase enzyme to inactivate the antibiotic. (B) The organisation of the erythromycin resistance cassette is different from that of the *tetR/tetA* cassette. Here, *mphR* lies downstream of *mphA* and they form part of an operon. (C) The structure of the MphR(A) dimer with erythromycin bound. (D) The ligand-binding pocket of MphR(A) with erythromycin bound (adapted from Zheng *et al*., 2009).

would impose positive feed back on the system. The binding of erythromycin closes the ligand-binding pocket, reducing its volume and impeding the release of the bound ligand.

1.7.5 DesT

In the examples above, we have mainly seen the role of TFRs in regulating the export of secondary metabolites. However, TFRs are also widely involved in controlling important primary metabolic processes. An example is DesT, which is involved in the regulation of membrane lipid homeostasis in *Pseudomonas aeruginosa* (Zhu *et al*., 2006). DesT regulates the ratio of unsaturated and saturated fatty acids available for membrane lipid biosynthesis by controlling the expression of the *desCB* operon (Zhu *et al*., 2006). This operon encodes an oxygen-dependent acyl-CoA desaturase that converts saturated fatty acids (16:0 or 18:0-CoA) into Δ9 unsaturated fatty acids (16:1 Δ9- or 18:1 Δ9-CoA). *P. aeruginosa* is able to convert extracellular fatty acids into their CoA thioesters and incorporates them into membrane phospholipids (Zhu *et al*., 2006). The membrane of *P. aeruginosa* normally contains a high proportion of unsaturated fatty acids and therefore the incorporation of an abundant extracellular saturated fatty acid could potentially disturb membrane homeostasis. The regulated expression of *desCB* serves to modulate membrane lipid composition to avoid this potential problem. The CoA thioesters of saturated fatty acids induce DesT dissociation from the DNA, and hence the expression of *desBC* (Zhang *et al*., 2007). In contrast, the CoA thioesters of unsaturated fatty acids promote association of DesT to the DNA (Zhang *et al*., 2007). Since DesT binds saturated and unsaturated fatty acid-CoA with equal affinity, it has been postulated that DesT senses the ratio of saturated and unsaturated fatty acid thioesters, rather than absolute concentrations (Zhang *et al*., 2007). Two structures of DesT have been solved, one in complex with an unsaturated fatty acid-CoA and DNA, the other in complex with a saturated fatty acid-CoA. These structures elucidated the DNA-binding mechanism and also explained how the protein could differentiate between ligands that differ only by the presence or absence of a double bond, and why they have opposing effects on DNA binding (Miller *et al*., 2010).

The overall structure of DesT is most similar to the structures of QacR and EthR (**Fig. 1.22**) (Miller *et al*., 2010). Unlike TetR, DesT does not possess the dimerisation wrapping arm. Helices α4, α5, α7, α8 and α9 of DesT run parallel to each other and perpendicular to the bound DNA. These helices form the long tunnel for ligand binding. The L-shaped oleoyl (18:1) moiety of the ligand is buried in this tunnel and complements the hydrophobic lining of the tunnel (**Fig. 1.22**). The CoA moiety is not seen, presumably due to disorder. The kink at the 18:1 chain due to the presence of the double bond bends the rest of the chain towards the periphery. This distal end is projected between α4 and α7. In the crystal structure of DesT and 16:0-CoA, the whole ligand is visible, including the CoA moiety. The entrance of the ligand-binding pocket does not change on swapping the unsaturated fatty acid-CoA for the saturated one. This is because the first seven carbons of the 18:0 saturated chain form the same interactions with the binding pocket as the first seven carbons of the 16:1 unsaturated fatty acid-CoA. However, since the saturated chain is essentially linear, the distal part of it inserts directly into the hydrophobic core of DesT without kinking (**Fig. 1.22**) (Miller *et al*., 2010). This action creates the new binding pocket for the saturated ligand. This pocket is now very similar to the pocket of EthR with bound hexadecyl octanoate in that both ligands run directly from top to bottom of the TFR, perpendicular to the bound DNA. This action by the saturated ligand but not by the unsaturated ligand captures the DNA-binding incompatible conformation of DesT, in which helix α4 and α7 are forced to pack closer together, and the recognition helices α3 and α3' rotate by ~5^{o} with a concomitant increase in the distance between them from 36.7 Å (DesT in complex with unsaturated ligand and DNA) to 41.9 Å (DesT in complex with saturated ligand) (Miller *et al*., 2010). This is similar to the movement seen in other TFRs.

DesT-oleoyl-CoA binds to a 30-bp duplex that contains an 18-bp pseudo-palindromic operator (Miller *et al*., 2010) (**Fig. 1.22**) with helix α3 again functioning as the recognition helix. Unlike in TetR or QacR, the N-terminal residue of α1 in DesT interacts with the minor groove. In other TFRs, helix α1 only serves as a structural element to stabilise the HTH core. Conversely, the N-terminus of helix α4 lacks the conserved lysine residue seen in TetR, and therefore this helix is not involved in DNA binding in DesT. DesT interacts with DNA mainly through the phosphate backbone, except for specific interactions with a pair of guanines in each

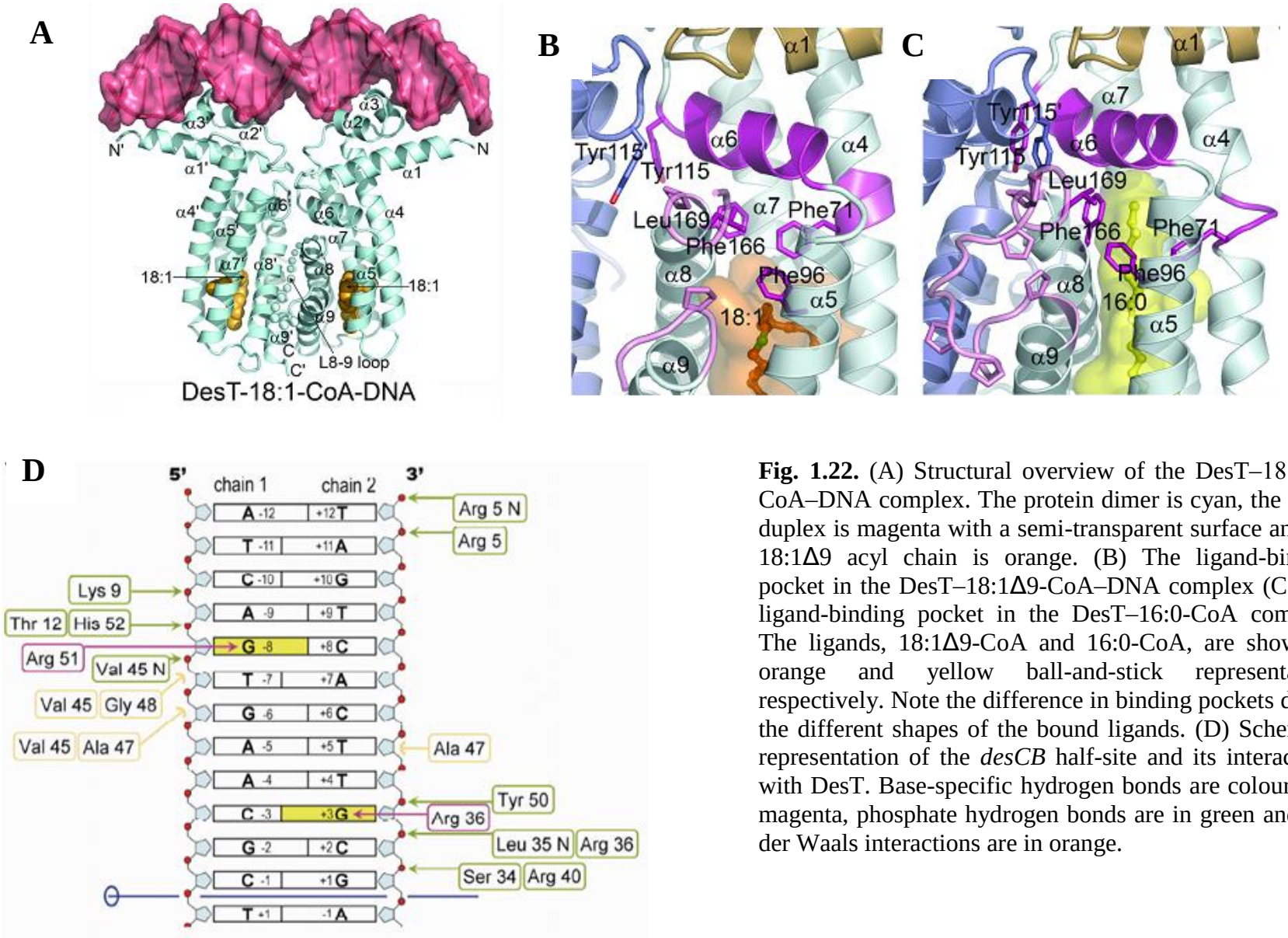


Fig. 1.22. (A) Structural overview of the DesT–18:1Δ9-CoA–DNA complex. The protein dimer is cyan, the DNA duplex is magenta with a semi-transparent surface and the 18:1Δ9 acyl chain is orange. (B) The ligand-binding pocket in the DesT–18:1Δ9-CoA–DNA complex (C) The ligand-binding pocket in the DesT–16:0-CoA complex. The ligands, 18:1Δ9-CoA and 16:0-CoA, are shown in orange and yellow ball-and-stick representation, respectively. Note the difference in binding pockets due to the different shapes of the bound ligands. (D) Schematic representation of the *desCB* half-site and its interactions with DesT. Base-specific hydrogen bonds are coloured in magenta, phosphate hydrogen bonds are in green and van der Waals interactions are in orange.

half-site (**Fig. 1.22**). The DNA is bent by only 3.3^{o} and is slightly deformed from the canonical B conformation, with the DNA unwound such that the major groove is widened 0.4 to 1.7 Å and the helical repeat is increased from 34 Å to 36Å (Miller *et al*., 2010).

1.7.6 CgmR

CgmR controls expression of the multi-drug transporter CgmA in *Corynebacterium glutamicum*, and the structures of CgmR in complex with two distinct inducers (methylene blue and ethidium bromide) have been solved (**Fig. 1.23**) (Itou *et al*., 2010). In the CgmR-ligand complexes the ligand is proposed to act as a rigidifying 'wedge' that traps the repressor in a state incompatible with DNA binding (Itou *et al*., 2010). The authors suggest that any ligand with sufficient volume and a favourable interaction with the binding pocket would lock CgmR in the derepressed state. The structure of the CgmR-operator complex shows that CgmR dimers bind to opposite faces of the nested DNA operator (**Fig. 1.23**) (Itou *et al*., 2010) in a manner reminiscent of the QacR system (Schumacher *et al*., 2002). The length of the spacer between the inverted repeats within the CgmR operators is 5 bp, compared to 4 bp in the case of QacR.

1.7.7 Other TFRs

Approximately 234 TetR-like structures from various organisms have been deposited in the Protein Database Bank (PDB). However, because almost all of these structures are of the apo-form, and because in almost all cases their biological ligands are not known, they do not contribute significantly to our understanding of ligand recognition, DNA-binding, or the mechanism of derepression. As a result, these structures are not discussed here.

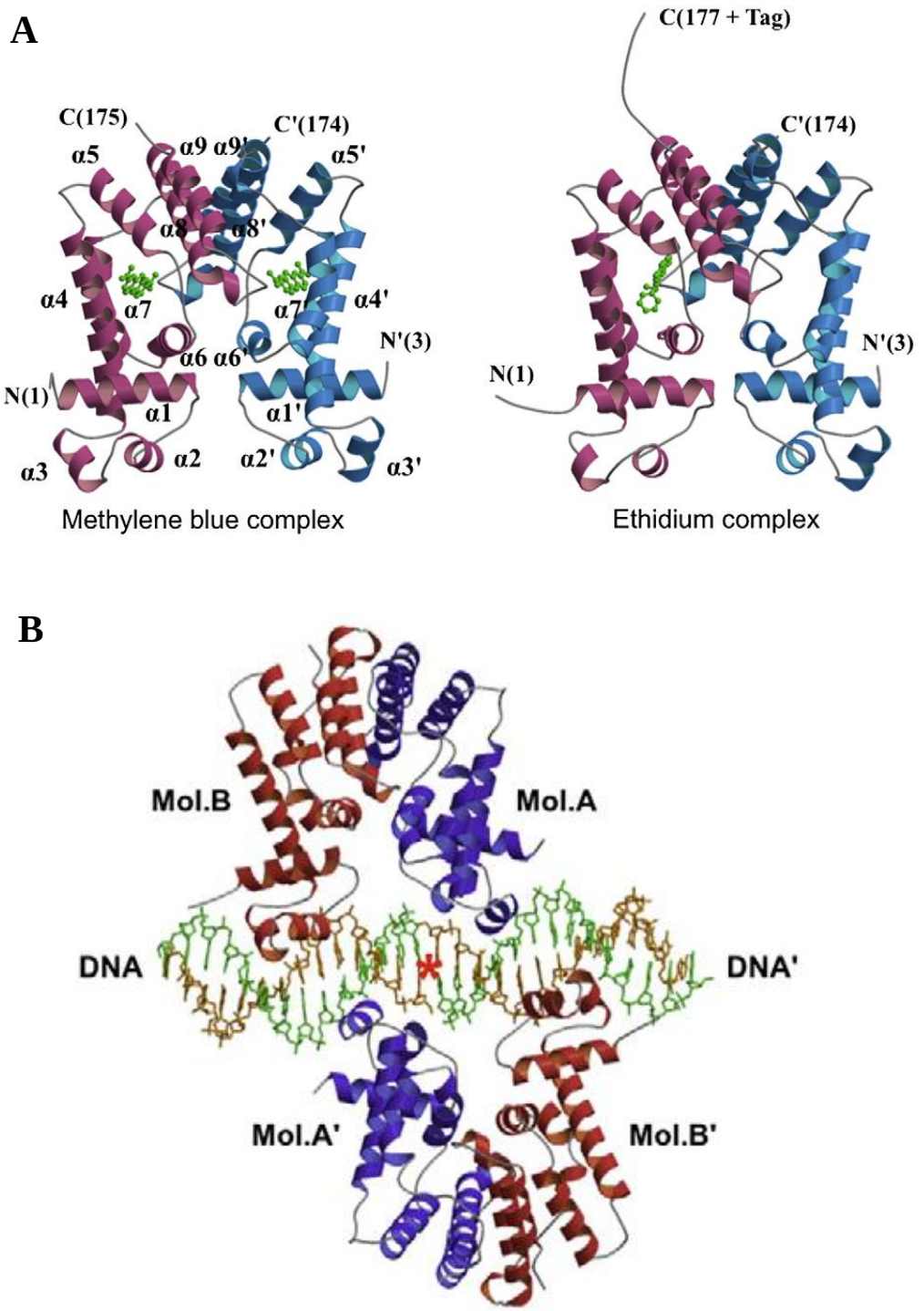


Fig. 1.23. (A) The overall structure of the CgmR dimer in complex with its ligands: methylene blue (left panel) and ethidium bromide (right panel). The overall structure of CgmR is more similar to QacR than TetR due to the missing helical wrapping arm. (B) The structure of two CgmR dimers bound to the nested operator DNA sequence. This is reminiscent of the QacR system in which two dimers are also bound to nested operator sequences (adapted from Itou *et al*., 2010).

1.8 Aims of this project

- To investigate the mechanism of self-resistance to the DNA gyrase inhibitor, simocyclinone D8, in the producing organism *S. antibioticus*.
- To investigate the role of a TetR-like repressor, SimR, in coupling the biosynthesis and the export of simocyclinones in *S. antibioticus*.
- To investigate at the molecular level, using X-ray crystallography, how SimR recognises simocyclinones and its cognate DNA-binding sites, and to elucidate the mechanism of ligand-mediated derepression.

Chapter 2
Materials and Methods

2.1 Bacterial strains and plasmids

2.1.1 *E. coli* strains

Derivatives of *E. coli* K12

Strain	Genotype	Source/ Reference
DH5α	F' *supE44 ΔlacU169* (*Φ80lacZ ΔM15*) *hsdR17 recA1 endA1 gyrA96 thi-1 relA1*	Hanahan (1983)
ET12567	F' *dam13::Tn9 dcm6 hsdM hsdR recF143::Tn10 galK2 galT22 ara-14 lacY1 xyl-5 leuB6 thi-1 tonA31 rpsL hisG4 tsx-78 mtl-1 glnV44*	MacNeil *et al* (1992)
ET12567 /pUZ8002	ET12567 containing helper plasmid pUZ8002	Paget *et al* (1999a)
BW25113	*Δ*(*araD-araB*)*567ΔlacZ4787*(*::rrnB-4*)*lacIp-4000*(*lacI*^{Q}) *λ-rpoS369*(*Am*) *rph-1Δ*(*rhaD-rhaB*)*568 hsdR514*	Datsenko and Wanner, (2000)
BL21 (DE3) /pLysS	F^{-} *omp*T *gal dcm lon hsd*S$_B$(r$_B^-$ m$_B^-$) λ(DE3 [*lac*I *lacUV5*-T7 gene 1 *ind*1 *sam*7 *nin*5]) pLysS (Cam^R)	Studier and Moffatt (1986)

2.1.2 *Streptomyces* strains

Strain	Genotype	Source/ Reference
Streptomyces lividans 1326	Wild type SLP2$^+$ SLP3$^+$	Kieser *et al*. (2000)
Streptomyces antibioticus Tü 6040	Environmental isolate; original producer of simocyclinone D8	Schimana *et al* (2000)

Streptomyces antibioticus Tü 6040 Δ*simX*	*simX::apr*	This thesis

2.1.3 Plasmids

Plasmid	Description	Source/ Reference
pIJ773	pBluescript KS (+) containing the apramycin resistance gene *apr* and *oriT* of plasmid RP4, flanked by FRT sites (Apr^R)	Gust *et al* (2003)
pIJ790	Modified λ RED recombination plasmid pKD20 (Cam^R)	Gust *et al* (2003)
pSET152	Plasmid cloning vector for the conjugal transfer of DNA from *E. coli* to S*treptomyces* spp. Integrates site-specifically at the ΦC31 attachment site (Apr^R)	Bierman *et al* (1992)
pUC19	*E. coli* multicopy cloning vector with *lacZ* selection (Bla^R)	Yanisch-Perron *et al* (1985)
pUZ8002	Non-transmissible *oriT*-mobilising plasmid (Kan^R)	Paget *et al* (1999a)
pMS82	Plasmid cloning vector for the conjugal transfer of DNA from *E. coli* to S*treptomyces* spp. integrates site-specifically at the ΦBT1 attachment site (Hyg^R)	Gregory *et al*. (2003)
pIJ10469	pMS82 carrying *simR* under the control of its native promoter	This thesis
pIJ10257	Plasmid integrating at the ΦBT1 *attB* attachment site of *S. coelicolor* and containing the strong *ermEp** promoter (Hyg^R)	Hong *et al.* (2005)
pIJ10480	pIJ10257 *ermEp*-simX*	This thesis
pIJ10481	pIJ10257 *ermEp*-simEX2*	This thesis

pIJ5972	Integrative *Streptomyces* promoter-probe plasmid based on TTA codon-free derivatives of the *luxAB* reporter genes	This thesis
pIJ10465	pIJ5972 *simXp-luxAB*	This thesis
pIJ10466	pIJ5972 *simRp-luxAB*	
pET28a	Plasmid for overexpression of N-terminally His-tagged and Thrombin cleavage tagged protein (Kan^R)	Novagen
pET20b	Plasmid for overexpression of C-terminally His-tagged protein (Bla^R)	Novagen
pIJ10490	pET28a derivative expressing His_6-tagged SimR	This thesis
pIJ10499	pET20b derivative expressing full-length wild-type SimR (residue 1-259) with C-terminal His-tag	This thesis
pIJ10500	pET20b derivative expressing SimR-ΔN10 (residue 11-259) with C-terminal His-tag.	This thesis
pIJ10501	pET20b derivative expressing SimR-ΔN15 (residue 16-259) with C-terminal His-tag.	This thesis
pIJ10502	pET20b derivative expressing SimR-ΔN22 (residue 23-259) with C-terminal His-tag.	This thesis
pIJ10503	pET20b derivative expressing SimR-ΔN25 (residue 26-259) with C-terminal His-tag.	This thesis

2.2 Growth conditions and storage of bacterial strains

2.2.1 *E. coli* strains

E. coli was grown on solid or in liquid media at 37°C (Sambrook *et al.*, 1989). Glycerol stocks were made from fresh overnight cultures by adding 40% (v/v) glycerol to an equal volume of culture and storing at -80°C.

2.2.2 *Streptomyces* strains

Streptomyces lividans cultures were grown on SFM for the preparation of spores. MYMTAP or OBM media were used for culturing *Streptomyces antibioticus*. A single colony was resuspended in 500 µl dH_2O and briefly vortexed and spun. The supernatant was spread on a solid media to yield a confluent lawn. The plates were incubated at 30°C for about six days or until confluent lawns of grey spores were visible. The plates were not left for more than two weeks to prevent significant loss of spore viability. The spores were harvested as described by Kieser *et al*. (2000) and stored at -20°C. The viable spore concentration was determined by plating out a dilution series on SFM plates. All *Streptomyces* strains were grown on solid media at 30°C, and were grown in liquid media at 30°C with shaking at 240 rpm (Kieser *et al*., 2000).

2.2.3 Antibiotic concentrations for *E. coli* and *Streptomyces* strains

The references for antibiotic concentrations are Sambrook *et al*. (1989) and Kieser *et al*.
(2000).

Antibiotic	Stock (mg/ml)	Final concentration (µg/ml)			
		S. lividans SFM	*S. antibioticus* MYMTAP	*E. coli* L	LB
Carbenicillin	200	-	-	100	100
Chloramphenicol	25 (in ethanol)	-	-	25	25
Kanamycin	50	-	-	50	50
Nalicidic acid	50 (in 0.25 M NaOH)	25	12.5	-	-
Apramycin	50	50	25	50	50
Hygromycin	40	50	50	50	-

2.3 Culture media, buffers, solutions

2.3.1 Solid media

Unless otherwise stated, the reference for all media is Kieser *et al*. (2000).

SFM medium (Soya flour mannitol medium)

This medium is used to prepare spores of *Streptomyces lividans* strains.

Agar .....20 g

Mannitol .....20 g

Soya flour .....20 g

Tap water to …….1000 ml

The mannitol was dissolved in the water and poured (100 ml) into 250 ml Erlenmeyer flasks each containing 1 g agar and 1 g soya flour. The flasks were closed and autoclaved twice (115°C, 15 min), with gentle shaking between the two runs.

MYMTAP medium

This medium is used to prepare spores of *Streptomyces antibioticus* strains.

Maltose…….4 g

Yeast Extract…….4 g

Malt Extract……..10 g

Bacto Agar……20 g

Made up in 50% tap water and 50% distilled water. After autoclaving 0.4 ml of R2 trace elements per were added for 200ml MYM

OBM medium (Oat-bran meal medium)

This medium is used to prepare spores of *Streptomyces antibioticus* strains.

Rolled Oat.....30g

Bacto agar.....15g

Tap water to …….1000 ml

The oat-bran meal was autoclaved twice in tap water (115°C, 15 min), with gentle shaking and addition of agar after the first run.

R2 medium

The following solutions were made up:

Sucrose …….103 g

K_2SO_4 ……...0.25 g

$MgCl_2.6H_2O$ …….0.12 g

Glucose ……10 g

Difco Casamino acids ……0.1 g

dH_2O (Distilled water) to ……..800 ml

80 ml of the solution was poured into 250 ml Erlenmeyer flasks each containing 2.2 g Difco Bacto agar. The flasks were closed and autoclaved. At the time of use, the medium was re-melted and the following autoclaved solutions were added into each flask in the order listed:

KH_2PO_4 (0.5% w/v in dH_2O) .....1 ml

$CaCl_2.2H_2O$ (3.68% w/v) ……..……8 ml

L-proline (20% w/v) …….1.5 ml

TES buffer (5.73% w/v, pH 7.2) ……...10 ml

Trace element solution …….0.2 ml

NaOH (1M) (unsterilised) ……0.5 ml

Required growth factors for auxotrophs …….0.75 ml

DNA (Difco nutrient agar)

4.6 g Difco Nutrient Agar was placed in each 250 ml Erlenmeyer flask and added with 200 ml distilled water. The flasks were closed and autoclaved.

L Agar

This is used for *E. coli* culture.

Agar…….. 10 g

Difco Bacto tryptone…….5 g

Difco yeast extract…...…….5 g

NaCl …….5 g

Glucose… ……1 g

dH_2O to..…….. ….1000 ml

The ingredients, except agar, were added in the distilled water and 200 ml was poured into 250 ml Erlenmeyer flasks each containing 2 g agar. The flasks were closed and autoclaved. When required, X-gal and IPTG were added to final

concentrations of 40 µg/ml and 200 µM, respectively.

2.3.2 Liquid Media

2×PG (Double strength germination medium)

Used for germinating *Streptomyces* spores.

Difco Casaminoacids..…….10 g

Difco yeast extract..……..10 g

$CaCl_2$ (added after autoclaving)..…….1.1 g

dH_2O to..……..1000 ml

The ingredients were dissolved, except $CaCl_2$, in the distilled water and 10 ml was aliquoted into universal bottle and autoclave. $CaCl_2$ was prepared and autoclaved separately as 1M solution and 100 µl was added to 10 ml yeast extract/Casaminoacids solution at time of use.

NMMP (Minimal liquid medium)

This medium is used to obtain dispersed growth, with specific carbon sources.

The following solutions were made up:

$(NH_4)_2SO_4$ ……..2 g

Difco Casaminoacids …… 5 g

$MgSO_4.7H_2O$ ……0.6 g

PEG 6000 …… 50 g

Minor elements solution ……1 ml

dH_2O to ……800 ml

80 ml aliquots were dispensed and autoclaved. At time of use, the following solutions were added:

NaH_2PO_4/K_2HPO_4 buffer (0.1M, pH6.8) ……15 ml

Carbon source (20%) ……...2.5 ml

Any required growth factors …… ..2.5 ml

Spore inoculum (a heavy inoculum, usually pre-germinated)

Lennox broth (L broth)

Difco Bacto tryptone ……10 g

Difco yeast extract ……5 g

NaCl ……..5 g

Glucose ……1 g

dH_2O to ……1000 ml

The mixture was dispensed in 100 ml volumes and then autoclaved.

YEME (Yeast extract-malt extract medium)

Difco Bacto-peptone ……3 g

Difco yeast extract ……...5 g

Oxoid malt extract ……5 g

Glucose ……10 g

Sucrose ……..340 g (34% final)

dH_2O to ……1000 ml

After autoclaving, the following solution was added:

$MgCl_2.6H_2O$ (2.5 M) …….2 ml/litre (5mM final)

TSB (Tryptone soya broth)

Used for growing *S. lividans* on a small scale.

Oxoid Tryptone Soya Broth powder (CM129) ……..30 g

dH_2O to ……..1000 ml

Tu6040-TSB broth

Used for growing *S. antibioticus* for chromosomal DNA isolation or for conjugation with *E. coli*.

Oxoid Tryptone Soya Broth powder (CM129) ……..30 g

Succrose.....100 g

L-Glycine.....4 g

Tap water added to..... 1000 ml

2.3.3 Buffers and solutions

Trace element solution

$ZnCl_2$ …….40 mg
$FeCl_3.6H_2O$ ……...200 mg
$CuCl_2.2H_2O$ ……..10 mg
$MnCl_2.4H_2O$ ……..10 mg
$Na_2B_4O_7.10H_2O$ ……10 mg
$(NH_4)_6Mo_7O_{24}.4H_2O$ ……10 mg
dH_2O to ……1000 ml

10×TBE

Trizma base …….108 g
Boric acid …….55 g
0.5 M EDTA (pH 8) …….40 ml
dH_2O to ……1000 ml

This gave a solution of pH 8.3.

2.4 Fermentation, isolation and purification of SD8 and SC4

SD8 was isolated as described by Schimana *et al*. (2000). Briefly, *S. antibioticus* Tü 6040 was fermented in a complex medium consisting of 2% (w/v) mannitol and 2% (w/v) soybean meal in a 20 l fermentor, and simocyclinones were extracted from the mycelium with methanol. Pure SD8 was obtained after reversed-phase HPLC using Nucleosil-10 C-18 material and 0.01% trifluoroacetic acid-acetonitrile gradient elution, resulting in a dark yellow powder after drying. Pure SC4 was isolated using essentially the same procedure, but fermentation was carried out in a defined medium containing 20% (v/v) glycerol and 0.15% (w/v) L-arginine to maximise SC4 production (Theobald *et al*., 2000). Prof. Hans-Peter Fiedler (Mikrobiologisches Institut, Eberhard-Karls-Universität Tübingen, Auf der Morgenstelle 28, D-72076 Tübingen, Germany) kindly supplied us with purified SD8 and its biosynthetic intermediate SC4.

2.5 Nucleic acid isolation

2.5.1 Isolation of plasmid or cosmid DNA from *E. coli*

E. coli strains were grown overnight at 37°C in 10 ml LB broth containing the appropriate antibiotic for plasmid or cosmid selection, and harvested by centrifugation (5 min at 3000 rpm), resuspended in 200 µl of solution I (50 mM Tris-HCl pH 8, 10 mM EDTA), 400 µl 0.2 M NaOH, 1% (w/v) SDS were added and the mixture was left at room temperature for 5 min. 300 µl 3M potassium acetate (pH 5.5) were added, and the mixture was inverted a few times, and microcentrifuged for 5 min at top speed. The supernatant was extracted with an equal volume of phenol:chloroform (1:1, v/v), and the nucleic acid precipitated by adding an equal volume of isopropanol and left on ice for 10 min. The nucleic acid was harvested by centrifugation, washed in 70% (v/v) ethanol, dried, and resuspended in sterile water. Plasmids are routinely isolated from 3ml of overnight *E. coli* cultures, using Qiaprep Spin Miniprep Kit (Qiagen).

2.5.2 Isolation of chromosomal DNA from *Streptomyces*

Streptomyces cultures grown in a mixture of 4 ml of YEME containing 34% sucrose and 6 ml TSB for one day were harvested by centrifugation (10 min at 3000 rpm). The cells were washed in dH_2O and resuspended in 2 ml SET buffer (75 mM NaCl, 20 mM Tris-HCl, 25 mM EDTA pH 8), containing 68 µl of lysozyme (30 mg/ml). The suspension was incubated at 37°C for 30 min (or until the cells lysed if this occurred first). 56 µl of Proteinase K (20 mg/ml) and 240 µl of SDS (10% w/v) were added and the mixture was left at 55°C for 2 h. 800 µl 5 M NaCl and 2 ml chloroform were added, and the mixture was mixed on a wheel rotator for 30 min, and centrifuged for 15 min at 6000 rpm. The supernatant was extracted with 0.6 volumes of isopropanol, and the nucleic acid was spooled by using a sterilize glass rod. The nucleic acid was washed in 70% (v/v) ethanol, dried, and redissolved in sterile water.

2.5.3 Isolation of total RNA from *Streptomyces antibioticus*

For RNA preparation, approximately 10^9 *S. antibioticus* spores were germinated by heat-shock treatment in 5 ml TES buffer (0.05 M, pH 8) at 50°C for 10 min, then diluted with an equal volume of double-strength germination medium [1% (w/v) Difco yeast extract, 1% (w/v) Difco casaminoacids, 0.01 M $CaCl_2$] and incubated with shaking at 37°C for 6 h (modified from Kieser *et al*., 2000). Germinated spores

were inoculated into NMMP (Kieser *et al*., 2000) and incubated with shaking for a further 15 h at 30^{o}C. RNA was prepared essentially as described by Hesketh *et al*. (2007), but with minor modifications taken from the Qiagen RNA Extraction Kit procedure (Qiagen). Essentially, 5 ml samples were collected and mixed with 10 ml of RNA protect bacteria solution (Qiagen) in Falcon tubes. The mixture was vortexed, incubated at room temperature for 5 min and spun down at 3000 rpm for 5 min. The pellet was resuspended in 1ml TE (10 mM Tris, pH 8, 1 mM EDTA, 15 mg/ml lysozyme) and incubated at room temperature for 60 min. 4 ml RLT buffer (Qiagen) was added and the samples were sonicated on ice for 3 cycles (20 s ON and 20 s OFF) with 18 microns amplitude. Samples were extracted twice with 2 ml phenol-chloroform-isoamylalcohol pH 8 (Sigma), and once with 4 ml chloroform. The upper layer was removed and mixed with 2.8 ml 100% ethanol and applied to a Qiagen Rneasy midi column and centrifuged at 3000 rpm for 1 min. The column was washed with 2 ml RW1 buffer (Qiagen) and DNA on the column was digested with DNaseI (Qiagen) for 60 min at room temperature. The column was then washed with 2 ml RW1 buffer (Qiagen) and centrifuged at 3000 rpm for 5 min. RNA on the column was washed twice with 2.5 ml RPE buffer (Qiagen) by spinning at 3000 rpm for 5 min each time. The ethanol was removed from the column by spinning without any buffer and RNA was eluted with 300 µl RNase-free water. The flow-through was re-applied to the column and centrifuged at 3000 rpm for 5 min to have more concentrated RNA.

2.6 Genetic manipulations

2.6.1 Preparation of electro-competent *E. coli* cells

A single colony or glycerol stock of the desired *E. coli* strain was used to inoculate 10 ml of L broth supplemented with the appropriate antibiotic and grown overnight at 37^{o}C with shaking (300 rpm). The overnight culture was diluted 100-fold in 100 ml of fresh media and grown at 37^{o}C with shaking to an OD_{600} of 0.2-0.4. After chilling on ice for 10 min, the cells were harvested by centrifugation at 4000 rpm for 5 min at 4^{o}C. The supernatant fluid was discarded and the pellet resuspended in 50 ml ice-cold 10% glycerol and centrifuged as above. The pellet was resuspended in 25 ml ice-cold 10% glycerol and centrifuged as above. The supernatant was discarded and the pellet was resuspended in the remaining 500 µl supernatant. 50 µl aliquots

were frozen and stored at -80^{o}C .

2.6.2 Transformation of commercial *E. coli* DH5*α* competent cells

For transformation, the frozen commercial competent cells (Invitrogen) were quickly thawed. DNA was added to 50 µl of competent cells, which were incubated on ice for 30 min. The suspension was heat-shocked at 42^{o}C for 25 s, then transferred to ice for 2 min. 0.6 ml of warm L broth was added to the suspension, which was incubated for 1 h at 37^{o}C. The transformed cells were plated out onto L agar plates containing the appropriate antibiotic and incubated at 37^{o}C.

2.6.3 Conjugation between *E. coli* and *Streptomyces spp.*

DH5α was used as the primary *E. coli* cloning host and constructs were then transferred to the methylation-deficient *E. coli* strain ET12567 (MacNeil *et al*., 1992) containing the helper plasmid pUZ8002 (Paget *et al*., 1999a) by transformation. ET12567/pUZ8002 was grown under chloramphenicol and kanamycin selection. Donor cultures of ET12567/pUZ8002 carrying the conjugation vectors or their derivatives were grown at 37^{o}C to an OD_{600} of 0.4 - 0.6 in L broth supplemented with chloramphenicol (25 µg/ml), kanamycin (50 µg/ml) and apramycin (50 µg/ml; to select for the integrative vector pSET152) or hygromycin (40 µg/ml; to select for the integrative vector pMS82 or pIJ10257). The bacteria were washed twice and resuspended with approximately 10^8 *Streptomyces* spores that had been activated for rapid germination by heating to 50^{o}C for 10 min in 2×YT and then cooled. In the case of *S. antibioticus*, young mycelium (grown for 12 hours in Tu6040-TSB broth) was used instead. Donors (500 µl; 2×10^8 *E. coli*) and recipients (500 µl mycelial suspension, or 10^8 *Streptomyces* spores) were spread on SFM (or OBM for *S. antibioticus*) containing 10 mM $MgCl_2$ and incubated at 30^{o}C overnight. Plates were flooded with 1 ml dH_2O containing 0.5 mg nalidixic acid and 1 mg apramycin and incubated at 30^{o}C. Exconjugants were picked from the plates 4 days (or 2 weeks for *S. antibioticus*) later and streaked for single colonies on SFM or OBM plates containing 25 µg/ml nalidixic acid (or 12.5 µg/ml for *S. antibioticus*) and 50 µg/ml apramycin (or 25 µg/ml for *S. antibioticus*). 50 µg/ml hygromycin was used to select for pMS82 and pIJ10257.

2.7 *In vitro* manipulations of DNA

2.7.1 Agarose gel electrophoresis of DNA

DNA was subjected to electrophoresis on agarose gels after addition of 1/5 volume 5× loading dye. Agarose gels were prepared and run submerged with 1×TBE buffer. For most routine work, 1% agarose was used, but concentrations down to 0.5% and up to 2% were used for analysis of large or small fragments, respectively. EtBr at 0.5-1 μg/ml in 1×TBE buffer was used to stain gels for approximately 10 min. Gels were photographed using UV illumination at a wavelength of 254 nm. The DNA size marker was the 1 kb or 500 bp ladder.

DNA loading dye: 0.25% (w/v) Bromophenol Blue, 0.25% (w/v) Xylene Cyanol FF, 30% (w/v) glycerol.

2.7.2 Isolation of DNA fragments from agarose

Restriction fragments were isolated from agarose gels that were run and loaded as described above. Bands were visualised using long-wavelength UV light (310 nm) to minimise nicking of the DNA molecules. Fragments were excised with a cleaned razor blade. The DNA fragments were then extracted using the QIAquick Gel Extraction Kit from Qiagen, using a microcentrifuge. Recovery was estimated by Nanodrop (Thermo Scientific).

2.7.3 Ligation of DNA

Vector and insert DNA were mixed at a molar ratio of 1:2, respectively, with 1/10 volume 10× ligation buffer and 5 U T4 DNA ligase in 10 μl total volume. The mixture was incubated at 4^{o}C overnight. The ligated DNA was used to transform commercial *E. coli* DH5α competent cells.

2.7.4 Polymerase Chain Reaction (PCR)

Synthetic oligonucleotides primers (Genosys or Sigma) were used in the PCR. The reaction mixture contained 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM $MgC1_2$, 200 μM final concentration of each of the four dNTPs, 2.5 U Phusion polymerase, 50 pmol each primer, and 50 ng template DNA in a final volume of 50 μl. After denaturation at 98^{o}C for 30s, the samples were subjected to 25 cycles of denaturation

(98^{o}C, 45 s), annealing (55-62^{o}C, 45 s), and extension (72^{o}C, 60 s) and then incubated for 5 min at 72^{o}C. The annealing temperature varied with the primers. Also the annealing and extension time varied; 15 s was used for PCR products between 200 and 500 bp, 30s was used for PCR products between 500 bp and 1 kb, and 60 s was used for PCR product between 1kb and 2 kb.PCR products were analysed by agarose (1% w/v) gel electrophoresis.

Table of primers used in this work

Name	Sequence (5´-3´)
pX-F-EcoRI	GAATTCGAGCACGAACTCCTGCTGGC
pX-R-BamHI	GGATCCGACCACCACTTCCTCGGACTGG
pR-F-BamHI	GGATCCTCCCCAGCTGCTCTGGCGTACACC
pR-R-EcoRI	GAATTCGTGAACGTACCGACCATCAGGCCG
MS82-simR-F-HindIII	GGCAAGCTTTCAAGCCAGTGCTGGACGTTCC
MS82-simR-R-KpnI	AACGGTACCAACGGCATCCTCATCTGGCATGACC
intRX-138-F	AAAGATATCCTCGTTCATCCACACTCCCC
intRX-138-R	AAAGGATCCATCTGGCATGACCACCACTTC
intOX-123-F	CCAATTGCGCTACGCTCCTTC
intOX-123-R	CCTGCGCGGAGCCTCCGGAC
intOR-130-F	CACCCTCGGTGTCGGCCACC
intOR-130-R	AACGAGAACGAACCCGTCAG
simEX1-F-NdeI	GAGCATATGCCAGATGAGGATGCCGTTGC
simEX1-R-HindIII	TAGAAGCTTCTATCCGGCATTCCGAGCCG
simEX2-F-NdeI	GGGCATATGACCAGTTTCCAAGTCCAG
simEX2-R-HindIII	GGGAAGCTTACCTCCCGGCCGCAGTAGACC
DelsimX-F	ATGAATACGACGCGCCAGTCCGAGGAAGTGGTGGTCAT GATT CCG GGG ATC CGT CGA CC
DelsimX-R	CTAGTAGCCCGCTAATTCGACTTAGCGCTCCACTAGCTA TGT AGG CTG GAG CTG CTT C
simR-int-simX-R	CTATCCGGCATTCCGAGCCG
S1-probeX-F	GTAGAGGGACATCGTGCCGGC

S1-probeX-R	GGCCGAGCAGTACGGCCAGC
S1-probeR-F	GTGTCGGCCACCTTGACGGC
S1-probeR-R	CCACCGAGCTCTCCGACGATCG
Invitro1-F	GAACGATCTGGTCACGGCTC
Invitro1-R	AGCAGTACGGCCAGCACCCC
Invitro2-R	GCAGCGCCGTACCGACGATCACC
Invitro3-F	CATCGACGCCGCCTCGACACCC
R2-full-CtagHis-F	GATCATATGAACGAAAACGAACCGGTGAG
R2-full-CtagHis-R	GATCTCGAGCGCCAGCGCCGGGCGTTCGC
R2-M10-trunc-F-NdeI	GCCCATATGATGCATCCGGAACCGGCCGG
R2-A15-trunc-F-NdeI	GCCCATATGGCCGGTCGTCGCAGCGCGCG
R2-S22-trunc-F-NdeI	GCCCATATGAGCCACCGTACCCTGAGCCG
R2-T25-trunc-F-NdeI	GCCCATATGACCCTGAGCCGCGATCAGATTG

2.7.4 Polyacrylamide gel electrophoresis (PAGE)

Polyacrylamide gels were prepared by mixing 40 ml acrylamide/bis-acrylamide [6% (w/v), ratio 19:1], 7 M urea, 1×TBE solution (Severn Biotech Ltd., UK) with 200 µl ammonium persulphate 10% (w/v) and 80 µl TEMED to promote polymerisation. Sequencing reactions were denatured at 90^{o}C for 5 min and resolved on 40 cm long gels run in 1×TBE at a constant power of 1.2 kV. The gels were transferred to Whatman 3MM paper and dried on a vacuum drier. The sequencing ladder was visualised by autoradiography for 1 to 3 days.

2.7.5 S1 nuclease protection analysis

The isolation of RNA for S1 mapping was described in 2.5.3. For all assays, 30 µg RNA and 25 pmol labelled probe were dissolved in 20 µl NaTCA buffer and hybridized at 45^{o}C overnight, following denaturation at 65^{o}C for 10 min. Probes for S1 nuclease protection analysis were generated by PCR using a 5′ end-labelled oligonucleotide internal to the ORF and an unlabelled upstream primer [for *simXp*, primers S1-probeX-F and S1-probeX-R; for *simRp*, primers S1-probeR-F and S1-probeR-R; (section 2.7.4)]. A G+A sequencing ladder was generated by chemical sequencing (Maxam and Gilbert, 1980).

2.7.6 *In vitro* run-off transcription

This technique is useful to confirm the identification of a promoter. RNA polymerase containing a mixture of sigma factors was purified as described previously (Buttner *et al*., 1988) from *S. coelicolor* M600 grown to exponential phase in YEME (Kieser, 2000). Three different templates for *in vitro* run-off transcription were generated by PCR that differed in their left or right ends, to allow the *simR* and *simX* transcripts to be identified unambiguously. Templates were generated using the following primers: Template 1, Invitro1-F and Invitro1-R; Template 2, Invitro1-F and Invitro2-R; Template 3, Invitro3-F and Invitro1-R (section 2.7.4). Transcripts were analyzed on a 6% polyacrylamide-7 M urea gel using a heat-denatured, ^{32}P-labelled *Hin*fI digest of ΦX174 as size standards (Kieser *et al*. 2000).

In vitro transcription was carried out as described in Buttner *et al*. (1987). Essentially, 7 μl water and 5 μl DNA solution were mixed with 25 μl reaction mix and incubated at 30^{o}C for 2 min. 1 μl of RNA polymerase was added to the mixture and incubated at 30^{o}C for 5 min. After the incubation, 1 μl (=10 μCi) [α-^{32}P] CTP (>600 Ci/mmol: Perkin Elmer) and 1 μl RTP mix were added and incubated at 30^{o}C for 2 min. 1 μl 5 mg/ml heparin was added and incubated at 30^{o}C for 5 min. 2 μl CTP mix was added and incubated at 30^{o}C for 10 min. The sample was moved onto ice, 10 μl precipitation mix and 70 μl isopropanol were added and the sample was left at -20^{o}C for at least 1 h. The precipitated sample was then centrifuged at 13,000rpm, 4°C for 30 minutes and the supernatant was discarded. The pellet was washed with 100ul 70% ethanol and air- dried before being resuspended in 6 μl formamide loading dye. Sample was heated at 95°C for 2 minutes. Only 3 μl was loaded on 6% polyacrylamide gel /urea sequencing gel with appropriate radiolabelled DNA size markers.

2×Master mix: 1.21 g Tris base, 9.3 mg Na_2EDTA and 487.2 mg $MgCl_2{\cdot}6H_2O$ were dissolved in 35 ml water and the pH was adjusted to pH 7.9 with HCl. 25 ml 4 mM potassium phosphate (pH 7.5) was added and pH was adjusted again to pH 7.9. Water was added to 70 ml as a final volume and filtered and stored in 1 ml aliquots at -20^{o}C.

Reaction mix: 115 μl of 2×Master mix, 6 μl BSA (RNase-free, 50 mg/ml), 6 μl 100 mM DTT, 2 μl 100 mM EDTA, 80 μl glycerol (sterilised by autoclaving), 41 μl water.

RTP mix: RTP mix is an equal mixture of 30 mg/ml ATP, 30 mg/ml GTP, 30 mg/ml UTP.

CTP mix: 15 mg/ml CTP

Precipitation mix: 10 μl tRNA (10 mg/ml), 53 μl 3M sodium acetate (pH 6.0), 17 μl water.

2.7.7 PCR for ABI automated sequencing

The PTC-100 Programmable Thermo Controller (MJ Research, Inc.) was used in all PCR reactions. PCR sequencing reactions were prepared by adding 0.2 to 2 μg of plasmid DNA, or 15 to 30 ng of PCR product DNA, 1 to 5 pmol of primer, 0.5 μl of 100% DMSO, 2 μl of 5x sequencing buffer, 1 μl of Big Dye reaction mix (Big Dye Terminator v3.1, Applied Biosystems) and 1 μl Half Term, with the total volume made up to 10 μl in 200 μl tubes. The PCR program was 25 cycles of 30 s at 96°C, 30 s at 50°C, and 4 min at 60°C. When the PCR was finished, the reactions were sent for automated sequencing (TGAC, Norwich).

2.7.8 Analysis of sequences

Sequences were compiled and analysed using VectorNTI (Invitrogen).

2.7.9 Phosphorylation of DNA primers

Primers were phosphorylated using T4 Polynucleotide kinase (T4 PNK). The reaction mix contained: 2 μl 100 μM primers, 2 mM ATP or 2 mM [γ-^{32}P] ATP for radioactive labelling, 2 μl 10x T4 PNK buffer and 2 units of T4 PNK. The reaction volume was brought up to 20 μl with dH_2O. The mixture was incubated at 37^{o}C for 1 h followed by incubation at 65^{o}C for 20 min for deactivation of the enzyme. 1 μl of primer was used for PCR reactions directly from the phosphorylation mix.

2.8 PCR-targeted mutagenesis for *simX* knock-out in *S. antibioticus*

2.8.1 Introduction

This system has been adapted for use in *S. antibioticus* from the method developed for *S. coelicolor* (Gust *et al*., 2004) and originally for *E. coli* (Datsenko and Wanner, 2000). Many bacteria are not readily transformable with linear DNA because of the presence of an intracellular exonuclease that degrades linear DNA. However, the λ RED (*gam, bet, exo*) functions promote a greatly enhanced rate of recombination when using linear DNA. The method of Datsenko and Wanner (2000) allows the disruption of genes in *E. coli* chromosome using a selectable marker generated by PCR using primers that have long extensions homologous to the targeted genes. The linear product can then be introduced directly into a strain expressing the λ RED recombinase.

The strategy for PCR-targeting for mutagenesis of *S. antibioticus* is to replace a chromosomal sequence within a *S. antibioticus* SupercosI-derived cosmid (Galm *et al*., 2002) by a selectable marker that has been generated by PCR using primers with 39 nt extensions homologous to the target gene. The inclusion of *oriT* (RK2) in the disruption cassette allows conjugation to be used to introduce the PCR-targeted cosmid DNA into *S. antibioticus*. Conjugation is much more efficient than transformation of protoplasts and it is readily applicable to many actinomycetes (Matsushima *et al*., 1994). The potential methyl-specific restriction system of *S. antibioticus* is circumvented by passing DNA through a methylation-deficient *E. coli* host ET12567 (Paget *et al*., 1999a). Vectors containing *oriT* (RK2; Pansegrau *et al*., 1994) are mobilisable *in trans* in *E. coli* by the non-transmissible pUZ8002, which lacks a *cis*-acting function for its own transfer (Paget *et al*., 1999a).

To adapt the procedure of λ RED-mediated recombination for *Streptomyces*, cassettes for gene disruption were constructed (Gust *et al*., 2004) that can be selected both in *E. coli* and in *Streptomyces.* The λ RED recombination plasmid pKD20 (*E. coli* Genetic Stock Centre CGSC Strain # 7637) was modified by replacing the ampicillin-resistance gene *bla* with the chloramphenicol-resistance gene *cam*, generating pIJ790, to permit selection in the presence of SupercosI-derived cosmids (carrying ampicillin and kanamycin resistance). **Figure 2.1** illustrates the PCR-targeted gene disruption method.

2.8.2 Introduction of the *S. antibioticus* cosmid into *E. coli* BW25113/pIJ790 by electroporation

An overnight culture of *E. coli* BW25113 containing the λ RED recombination plasmid pIJ790 was grown in LB containing 25 µg/ml chloramphenicol at 30°C. The overnight culture was diluted 1:100 in SOB (Hanahan, 1983) containing 20 mM $MgSO_4$ and 25 µg/ml chloramphenicol and incubated for 3-4 h at 30°C with shaking at 250 rpm until the culture reached an OD_{600} of 0.6. The cells were pelleted by centrifugation for 5 min at 4,000 rpm at 4°C and resuspended by gentle mixing in 10 ml ice-cold 10% glycerol. The centrifugation step was repeated twice more, the supernatant removed and the cells resuspended in the remaining about 100 µl 10% glycerol. The cell suspension was mixed with 100 ng cosmid DNA (VII-8g cosmid) (Galm *et al*., 2002) and electroporation was carried out in a 0.2 cm ice-cold electroporation cuvette using a BioRad GenePulse II set to 200 Ω, 25 µF and 2.5 kV. 1 ml ice-cold LB was immediately added to the shocked cells and they were incubated with shaking at 30°C for 1 h. The cells were spread onto L agar containing 100 µg/ml carbenicillin, 50 µg/ml kanamycin and 25 µg/ml chloramphenicol and incubated overnight at 37°C.

2.8.3 PCR amplification of the extended resistance cassette

PCR amplification of the extended resistance cassette was performed using the Expand High Fidelity PCR system (Roche). The reaction contained 1×buffer 2, 50 ng template DNA, 50 pmol of each primer (delsimX-F and delsimX-R) (section 2.7.4), 5% DMSO, 250 µM final concentration of each of the four dNTPs, and 2.5 U DNA polymerase mix in a total volume of 50 µl. The cycle conditions consisted of an initial denaturation step of 94°C for 2 min and then 10 cycles of denaturation at 94°C for 45 s, primer annealing at 50°C for 45 s and extension at 72°C for 90 s followed by another 10 cycles of denaturation at 94°C for 45 s, primer annealing at 55°C for 45 s and extension at 72°C for 90 s before a final extension step of 72°C for 5 min. The PCR product was purified using a Qiaquick PCR purification Kit (Qiagen) and eluted in 30 µl deionised water.

2.8.4 PCR-targeting of the *S. antibioticus* cosmid

A single colony of *E. coli* BW25113/pIJ790 containing the *S. antibioticus* cosmid VII-8g (Galm *et al*., 2002) from section 2.8.3 was used to inoculate 5 ml LB containing 100 µg/ml carbenicillin, 50 µg/ml kanamycin and 25 µg/ml chloramphenicol and was incubated overnight at 30°C. The overnight culture was diluted 1:100 in SOB lacking $MgSO_4$ but containing 100 µg/ml carbenicillin, 50 µg/ml kanamycin, 25 µg/ml chloramphenicol and 10 mM L-arabinose and incubated for 3-4 h at 30°C with shaking at 250 rpm until the culture reached an OD_{600} of 0.6. The cells were recovered and washed with 10% glycerol as before (section 2.8.3). The cell suspension was mixed with 1-2 µl of the PCR product and electroporation was carried out as before (section 2.8.3). 1 ml of LB was immediately added to the shocked cells and they were incubated with shaking at 37°C for 1 h. The cells were spread onto L agar containing 100 µg/ml carbenicillin, 50 µg/ml kanamycin and 50 µg/ml apramycin and incubated overnight at 37°C.

2.9 Site-Directed Mutagenesis

The coding region of *simR* was PCR amplified using primers R2-full-CtagHis-F and R2-full-CtagHis-R (Section 2.7.4) and cloned into *Sma*I site of pUC19. The correct sequence of the insert was confirmed by sequencing. Abutting primers were designed to amplify the entire construct with the sense primer carrying the new codon at its 5´ end. The primers were phosphorylated (as described in 2.7.9) prior to PCR amplification. Hi-fidelity Phusion Taq polymerase (Finzymes) was used to generate the linearised blunt-end product and the PCR product was gel-purified and self-ligated overnight (**Fig. 2.2**). *Dpn*I was added to the ligation mix to remove any template DNA (plasmid purified from *E. coli* is always methylated) carried over from the gel purification step. Competent *E. coli* DH5α was transformed with 5 µl of the ligation mix and plated on LB-carbenicillin for selection (section 2.6.2). Transformants were grown either for 8 hours or overnight and plasmids were purified by miniprep column (Qiagen). The replacement (or deletion) of the targeted codons was confirmed by sequencing.

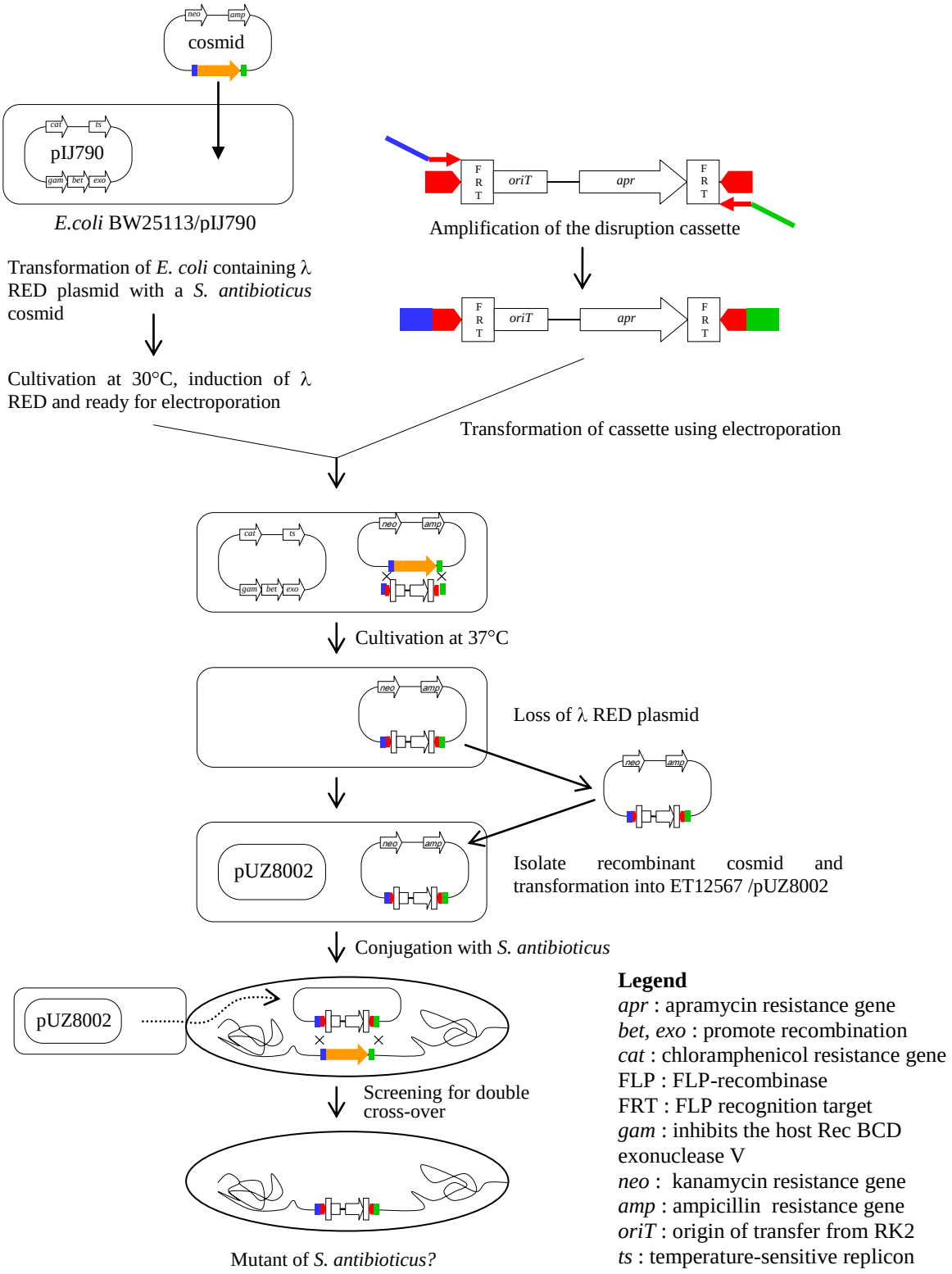


Fig.2.1. Gene disruption in *S. antibioticus* by PCR-targeting (Modified from Gust *et al*., 2003).

2.10 Antibiotic sensitivity assays

2.10.1 Construction of *simX* and *simEX2* expression plasmids

simX and *simEX2* were amplified by PCR using primers carrying *Nde*I (upstream) and *Hin*dIII (downstream) sites [for *simX*, simEX1-F-NdeI and simEX1-R-HindIII; for *simEX2*, simEX2-F-NdeI and simEX2-R-HindIII (section 2.7.4)]. The *simX* and *simEX2* fragments were cloned into *Nde*I-*Hin*dIII-cut pIJ10257 to generate pIJ10480 and pIJ10481, respectively, and transferred by conjugation into *S. lividans*.

2.10.2 Minimal inhibitory concentration (MIC) determinations

For MIC determinations, approximately $5x10^4$ spores were added to each well of a 96-well microtiter plate (Sterilin), containing a two-fold SD8 dilution series in SMMS medium (Kieser *et al*., 2000). The plates were evaluated everyday for five days after inoculation, and the MIC was defined as the lowest concentration at which no growth was observed. SD8 was dissolved in DMSO with the final concentration of DMSO in the agar not to exceed 1% (v/v).

2.11 Luciferase activity assays

2.11.1 Construction of *luxAB* reporter plasmids and luciferase activity measurements.

To probe *simXp* and *simRp* activities in the absence of SimR, the promoter regions (500 bp upstream of the translation start codons) were amplified by PCR using upstream primers carrying an *Eco*RI site and downstream primers carrying a *Bam*HI site [primers pX-F-EcoRI and pX-R-BamHI for *simXp*; pR-F-BamHI and pR-R-EcoRI for *simRp* (section 2.7.4)]. The *simXp* and *simRp* promoter fragments were cloned into *Eco*RI-*Bam*HI-cut pIJ5972, an integrative, *Streptomyces* promoter-probe plasmid based on TTA codon-free derivatives of the *luxAB* reporter genes (Aigle *et al*., 2000; M. Paget, pers. comm.), to create plasmids pIJ10465 and pIJ10466, respectively, and these reporter constructs were transferred by conjugation into *S. lividans*. To probe *simXp* and *simRp* activities in the presence of SimR, a plasmid carrying *simR* under the control of its own promoter (pIJ10469) was transferred by

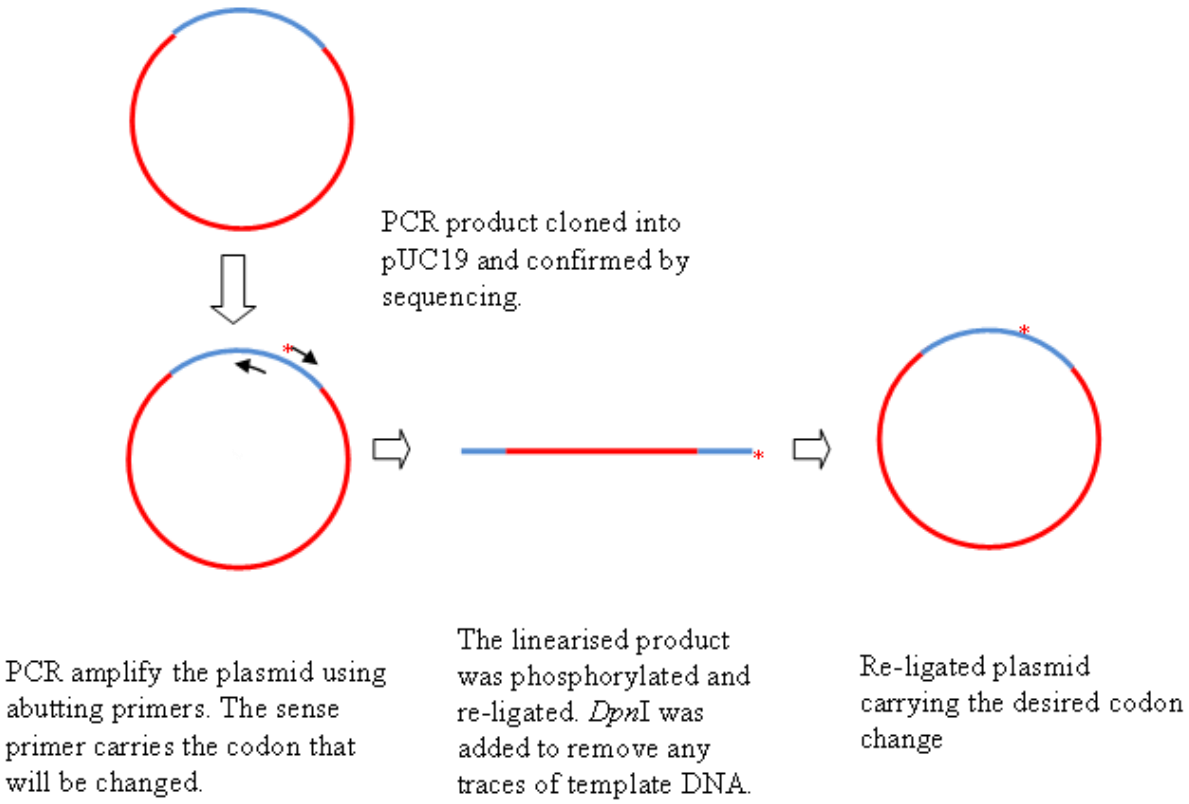


Fig. 2.2. Illustration for the site-directed mutagenesis strategy.

conjugation into strains already harbouring the promoter-probe plasmids pIJ0465 and pIJ10466. To construct pIJ10469, *simR* with its promoter was amplified by PCR using an upstream primer carrying a *Hin*dIII site and a downstream primer carrying a *Kpn*I site [primers MS82-simR-F-HindIII and MS82-simR-R-KpnI (section 2.7.4)] and the fragment was cloned into *Hin*dIII-*Kpn*I-cut pMS82.

Plasmid-containing strains were grown on Difco Nutrient Agar in single wells of a 96-well microtiter plate (Sterilin) for three days. Each well was inoculated with approximately $5x10^4$ spores. Plates were exposed to filter paper impregnated with n-decanal for 5 min and luciferase activities were quantified using a NightOwl camera (Berthold) equipped with WinLight software (Berthold) using a 1 min exposure time. Values given correspond to the average of three biological replicates from three different spore stocks.

2.12 Protein-related experiments

2.12.1 Overexpression and purification of native SimR

The *simR* gene of *S. antibioticus* Tü 6040 encoding a 259 amino acid protein was chemically synthesized with codon optimization (Genscript) for expression in *E. coli* and ligated into pET28a (Novagen) to give pIJ10490. This encodes a protein with an additional 26 amino acids at the N-terminus of the native sequence (with sequence MK-H_6-PMSDYDIPTTENLYFQGA), giving a total molecular weight of 32,222 Da. For expression of SimR with a C-terminal His-tag, the gene was amplified by PCR using an upstream primer carrying an *Nde*I site: R2-full-CtagHis-F (section 2.7.4) and a downstream primer carrying a *Xho*I site: R2-full-CtagHis-R (section 2.7.4). The amplified DNA fragment was 5'-phosphorylated, cloned into *Sma*I-cut pUC18 and verified by DNA sequencing. The *simR* coding sequence was excised by *Nde*I/*Xho*I digestion and cloned into *Nde*I-*Xho*I-cut pET20b, giving rise to the overexpression plasmid pIJ10499. This encodes a protein with an additional 8 amino acids at the C-terminus of the native sequence (with sequence LE-H_6), giving a total molecular weight of 30,197 Da.

For N-terminally His-tagged SimR expression, vector pIJ10495 was introduced into *E. coli* BL21(DE3)/pLysS and a 7 ml overnight culture was used to inoculate 500 ml Luria-Bertani medium containing 50 mg kanamycin and 15 mg chloramphenicol. The cells were grown at 37^{o}C to OD_{600} of ~0.4. The culture was cooled to 20^{o}C and 0.7 ml dimethylsulfoxide (DMSO) was added before induction of protein expression by the addition of isopropyl-*β*-D-thiogalactopyranoside (IPTG) to a final concentration of 0.3 mM. The addition of DMSO was essential for a high yield of soluble SimR, presumably because it artificially induces expression of chaperones that assist in the folding of SimR *in vivo*. The culture was left shaking for 3 h at 30^{o}C. Harvested cells were resuspended in 50 mM Tris-HCl (pH 8.4), 300 mM NaCl, 5% (v/v) glycerol, containing a Complete EDTA-free protease inhibitor cocktail (Roche) and lysed by sonication (three cycles of 20 s with 40 s resting on ice in between each cycle). The cell debris was removed by centrifugation at 84,000*g* for 30 min and the supernatant was filtered through a 0.45 µm membrane before being applied to a 1-ml Ni-loaded Hi-Trap Chelating HP column (GE Healthcare) that had been equilibrated with buffer A [50 mM Tris-HCl (pH 8.4), 300 mM NaCl, 50 mM imidazole]. Protein was eluted from the column using an increasing (50 mM to 500 mM) imidazole gradient in the same buffer. SimR fractions were identified using SDS-PAGE, pooled together, and concentrated to approximately 5-10 mg ml^{-1} using a Vivaspin 6 10-kDa cut-off concentrator (Vivascience). The concentrated protein was immediately exchanged into crystallization buffer [25 mM Tris-HCl (pH 8.4), 300 mM NaCl] using a Zeba desalting micro-column (Thermo Scientific).

2.12.2 Overexpression and purification of Selenomethionine-labelled SimR

Selenomethionine-(SeMet)-labelled SimR was obtained by the metabolic inhibition method. *E. coli* BL21(pLysS) cells containing pIJ10495 were grown in 2 l of M9 medium at 37^{o}C. When OD_{600} reached ~0.4, lysine, phenylalanine, and threonine were added to a final concentration of 100 mg l^{-1} each, and isoleucine, leucine, valine and SeMet were added to a final concentration of 50 mg l^{-1} each. The culture was incubated for a further 30 min at 37^{o}C and then induced with 1 mM IPTG for 3 h. The purification of SeMet-labelled SimR followed the protocol described above except that all buffers were degassed and contained 0.5 mM DTT to prevent protein oxidation. C-terminally His-tagged SimR was produced from the overexpression

plasmid pIJ10499 using the same procedure as for the N-terminally His-tagged protein, except that the medium was supplemented with 25 mg carbenicillin in place of kanamycin.

2.12.3 SDS-PAGE

Resolving gels (12.5% w/v) were prepared by mixing acrylamide/bis-acrylamide (30% w/v), 2 ml 1.5 M Tris-HCl (pH 8.8), 50 µl SDS (20% w/v) and dH_2O to make up 10 ml. 100 µl ammonium persulphate [10% (w/v)] and 10 µl TEMED was added to promote polymerisation. The resolving gel was overlayed with isopropanol at room temperature. After 20 min, the isopropanol was poured off and the gel was rinsed gently with dH_2O before addition of stacking gel (4% w/v). Stacking gels were prepared by mixing 0.67 ml acrylamide/bis-acrylamide (30% w/v), 0.625 ml 1 M Tris-HCl (pH 6.8), 20 µl SDS 20% (w/v) and dH_2O to make up 5 ml. 100 µl ammonium persulphate [10% (w/v)] and 10 µl TEMED were added to promote polymerisation. The stacking gel was poured on the resolving gel and a comb was inserted to create wells. The gel was left at room temperature for at least 30 min before use.

2.12.4 Electrophoretic mobility shift assays (EMSAs) and determination of dissociation constants (K_ds)

The EMSA DNA probe spanning the entire *simR-simX* intergenic region was amplified by PCR using primers intRX-138-F and intRX-138-R (section 2.7.4) and then 5′-end labelled using [γ^{32}-P] ATP and T4 polynucleotide kinase (New England Biolabs). The competitor DNA carrying only O_R was amplified using primers intOR-130-F and intOR-130-R, and the competitor DNA carrying only O_X was amplified using primers intOX-123-F and intOX-123-R (section 2.7.4). Binding of SimR to DNA was carried out in 20 µl EMSA Buffer [20 mM Tris, pH 8.0, 1 µg poly(dI-dC), 1 mM EDTA, 100 mM NaCl, 0.5 mM DTT, 8% (v/v) glycerol] containing 0.1 nM radiolabelled DNA (approximately 8000 cpm) and varying amounts of SimR. After incubation at 30^{o}C for 5 min, the binding reaction mixtures were loaded on 5% (w/v) native polyacrylamide gels and run in TBE buffer at 120 V for 45 min. The effect of SD8 and SC4 on the ability of SimR to bind DNA was tested by adding the compounds to the EMSA Buffer. The SD8 and SC4 stock solutions were made up in 100% DMSO.

EMSA data were collected and analysed on a PhosphoImager (FujiFilm) using Multi Gauge image analysis software (FujiFilm). Two independent EMSAs were carried out for each probe and mean values were calculated. In order to calculate K_ds for SimR binding to each operator when the two operator sites were separated, saturation curves (percentage of probe bound against concentration of protein) were fitted using SigmaPlot (one site saturation model). For K_ds of each operator when the two sites were coupled, a random-order binding model was used, where: $Y = [S]/(2*(K_{d1} + [S])) + [S]/(2*(K_{d2} + [S]))$, in which Y is the fractional saturation, K_{d1} and K_{d2} are dissociation constant of SimR binding to O_X and O_R operators, respectively, and [S] is the concentration of protein. The fractional saturation was calculated from intensities of EMSA bands in each lane as followed, Y = (0.5*intensity of middle band + intensity of top band) / (intensity of all bands in a lane). The equation was then fitted using SigmaPlot to determine K_ds.

2.12.5 DNaseI footprinting

Templates for DNaseI footprinting were amplified by PCR using one unlabelled primer and one primer 5′-end labelled using [γ^{32}-P] ATP and T4 polynucleotide kinase (New England Biolabs). The primers were intRX-138-F and intRX-138-R (section 2.7.4), the same pair used to generate the *simR-simX* intergenic region probe for the EMSA experiments. DNaseI footprinting assays were performed in 40 µl EMSA Buffer containing approximately 180,000 cpm radiolabelled DNA and varying amounts of SimR. After incubation at room temperature for 5 min, 10 µl DNaseI (10 units in 10 mM $CaCl_2$) was added and the incubation was continued for a further 60 s. Reactions were stopped by adding 140 µl DNaseI stop solution (200 mM unbuffered sodium acetate, 30 mM EDTA, 0.15% SDS and 0.1 mg/ml yeast tRNA), the samples were precipitated with ethanol, and the pellets were dried and dissolved in 5 µl Sequencing Loading Dye [80% (v/v) formamide, 10 mM NaOH, 1 mM EDTA, 0.1% (w/v) xylene cyanol and 0.1% (w/v) bromophenol blue]. After heating at 80^{o}C for 3 min and cooling on ice, the samples were run on a 6% (w/v) polyacrylamide/8 M urea sequencing gel, which was dried and analysed using a PhosphoImager (FujiFilm). A G+A sequencing ladder was generated from the template DNA by chemical sequencing (Maxam and Gilbert, 1980).

2.12.6 Limited proteolysis and protease protection assays

For limited proteolysis assays, 1 nmol of wild-type SimR was incubated with 1 pmol bovine trypsin (Sigma) in a total volume of 100 µl buffer [50 mM Tris (pH 8.0), 20 mM $CaCl_2$ and 150 mM NaCl] at 4 ^oC. For protease protection assays, 1 nmol wild-type SimR was incubated with equimolar amounts of 15mer, 25mer or 31mer double-stranded oligonucleotide containing the SimR O_X operator in a total reaction volume of 100 µl for 5 min at 4 ^oC before addition of 1 pmol bovine trypsin. 20 µl samples were then taken at 5 min time intervals. Reactions were stopped by adding SDS-PAGE loading buffer, boiled for 5 min, and analysed using SDS-PAGE. Proteins were transferred to PVDF membrane by electroblotting, stained with Coomasie blue, and proteolytically-resistant species were identified by N-terminal sequencing at the Protein & Nucleic Acid Chemistry Facility, University of Cambridge.

2.13 Protein crytallography: crystal methods

2.13.1 Protein crystallization and cryoprotection of SimR-apo and the SimR-SD8 and SimR-SC4 complexes

Prior to crystallisation, dynamic light scattering (DLS) was used to monitor the solution properties of the purified SimR sample. For this purpose, approximately 30 µl protein was centrifuged through a 0.1 µm Ultrafree filter (Millipore) to remove particulate material before introduction into a 12 µl microsampling cell. The cell was inserted into a DynaPro-Titan molecular-sizing instrument at 20 ^oC (Wyatt Technology). A minimum of 10 scattering measurements were taken and the resulting data were analysed using the *DYNAMICS* software package (Wyatt Technology).

Crystallization trials of His-tagged SimR were set-up using an OrxyNano robot (Douglas Instruments Ltd) in sitting-drop vapour diffusion format with 96-well MRC plates (Molecular Dimensions) using a variety of commercially available screens (Molecular Dimensions and Qiagen) at a constant temperature of 20°C. Drops consisted of 0.3 µl protein solution mixed with 0.3 µl precipitant solution and the reservoir volume was 50 µl; the protein concentration was approximately 5 mg ml^{-1}.

Improved crystals were subsequently obtained by refining the successful conditions in a hanging-drop format using 24-well VDX plates (Molecular Dimensions) over a reservoir volume of 1 ml.

Crystals of N-terminally His-tagged SimR-apo (both native and SeMet labelled) were obtained from 2% (w/v) polyethylene glycol (PEG)-10000, 0.2 M ammonium acetate in 0.1 M Bis-Tris (pH 5.5). Subsequently, C-terminally His-tagged protein was found to yield better quality crystals from the same conditions. All SimR-apo crystals were cryoprotected by supplementing the crystallization solution with 30% (w/v) PEG-1500.

SD8 and SC4 were isolated from *S. antibioticus* Tü 6040 as described (Schimana *et al*., 2000; Holzenkampfer *et al*., 2002). Both compounds were dissolved in DMSO and then added to C-terminally His-tagged protein to a final concentration of 500 µM, with the final concentration of DMSO not exceeding 2% (v/v). Crystals of the SimR-SD8 complex grew from 25% (w/v) PEG 1000 in 0.1 M sodium acetate (pH 4.6) and were cryoprotected by supplementing the crystallization solution with 32% (w/v) PEG 1500. Crystals of the SimR-SC4 complex grew from 28% (v/v) PEG-400, 0.2 M $CaCl_2$ in 0.1 M HEPES (pH 7.5). Due to the presence of PEG-400 in the mother liquor, no further cryoprotection was required for these crystals.

2.13.2 Structure determination and refinement of SimR-apo and the SimR-SD8 and SimR-SC4 complexes

All crystals were flash-cooled by plunging into liquid nitrogen and then mounted in Unipuck cassettes before being robotically mounted onto the goniostat on either station I02, I03 or I04 at the Diamond Light Source (Oxford, UK) and were maintained at -173^{o}C with a Cryojet cryocooler (Oxford Instruments). Diffraction data were recorded using an ADSC Quantum 315 CCD detector. The resultant data were integrated using MOSFLM (Leslie, 2006) and subsequently scaled by SCALA (Evans, 2006). Native data were collected from an apo crystal of the N-terminally His-tagged SimR to 2.3-Å resolution in space group H32. Solvent content estimation gave a value of 45% based on a single copy of the monomer in the asymmetric unit. A single apo SeMet-labelled crystal of N-terminally His-tagged SimR was used to collect a selenium anomalous peak data set at 3.4 Å resolution with the wavelength

set to 0.9763 Å. Experimental phases were determined by the single-wavelength anomalous dispersion method using the SHELX suite (Sheldrick, 2008) implemented through the HKL2MAP graphical user interface (Pape and Schneider, 2004). Substructure solution was performed with SHELXD, and seven of the expected nine selenium sites in the native sequence were located. This solution gave a figure-of-merit of 0.541 to 3.4-Å resolution after phasing in SHELXE. After density modification with PARROT (Cowtan, 2010) a partial backbone trace was obtained with BUCCANEER (Cowtan, 2006). At the same time, an approximate molecular replacement solution was obtained from the 2.3-Å resolution native data set using the BALBES pipeline (Long *et al*., 2008) (with a molecular replacement score of 2.63), which selected the structure of Sco6599 as a template (PDB accession code 2HXI, sequence identity with SimR 25.4%). The BUCCANEER model and the BALBES solution were then brought to the same origin after visual inspection in COOT (Emsley and Cowtan, 2004). The BALBES model was subsequently rebuilt against the experimentally-phased electron density map at 3.4-Å resolution and then refined against the 2.3-Å resolution native data set using REFMAC5 (Murshudov *et al*., 1997). At this stage the model contained only 118 residues and had R_{work} and R_{free} values of 50.6 and 51.7%, respectively. The refined model phases were then combined with the experimental phase information using the program SIGMAA (Read, 1986) to give improved phases. Further iterations of rebuilding, refinement and phase combination were performed until a model consisting of 165 residues was obtained, with corresponding R_{work} and R_{free} values of 36.3% and 40.0%, respectively. The work-flow is summarised in the **Fig. 2.2**. Thereafter, the model was refined against a new data set (but for C-terminally His-tagged SimR-apo instead) collected to 1.95 Å resolution, whereupon the electron density maps were significantly improved. Additional cycles of rebuilding and refinement yielded a model consisting of 238 residues, with corresponding R_{work} and R_{free} values of 20.5% and 23.4%, respectively.

The SD8 and SC4 complexes with SimR each yielded new crystal forms both containing two SimR protomers per asymmetric unit. In both cases, the structures were solved by molecular replacement using the structure of the C-terminally His-tagged SimR-apo monomer as the search model. The SimR-SD8 structure was solved using MOLREP (Vagin and Teplyakov, 2000), and the SimR-SC4 structure

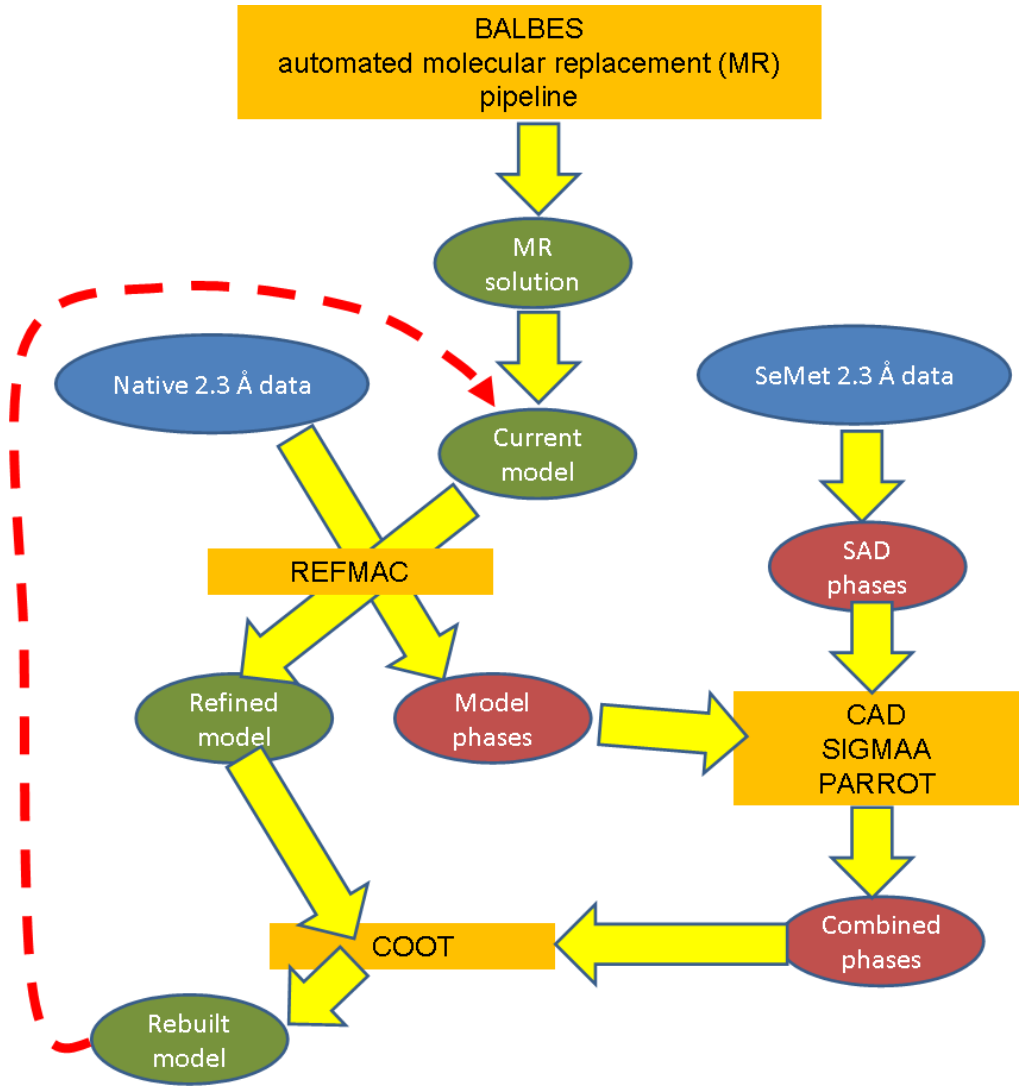


Fig. 2.2. Strategy for solving the structure of SimR-apo (Courtesy of Dr. David Lawson)

using PHASER (McCoy *et al.*, 2007). The two complex structures were then rebuilt and refined as for SimR-apo. All X-ray data collection and refinement statistics are summarized in **Table 5.1**.

Difference electron density 'omit' maps were generated for the SD8 and SC4 ligands using phases calculated from the final model minus the ligand coordinates after simulated annealing refinement. This was performed from a starting temperature of 5000 K after applying random shifts to the model ('shake' term set to 0.3) using PHENIX (Adams *et al.*, 2002). Structural figures were generated using PyMOL (DeLano, 2002).

Accession numbers

Coordinates and structure factors for SimR-apo, SimR-SD8 and SimR-SC4 described herein have been deposited in the Protein Data Bank with accession numbers 2Y2Z, 2Y30 and 2Y31, respectively.

2.13.3 Protein crystallization and cryoprotection of the SimR-DNA complex

C-terminally His-tagged SimR was purified and buffer-exchanged prior to crytallisation as described above.

Pairs of DNA oligonucleotides of different lengths (16 to 21 base pairs) and ends (blunt or sticky ends) were ordered from Oligos etc® and were reconstituted using the crystallisation buffer [25 mM Tris-HCl (pH 8.4), 300 mM NaCl]. Oligo pairs were annealed overnight to form duplexes. The final concentration of the duplexes was 2mM.

SimR and annealed oligonucleotides were mixed together in the ratio 1 SimR dimer to 1.2 double-stranded oligonucleotide and incubated at 20^{o}C for 10 minutes before crystallization screening. Crystallization trials of SimR-DNA were set-up in hanging-drop vapour diffusion format with 48-well VDX plates (Hampton Research) using a variety of commercially available screens (Emerald BioSystems and Hampton Research) at a constant temperature of 20 ^{o}C. Drops consisted of 1 µl SimR-DNA complex solution mixed with 1 µl precipitant solution and the reservoir volume was 150 µl. Improved crystals were subsequently obtained by refining the successful

conditions in a hanging-drop format using 24-well VDX plates (Molecular Dimensions) over a reservoir volume of 1 ml.

SimR-DNA crystals were obtained under several different screening conditions, but only with the blunt-ended 17mer DNA. The best crystals were obtained from solutions containing 10% (w/v) polyethylene glycol 8000, 0.2 M potassium chloride, 0.1 M magnesium acetate in 0.05 M sodium cacodylate (pH 6.5) two weeks after set-up. The crystals belonged to the orthorhombic space group $P2_12_12_1$. The SimR-17mer crystals were cryoprotected by a three-step transfer process in which ethylene glycol was added to the drop to a final concentration of 20%.

2.13.4 Structure determination and refinement of the SimR-DNA complex

All crystals were flash-cooled by plunging into liquid nitrogen and then mounted onto the goniostat at beamline 8.3.1 at the Advanced Light Source (Berkeley, California, US). The resultant data were integrated using MOSFLM (Leslie, 2006) and subsequently scaled by SCALA (Evans, 2006). Native intensity data were collected from a SimR-17mer crystal to 2.99 Å resolution. The reflections used to calculate the R-free value were selected in thin resolution shells to avoid bias resulting from the use of non-crystallographic symmetry (NCS) restraints in refinement. The structure of the complex was solved by molecular replacement using the structure of a subunit of C-terminally hexa-histidine-tagged apo SimR (PDB: 2Y2Z) and an idealised B-DNA of the correct sequence as the search models in PHASER (McCoy *et al*., 2007). SimR-17mer crystals contained two SimR dimer-DNA complexes in the asymmetric unit. The structure of the complex was then rebuilt in COOT (Emsley and Cowtan, 2004) and refined using REFMAC5 (Murshudov *et al*., 1997) and PHENIX (Adams *et al*., 2002) with NCS restraints. In the final stages, TLS refinement was used with a total of 20 TLS domains, which were defined using the TLS motion determination server (http://skuld.bmsc.washington.edu/~tlsmd/) (Painter and Merritt, 2006). X-ray data collection and refinement statistics are summarized in **Table 6.1**.

Structural figures were generated using PyMOL (DeLano, 2002). The local DNA helical parameters were calculated using Curves+ (Lavery *et al*., 2009).

Accession numbers

Coordinates and structure factors for SimR-DNA complex described herein have been deposited in the Protein Data Bank with accession number 3ZQL.

2.14 Global bioinformatic analysis of TFRs

I searched the PFAM database (http://pfam.sanger.ac.uk) for proteins that match the Hidden Markov Model profile PF00440, identifying 23,137 TFR candidates. Protein sequences longer than 300 amino acid residues were removed to eliminate false positives, and highly similar orthologous TFRs were removed using Jalview with a threshold of 99% identity (Waterhouse *et al*., 2009), resulting in a non-redundant set of 12,715 TFRs.

The non-redundant set of TFRs was divided into clusters of 200 sequences using USEARCH and UCLUST (Edgar, 2010). The amino acid sequences of the TFRs in each cluster were then aligned using MUSCLE (Edgar, 2004) to identify their N-terminal extensions, which were defined as the amino acid sequences preceding the conserved core DBDs (See **Fig. 6.1**). The globular body of the TFRs was defined by excluding the N-terminal extension from the whole protein sequence. In-house Perl scripts were used to quantify the length of the N-terminal extension and the fractions of R+K or D+E residues within these extensions. The sequences of the N-terminal extensions were concatenated together and submitted to the Regional Order Neural Network (RONN) programme (Yang *et al*., 2005) to predict the disorder probability for each residue. QtiPlot (http://soft.proindependent.com/qtiplot.html) was used to produce histograms.

Chapter 3

The *in vivo* target of SD8

3.1 Introduction

Although the interaction between SD8 and the GyrA had been dissected biochemically, a key remaining issue was to establish the *in vivo* target of the simocyclinones. Recent transcription profiling studies in *E. coli* suggested but did not prove that gyrase is the target (Oppegard *et al.*, 2009). To address this issue, I selected spontaneous resistant mutants in *E. coli* and mapped the mutations.

Wild-type *E. coli* and other Gram-negative bacteria are resistant to simocyclinones because the compounds cannot penetrate the outer membrane. The bilayer outer membrane of *E. coli* is asymmetrical in the sense that the outermost layer is composed of mainly lipopolysaccharides (LPS), while the innermost layer is comprised of phospholipids (**Fig. 3.1**). There are β-barrel proteins which span the bilayer, and there are lipoproteins that are attached to the inner surface of the outer membrane by a lipid anchor. The outer membrane serves as the general protective shield against many toxic compounds, such as detergents (e.g. bile salts) or antibiotics (e.g. chlorobiphenyl vancomycin or moenomycin).

To isolate spontaneous SD8-resistant mutants, I therefore used an *E. coli* strain (*imp4213*) that carries an in-frame deletion in the *imp* (*i*ncreased *m*embrane *p*ermeability) gene (also known as *ostA* for *o*rganic *s*olvent *t*olerant). *E. coli* cells carrying this allele are sensitive to antibiotics and detergents. *imp* is an essential gene that encodes a protein required for outer membrane assembly in *E. coli*. *E. coli imp* mutants have frequently been used to study the mechanism of action of compounds that cannot penetrate the outer membrane of wild-type *E. coli.* One such example is the use of an *E. coli imp* strain to study the glycopeptide mannopeptimycin, which is produced by *Streptomyces hygroscopicus* (Singh *et al*., 2003).

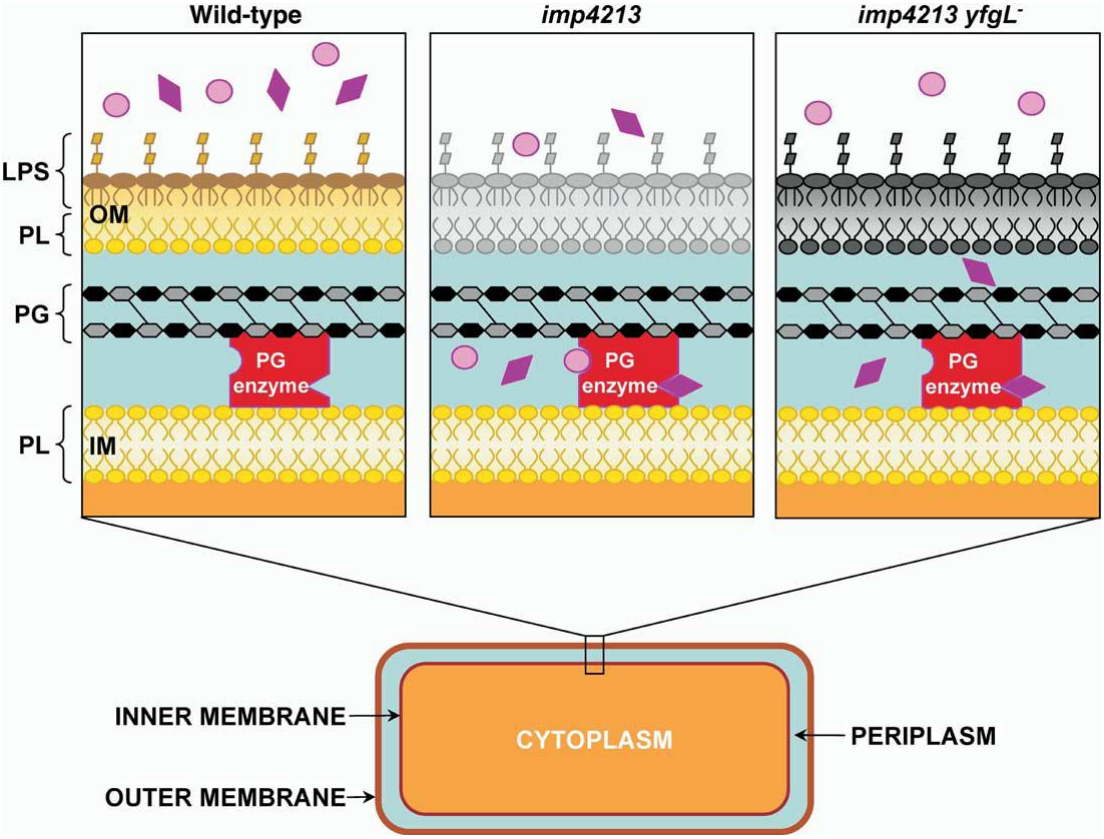


Figure. 3.1. The overall architecture of the cell envelope of wild-type *E. coli* and the *imp4213* mutant. There are three main layers in the cell envelope. They are the inner membrane (IM), the peptidoglycan layer (PG) and the outer membrane (OM). The outer membrane is composed of the outermost lipopolysaccharide layer (LPS) and the innermost phospholipid layer (PL). Membrane-bound proteins are omitted for clarity.

The wild-type OM acts as a barrier to some antibiotics (diamonds) or detergents such as bile salts (circles), which therefore cannot reach potential targets in the periplasm or cytoplasm. The *imp4213* has a defect in the OM which now has an increased permeability to antibiotics and bile salts. Second-site suppressor mutations in the *imp4213* mutant background (for example in this case in another OM biosynthetic gene, *yfgL*) can restore membrane impermeability. Because the *imp4213* mutation is an in-frame deletion, it cannot revert. The figure is adapted from Ruiz *et al*. (2005).

3.2 The isolation of spontaneous mutants resistant to SD8

Overnight cultures (300 µl) were plated on LB agar supplemented with SD8 to a final concentration of 10 µg/ml. Five independent overnight cultures were plated on separated SD8-supplemented plates in order to minimise the isolation of clonal mutants. Plates were incubated for 24-48 h and the resulting colonies were re-streaked onto SD8-supplemented agar to confirm their resistance. A ~500-bp region of *gyrA*, corresponding to residues M26 to S172 of the protein, was amplified by PCR from the resistant colonies and the resulting fragments were gel-purified and sequenced.

I isolated 31 spontaneous SD8-resistant mutants and found *gyrA* mutations in 22 of them, conferring one for the following amino acid changes: V44G, H45Y, H45Q, G81S and D87Y (**Table 3.1**). These amino acids are close to the bound SD8 molecule in the GyrA-SD8 crystal structure, consistent with gyrase being the *in vivo* target (**Fig. 3.2**). Unlike the 22 *gyrA* mutants, the remaining nine isolates had also acquired resistance to bile salts (**Table 3.1** and **Fig. 3.3**), showing they had restored outer membrane impermeability. Because the *imp4213* mutation is an in-frame deletion, these nine isolates cannot be simple revertants. However, several *imp4213* suppressor mutations that restore membrane impermeability to generic detergents such as bile salts have been mapped to other genes encoding further components of the outer membrane biosynthetic machinery (Ruiz *et al*., 2005). It is therefore likely that the remaining nine isolates are accounted for by such spontaneous second-site mutations.

The work contained in this chapter was published as part of Edwards *et al*. (2009).

MUTATIONS	**V44G**	**H45Y**	**H45Q**	**G81S**	**D87Y**
SP1					√
SP2					√
SP3					√
SP4					√
SP5				√	
SP6.1				√	
SP6.2					√
SP6.3					
SP6.4					
SP6.5					
SP6.6			√		
SP7.1					
SP7.2					
SP7.3					
SP7.6					√
SP8.1	√				
SP8.2					
SP8.3					√
SP8.4					√
SP8.5					
SP8.6		√			
SP9.1				√	
SP9.2					√
SP9.3					√
SP9.5					√
SP9.6					√
SP10.1					√
SP10.3					
SP10.4					
SP10.5					√
SP10.6				√	

Table 3.1. Summary of the spontaneous mutations conferring resistance to SD8 in the *E. coli imp* genetic background. The 22 mutants carrying mutations in *gyrA* are shown in black and the nine mutants in which no *gyrA* mutation was found are shown in red.

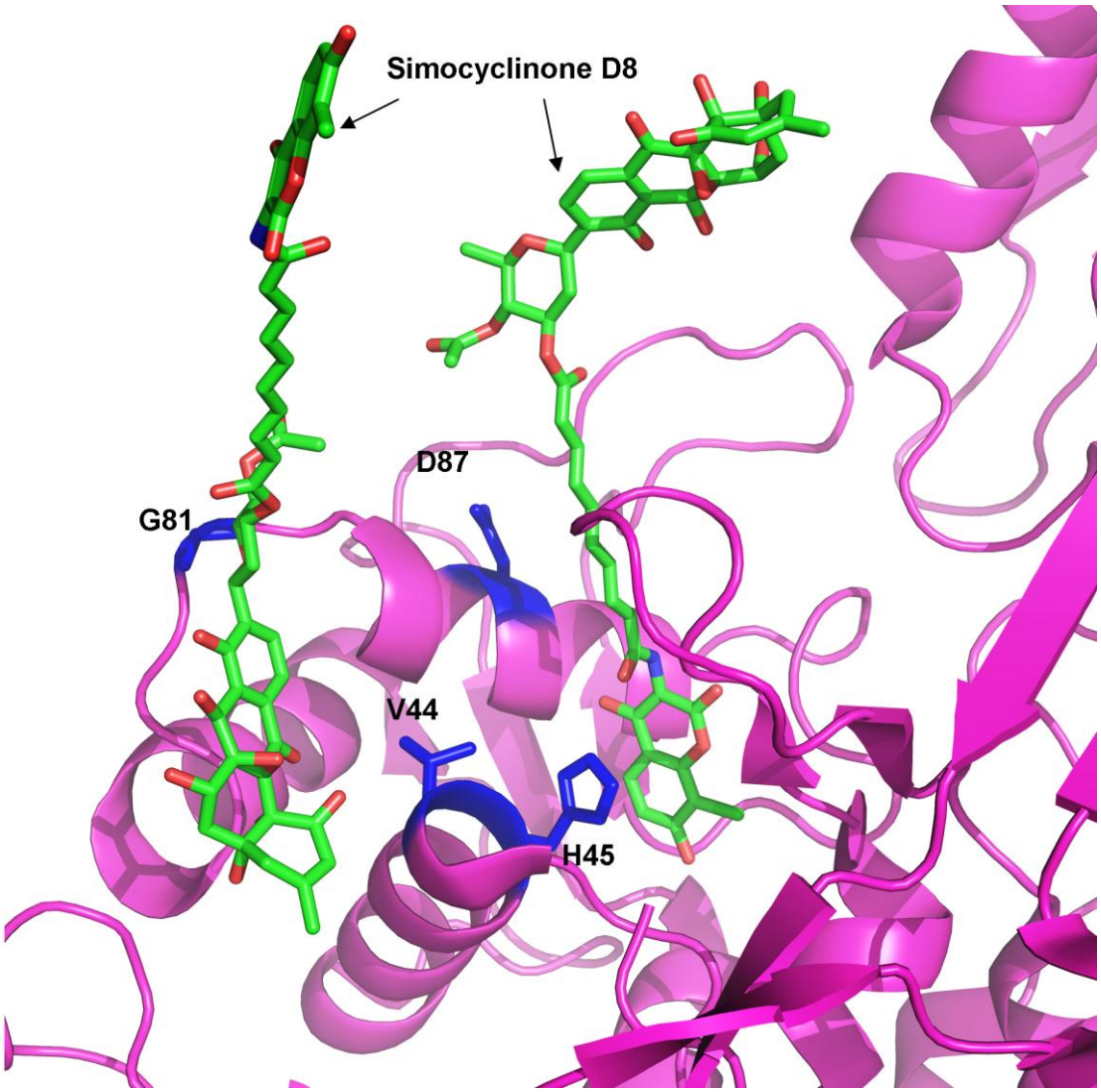


Fig. 3.2. The spontaneous amino acid substitutions in GyrA giving rise to SD8 resistance lie close to the bound antibiotic in the GyrA-SD8 crystal structure, consistent with gyrase being the *in vivo* target of SD8.

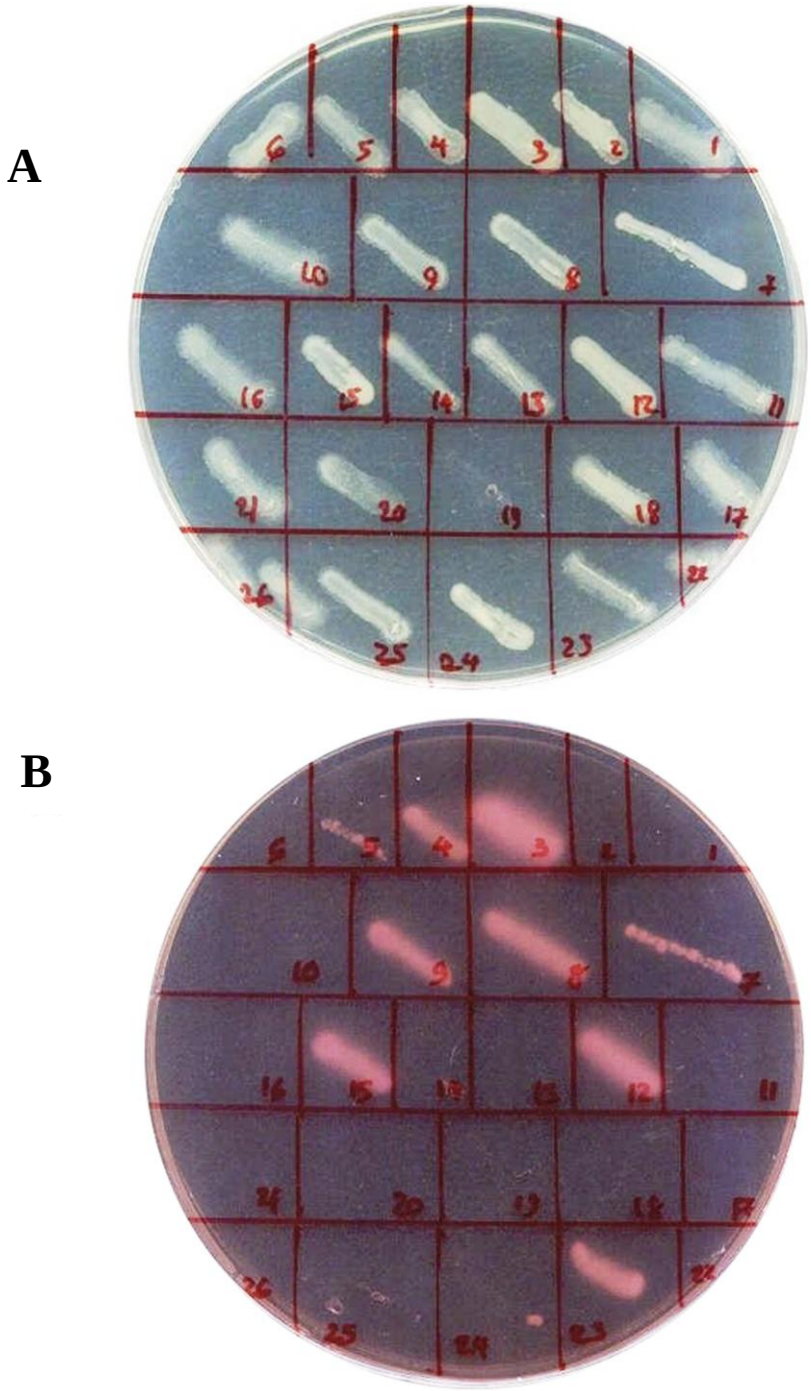


Fig. 3.3. Spontaneous SD8-resistant mutants grown on (A) LB agar, and (B) McConkey agar. Bile salts are present in McConkey agar but absent from LB agar. Unlike the 22 spontaneous SD8-resistant mutants that carried *gyrA* mutations, the remaining nine isolates had also acquired resistance to bile salts. This suggests that they could be accounted for by spontaneous second-site mutations that are known to restore outer membrane impermeability to the *imp* mutants (Ruiz *et al*., 2005).

Chapter 4

Coupling of the biosynthesis and export of the DNA gyrase inhibitor simocyclinone in *Streptomyces antibioticus*

4.1 Introduction

Aminocoumarin antibiotics are active against Gram-positive bacteria and function principally by inhibiting DNA gyrase (Maxwell and Lawson, 2003), an essential DNA topoisomerase found in all bacteria, which catalyses DNA supercoiling (Nollmann *et al*., 2007). In addition, a likely secondary target of these compounds is topoisomerase IV, which is involved in chromosome decatenation (Hardy and Cozzarelli, 2003; Oppegard *et al*., 2009). All four known aminocoumarins are produced by *Streptomyces* species. The first three to be discovered, novobiocin, clorobiocin and coumermycin A_1, each competitively inhibit the ATPase activity of the GyrB subunit of DNA gyrase and exhibit K_i values in the nanomolar range (Gormley *et al*., 1996). The most recently identified aminocoumarin antibiotic, SD8 (Schimana *et al*., 2000) (**Fig. 4.1**), also inhibits DNA gyrase but was unexpectedly found to have a completely novel mode of action, binding instead to the GyrA subunit of the enzyme and preventing its binding to DNA (Flatman *et al*., 2005; Edwards *et al*., 2009). SD8 is a potent inhibitor of DNA gyrase supercoiling with an IC_{50} lower than that of novobiocin (Flatman *et al*., 2005; Oppegard *et al*., 2009).

The genus *Streptomyces* accounts for the production of approximately two-thirds of the known antibiotics. They expel these compounds into their environment, typically the soil, most probably to give them a competitive advantage over other organisms that share the same ecological niche. Because the antibiotic is often potentially lethal to the producing organism, there must be mechanisms to ensure that the machinery responsible for export of the mature antibiotic is in place at the time of biosynthesis. This export machinery may be sufficient to confer resistance to the antibiotic, or there may be additional resistance mechanisms. For example, in the case of the novobiocin, clorobiocin and coumermycin A_1 producers, the principal mechanism of resistance is production of an aminocoumarin-resistant GyrB subunit ($GyrB^R$), encoded within the biosynthetic cluster and activated during antibiotic production (Thiara and Cundliffe, 1988, 1989, 1993; Schmutz *et al.*, 2003). The $GyrB^R$ subunit replaces the sensitive subunit ($GyrB^S$) in the $(GyrA)_2(GyrB)_2$ heterotetramer during the production phase. In addition, an aminocoumarin-resistant topoisomerase IV

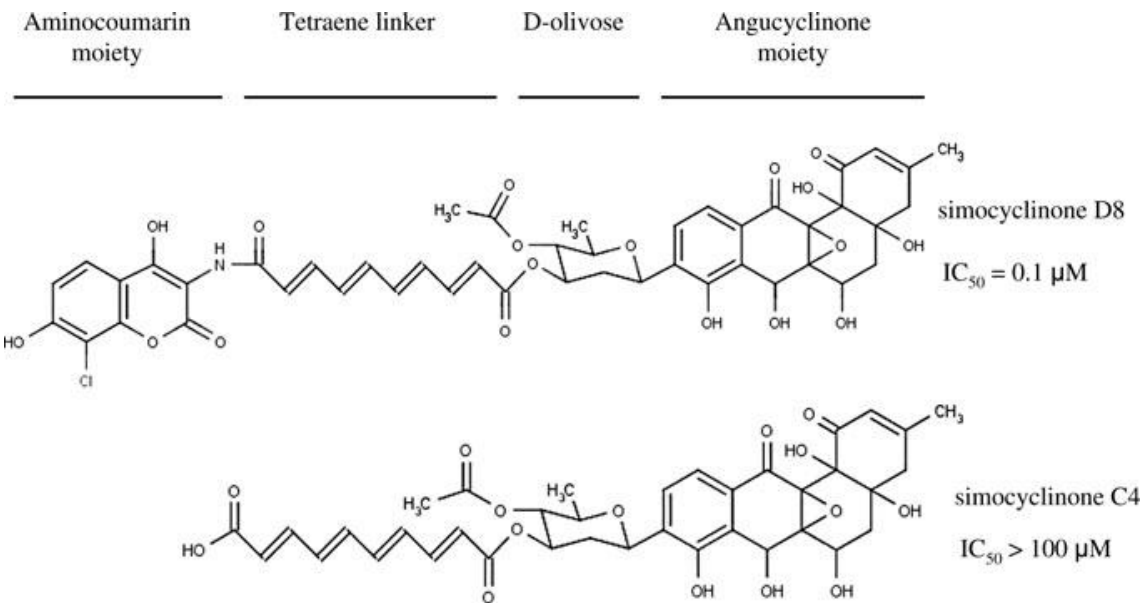


Fig. 4.1. Structures of simocyclinone D8 (SD8) and its biosynthetic intermediate, simocyclinone C4 (SC4), and their IC_{50}'s for *E. coli* DNA gyrase.

subunit, $ParY^R$, encoded within the clorobiocin and coumermycin A_1 biosynthetic clusters (but not the novobiocin biosynthetic cluster), also confers resistance to these antibiotics when introduced into a naïve host, presumably by an analogous mechanism (Schmutz *et al.*, 2003). However, no DNA gyrase or topoisomerase IV subunits are encoded within the *sim* biosynthetic cluster (Galm *et al*., 2002; Trefzer *et al*., 2002), leaving unknown the mechanism of resistance in the producing organism, *S. antibioticus* Tü 6040.

Within the *sim* cluster, among the genes responsible for the biosynthesis and linking of the four constituents of the antibiotic, are two divergently transcribed genes, *simR2* (hereafter, *simR*) and *simEX1* (hereafter, *simX*) (Galm *et al*., 2002; Trefzer *et al*., 2002). The SimR/SimX pair resembles the TetR/TetA repressor-efflux pump pair that causes widespread resistance to clinically important tetracyclines in several human pathogens (Chopra and Roberts, 2001). The similarity of SimR/SimX to TetR/TetA suggested they might be involved in simocyclinone efflux and, potentially, in simocyclinone resistance (Trefzer *et al*., 2002; Galm *et al*., 2002).

Here I show that *simX* encodes a simocyclinone efflux pump, and that transcription of *simX* is controlled by SimR, which directly represses the *simX* and *simR* promoters by binding to two operator sites in the *simR-simX* intergenic region. I show that SD8 abolishes DNA binding by SimR, providing an intimate mechanism that couples the biosynthesis of simocyclinone to its export. In addition, I show that an intermediate in the biosynthetic pathway, SC4, which is essentially inactive as a DNA gyrase inhibitor, also induces *simX* expression *in vivo* and relieves DNA binding by SimR *in vitro*, suggesting a potential 'feed-forward' mechanism (Tahlan *et al*., 2007) that might ensure expression of the SimX efflux pump prior to the build-up of a toxic concentration of the mature, active antibiotic.

4.2 *simX* encodes a simocyclinone efflux pump

There are two pump-like transmembrane proteins encoded in the *sim* cluster, SimEX1 (hereafter SimX) and SimEX2 (Trefzer *et al*., 2002). To determine if either of these two proteins is involved in SD8 efflux, I expressed *simX* and *simEX2* from

the strong constitutive promoter *ermEp** using the integrative, single-copy vector pIJ10257 (Hong *et al*., 2005). I introduced these constructs into the heterologous host *S. lividans* and compared the susceptibility of the resulting strains to SD8 (**Fig. 4.2**). The strain carrying *ermEp*-simEX2* had an MIC of 2 μg/ml, as did *S. lividans* alone, or *S. lividans* containing the parent vector, pIJ10257. In contrast, the strain carrying the *ermEp*-simX* construct had an MIC of 65 μg/ml. The *ermEp*-simX* construct did not confer resistance to the structurally related aminocoumarin DNA gyrase inhibitor novobiocin, nor to unrelated antibiotics such as erythromycin, lincomycin, rifampicin, bacitracin and nisin. Moreover, I constructed a *S. antibioticus simX* deletion mutant and showed that it was more sensitive to exogenous SD8 than the wild-type strain (**Fig. 4.3**). These results suggested that *simX* encodes a simocyclinone-specific efflux pump. Adjacent to *simX* is the divergent gene *simR*, encoding a potential transcriptional repressor of *simX*. I also cloned *simX* into an integrative vector under the control of its native promoter and in the absence of *simR*. This construct only mildly enhanced resistance to SD8 in *S. lividans*, giving an MIC of 4 μg/ml.

4.3 Mapping the transcription start points of *simR* and *simX*

High resolution S1 nuclease mapping of the *simR* and *simX* promoters was performed using RNA isolated from the SD8 producing organism, *S. antibioticus* Tü 6040 (**Fig. 4.4A**). A single *simR* promoter (*simRp*) was identified, initiating transcription 20 bp upstream of the *simR* ATG start codon, and a single *simX* promoter (*simXp*) was identified, initiating transcription 47 bp upstream of the *simX* ATG start codon (**Figs. 4.4A and 4.5**). *In vitro* run-off transcription experiments with purified *S. coelicolor* RNA polymerase confirmed the presence and locations of the *simR* and *simX* promoters (**Fig. 4.4B**).

4.4 SimR regulates expression of *simR* and *simX*

To investigate the regulation of *simR* and *simX*, *simR* and *simX* promoter activities were measured in the presence and absence of SimR, using the integrative luciferase

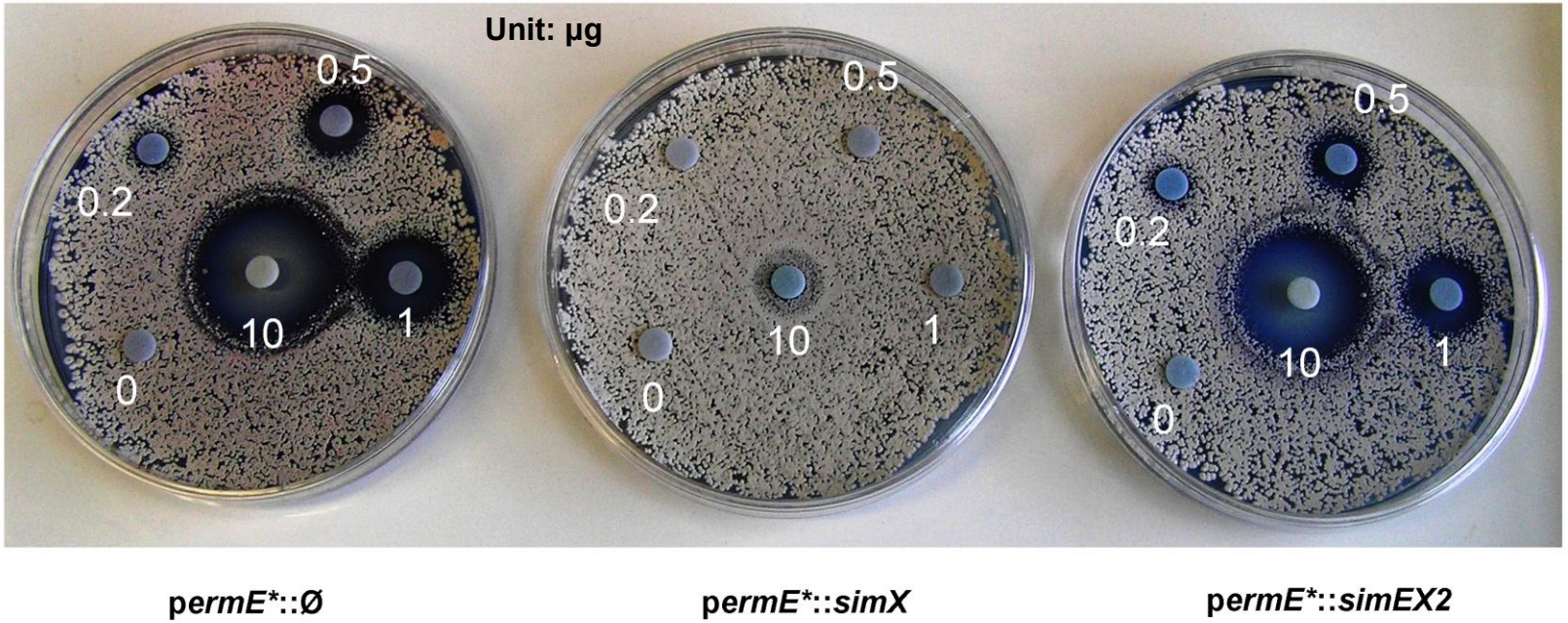


Fig. 4.2. Expression of the efflux pump *simX* confers SD8 resistance on a naive host, *S. lividans*. Approximately $5x10^4$ spores were spread on each R2 agar plate and paper discs impregnated with varying amounts of SD8 were applied to the freshly spread plates. The plates were incubated at 30^oC for 4 days before scoring.

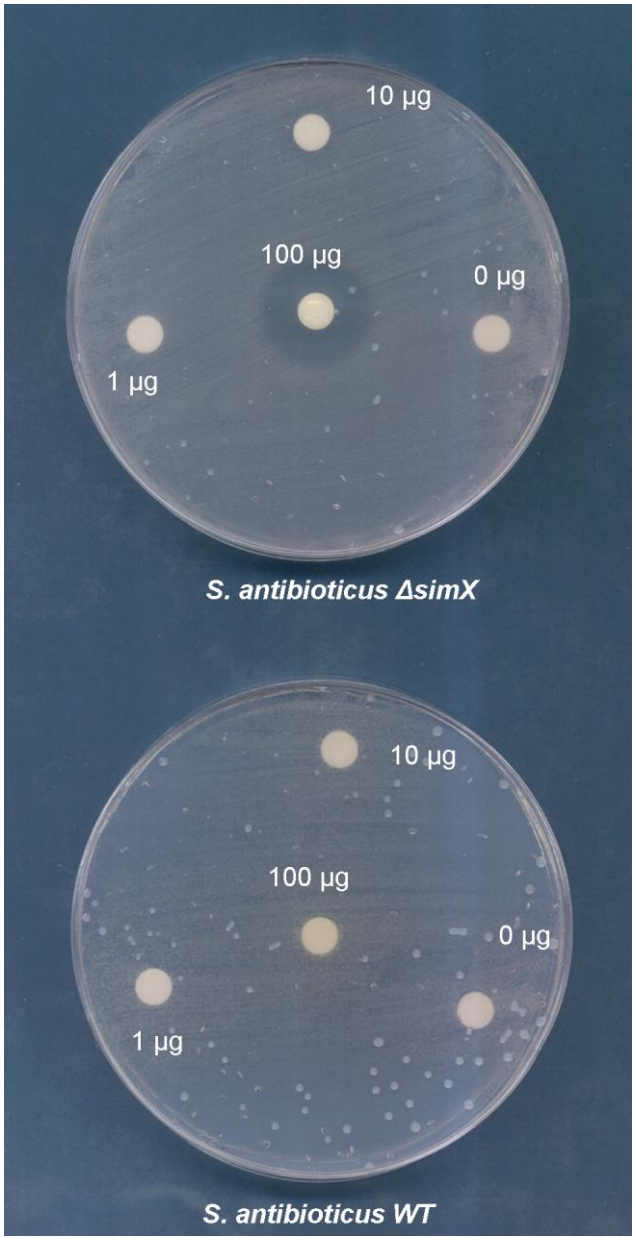


Fig. 4.3. The *S. antibioticus simX* deletion mutant is more sensitive to SD8 than the wild-type (WT) strain. Approximately $5x10^4$ spores were spread on each SMMS agar plate and paper discs impregnated with varying amounts of SD8 were applied to the freshly spread plates. The plates were incubated at 30^oC for 4 days before scoring.

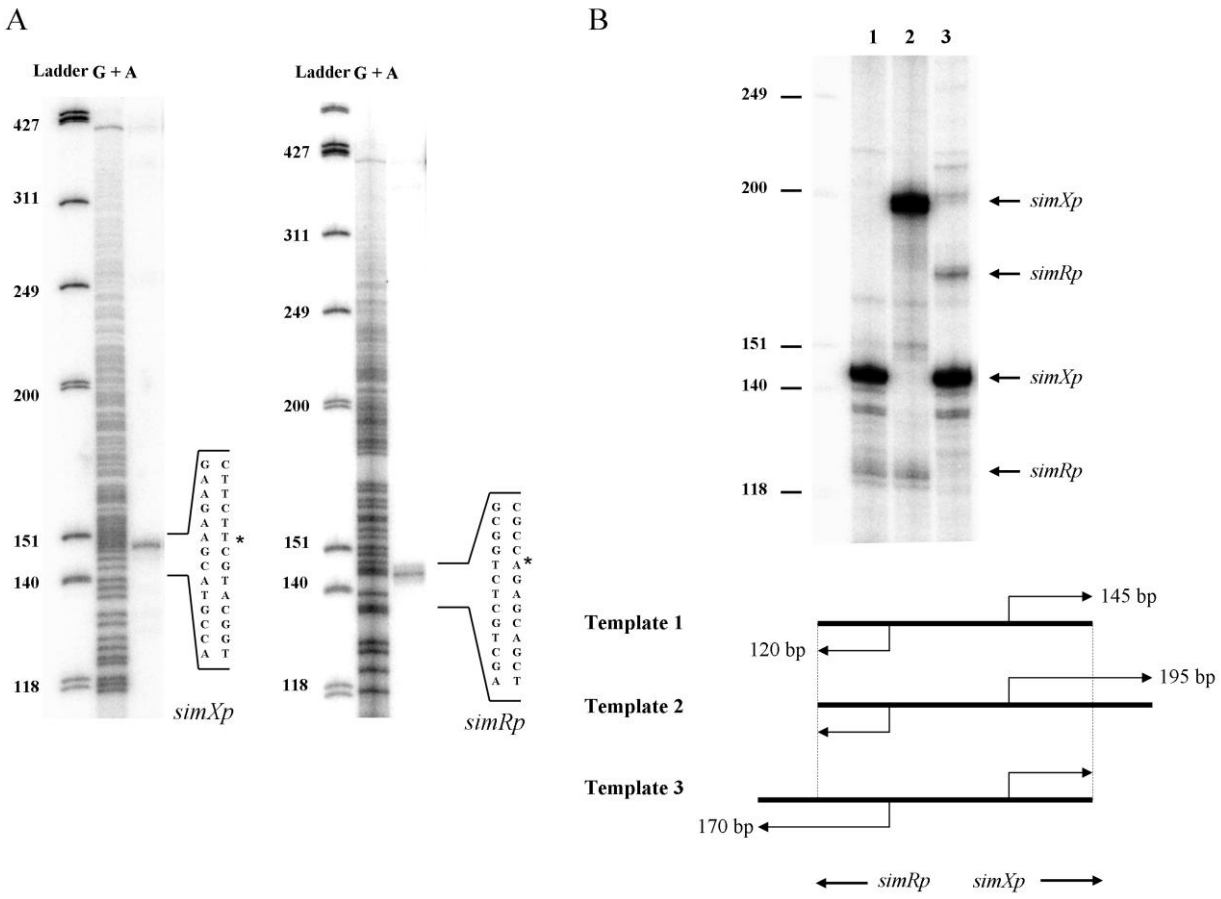


Fig. 4.4 (A) High-resolution S1 nuclease mapping of the 5' ends of the *simR* and *simX* transcripts using PCR-generated probes and RNA from the SD8 producing organism, *S. antibioticus* Tü 6040. The most likely transcription start points are indicated by the asterisks. The G+A Maxam-Gilbert chemical sequencing ladder was generated using the same probe as the one used for S1 nuclease mapping assays. The size markers are a radiolabelled *Hin*fI digest of ΦX174 DNA. (B) Run-off transcription from the *simR* and *simX* promoters *in vitro* using purified *S. coelicolor* RNA polymerase and the templates illustrated containing the *simR-simX* intergenic region. Lane numbers correspond to the different templates shown in the illustration below.

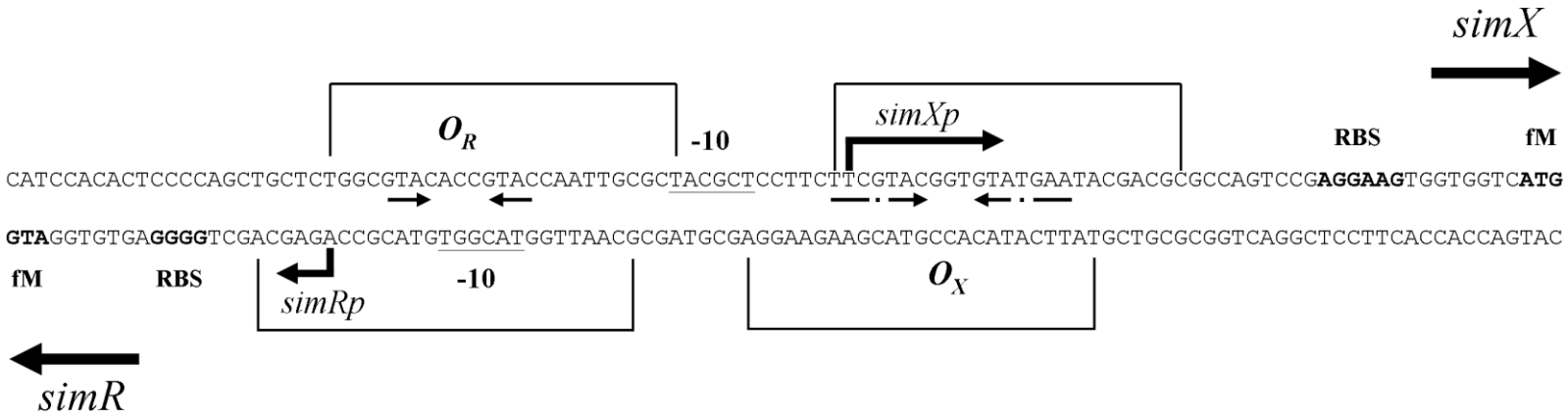


Fig. 4.5. Sequence of the *simR-simX* intergenic region showing the *simRp* and *simXp* transcription start points and putative -10 sequences, the *simR* and *simX* ribosome binding sites (RBS), the extent of the SimR DNaseI footprints on the O_X and O_R operators, and the imperfect inverted repeats within the footprints that may represent SimR binding motifs.

(*luxAB*) reporter plasmid pIJ5972 (Aigle *et al*., 2000; M. Paget, pers. comm.). Fragments carrying *simRp* and *simXp* were individually cloned into pIJ5972, and the resulting reporter constructs were introduced into *S. lividans* in order to probe *simR* and *simX* promoter activities in the absence of SimR. To measure *simR* and *simX* promoter activities in the presence SimR, an integrative plasmid (pIJ10469) carrying *simR* under the control of its own promoter was introduced into the strains already harbouring the promoter-probe plasmids. Transformants were grown on Difco Nutrient Agar to promote vegetative growth and delay aerial hyphal formation, which may interfere with diffusion of the luciferase substrate (n-decanal) and with light emission. **Fig. 4.6** shows that *simX* and *simR* promoter activities were repressed 100-fold and 12-fold, respectively, in the presence of *simR*.

4.5 Purified SimR binds to the *simR-simX* intergenic region at two distinct operator sites

The *lux* reporter data suggested that SimR is a repressor that regulates its own expression as well as that of *simX*. To test this idea, I monitored SimR binding to the *simR-simX* intergenic region by electrophoretic mobility shift assay (EMSA). An N-terminally His_6-tagged derivative of SimR was overexpressed in *E. coli* and purified to homogeneity (**Fig. 4.7**). Increasing concentrations of SimR were incubated with a radiolabelled probe spanning the *simR-simX* intergenic region and the complexes were resolved on a native gel. Purified SimR bound to the intergenic region at concentrations as low as 0.23 nM and, as the concentration of SimR increased, two sets of shifted protein-DNA complexes became evident, suggesting that there are two SimR binding sites in the *simR-simX* intergenic region (**Fig. 4.8A**).

DNaseI footprinting on both DNA strands was used to map precisely the SimR operator sites within the *simR-simX* intergenic region. Two separate SimR binding sites were observed, consistent with the two shifted species seen in the EMSA experiments: the operator closer to *simX* was designated O_X and the one closer to *simR* was designated O_R (**Figs. 4.5 and 4.9)**. Within the O_X and O_R footprints we identified imperfect inverted repeats that may represent the binding sequences for the

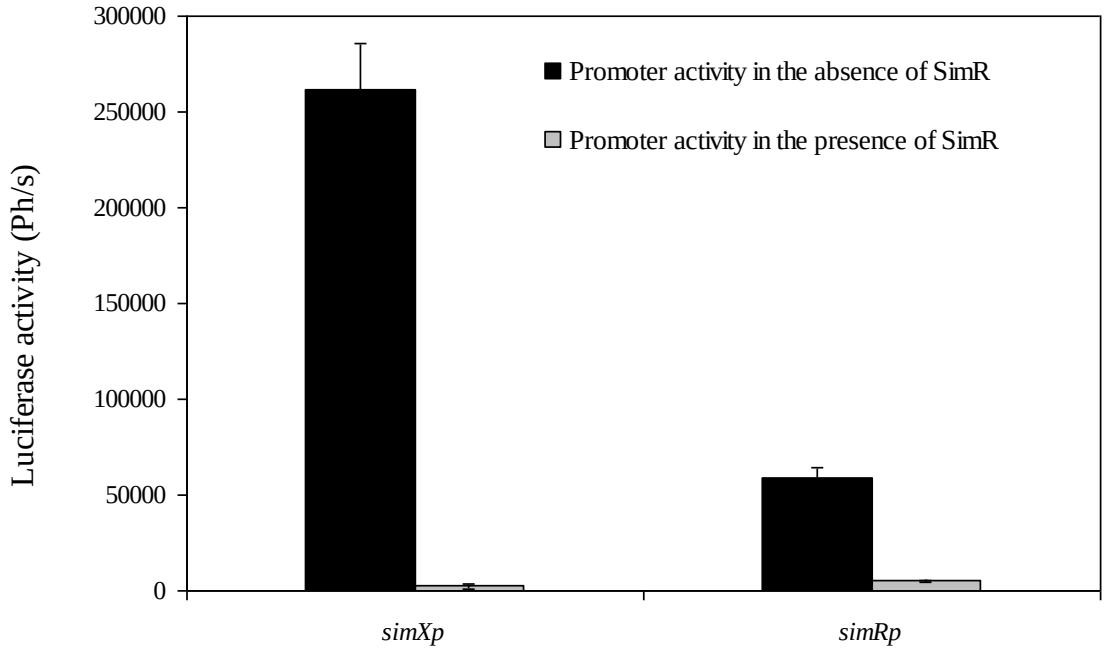


Fig. 4.6. Promoter activities of *simR* and *simX* in the presence and absence of *simR*. *simRp-luxAB* and *simXp-luxAB* transcriptional fusions were created in an integrative luciferase promoter-probe vector (pIJ5972) and assayed in *S. lividans*, in either the presence or absence of *simR*. Plasmid-containing *S. lividans* strains were grown on Difco Nutrient Agar in single wells of a 96-well microtiter plate (Sterilin) for three days. Each well was inoculated with approximately $5x10^4$ spores. Plates were exposed to filter paper impregnated with n-decanal for 5 min and luciferase activities were quantified using a NightOwl camera (Berthold) equipped with WinLight software (Berthold) using a 1 min exposure time. Values given correspond to the average of three biological replicates from three different spore stocks and standard errors are shown.

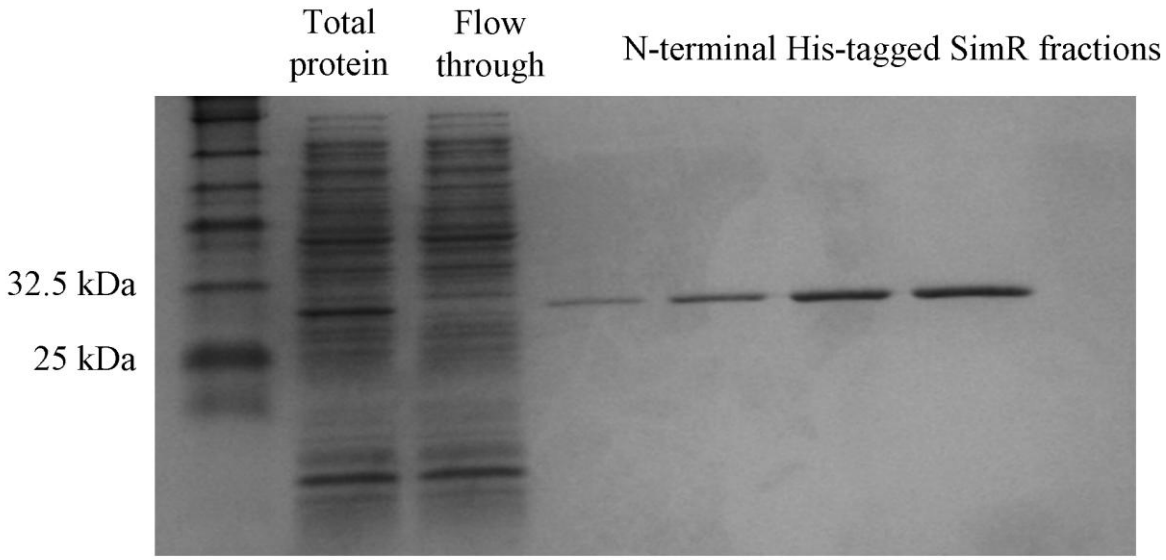


Fig. 4.7. Purification of N-terminally His-tagged SimR to apparent homogeneity.

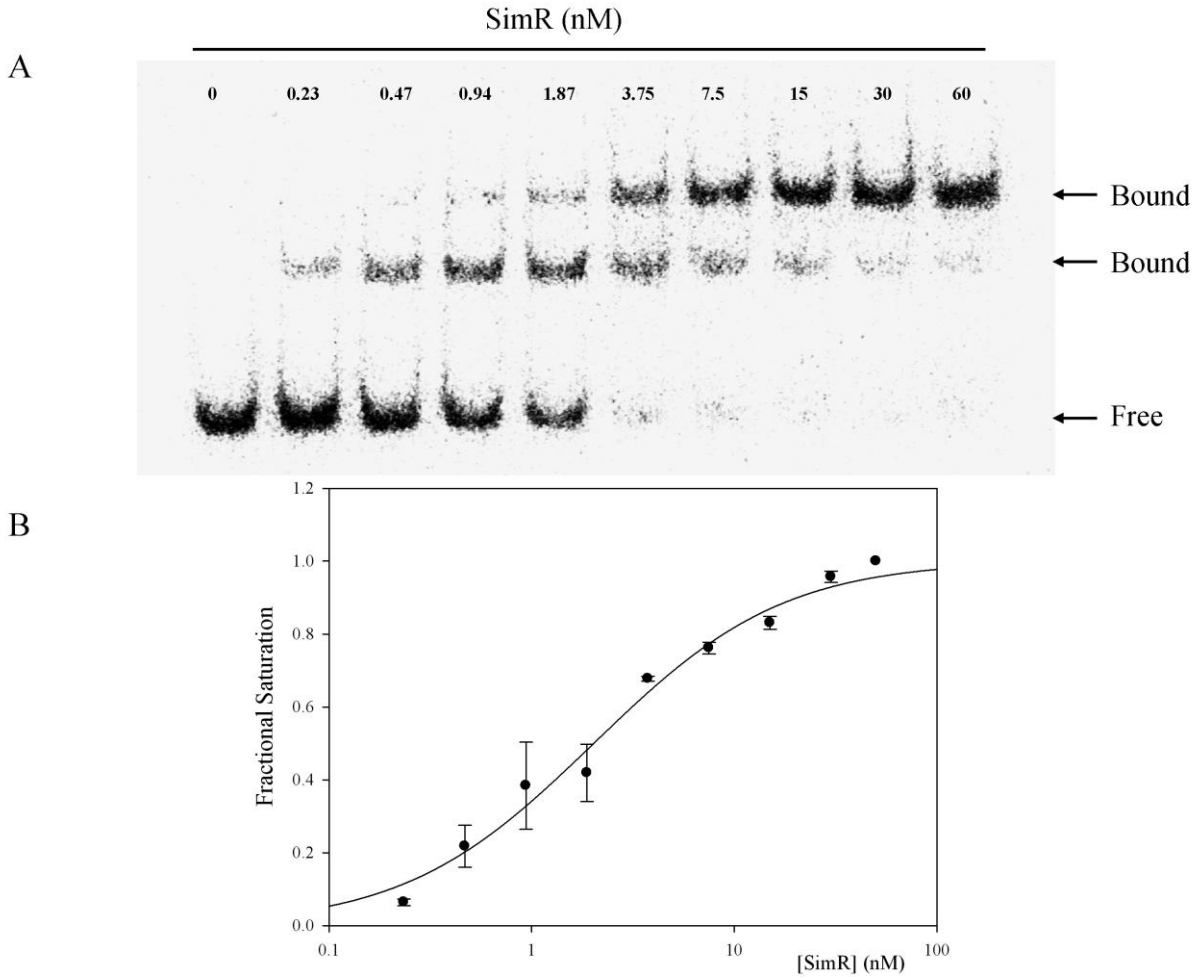


Fig. 4.8. (A) Electrophoretic mobility shift assay showing binding of purified SimR to the *simR-simX* intergenic region. Bands corresponding to protein-DNA complexes (Bound) and free DNA (Free) are indicated. The final concentration of SimR is indicated above each lane. (B) Saturation curve of the data from EMSA experiments. EMSA data were collected and analysed on a PhosphoImager (FujiFilm) using Multi Gauge image analysis software (FujiFilm). Two independent EMSAs were carried out [one of which is shown in A] and the mean values calculated. Standard errors are shown. Saturation curves (Saturation fraction against concentration of protein) were fitted with SigmaPlot (See Experiment Procedures) to determine K_ds. In all EMSA experiments, SimR was present in molar excess over the probe.

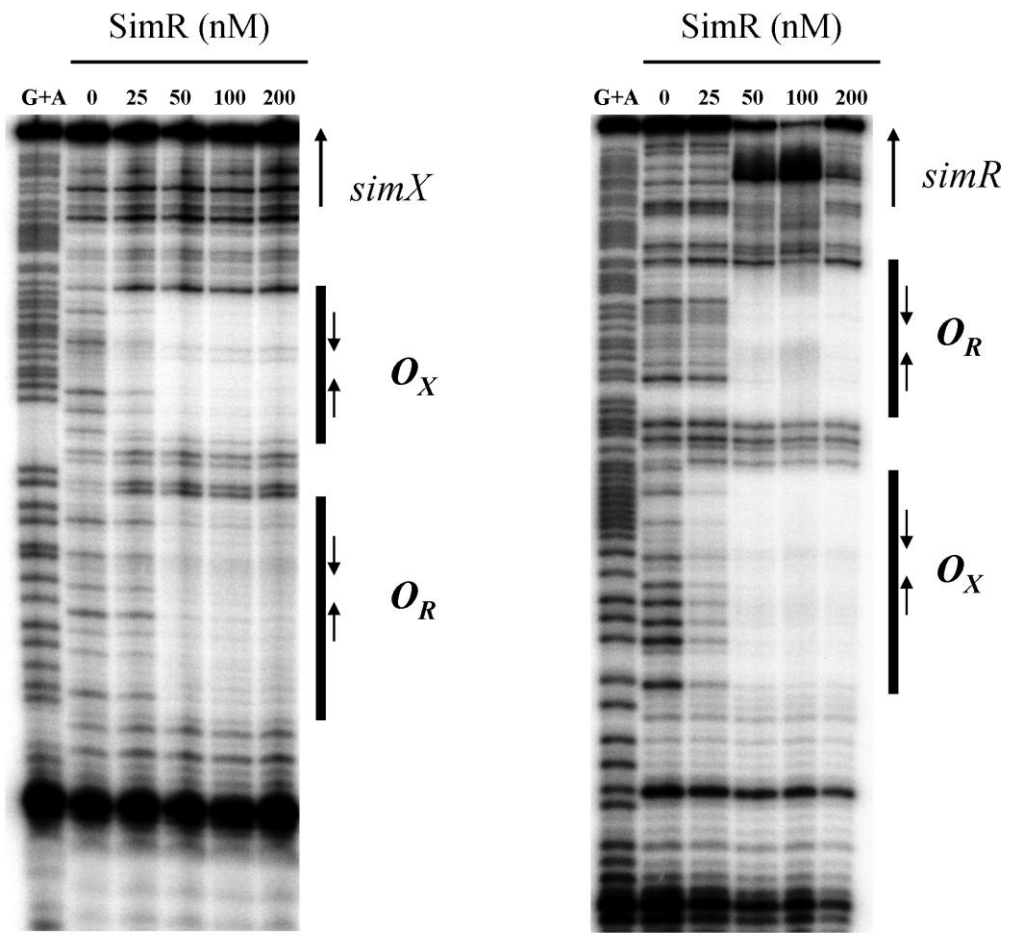


Fig. 4.9. DNaseI footprinting analysis of SimR binding to the *simR-simX* intergenic region. A DNA fragment containing the *simR-simX* intergenic region, 5'-end labelled on either the upper strand (left panel) or the lower strand (right panel), was exposed to DNaseI in the presence of increasing concentrations of SimR. The sequencing ladders were generated by subjecting the probes to Maxam-Gilbert G+A chemical sequencing. Regions protected from DNaseI cleavage (operators O_X and O_R) are indicated with vertical bars and inverted repeats within the DNaseI protected regions are indicated by convergent arrows; these features are also highlighted on the DNA sequence in **Fig. 4.5.**

SimR homodimer. **Fig. 4.5** shows these imperfect inverted repeats, the extent of the SimR DNaseI footprints on the sequence of the *simR-simX* intergenic region and the positions of the *simRp* and *simXp* transcription start points and putative -10 promoter sequences in relation to the two SimR operators.

4.6 SimR binding to the two operators is non-cooperative

In the DNaseI footprinting analysis, O_X was occupied at a lower concentration of SimR than was O_R, suggesting that O_X has a higher affinity for SimR than O_R. Competitive EMSA was employed to explore this issue further. In these experiments, unlabelled fragments containing either O_X or O_R were used to compete with a radioactively labelled *simR-simX* intergenic fragment containing both O_X and O_R. The final concentration of SimR used in the competitive EMSA experiments was set at 20 nM such that all the labelled probe was in complex with SimR in the absence of unlabelled competitor DNA. A thousand-fold excess of unlabelled O_X-containing fragment out-competed the labelled intergenic probe (no complex formation between SimR and the labelled intergenic probe) (**Fig. 4.10**). However, the same excess of O_R-containing fragment could not completely abolish SimR complex formation with the labelled intergenic probe (**Fig. 4.8**), confirming that SimR binds O_X more tightly than O_R.

Using the EMSA data shown in **Fig. 4.8**, we determined the approximate equilibrium dissociation constants (K_ds) for the two complexes to be 1.2 ± 0.4 nM for SimR-O_X and 3.5 ± 1.4 nM for SimR-O_R. In order to determine whether there is cooperativity between SimR binding at O_X and O_R, we also determined the K_ds of each SimR-operator complex by EMSA using probes containing only O_X or only O_R (data not shown), instead of the full *simR-simX* intergenic region. The K_ds were found to be 0.9 ± 0.2 nM for SimR-O_X and 3.6 ± 0.3 nM for SimR-O_R. The K_ds for the two SimR-operator complexes did not change substantially when the two operators were separated, suggesting that SimR binding to its two operators is non-cooperative.

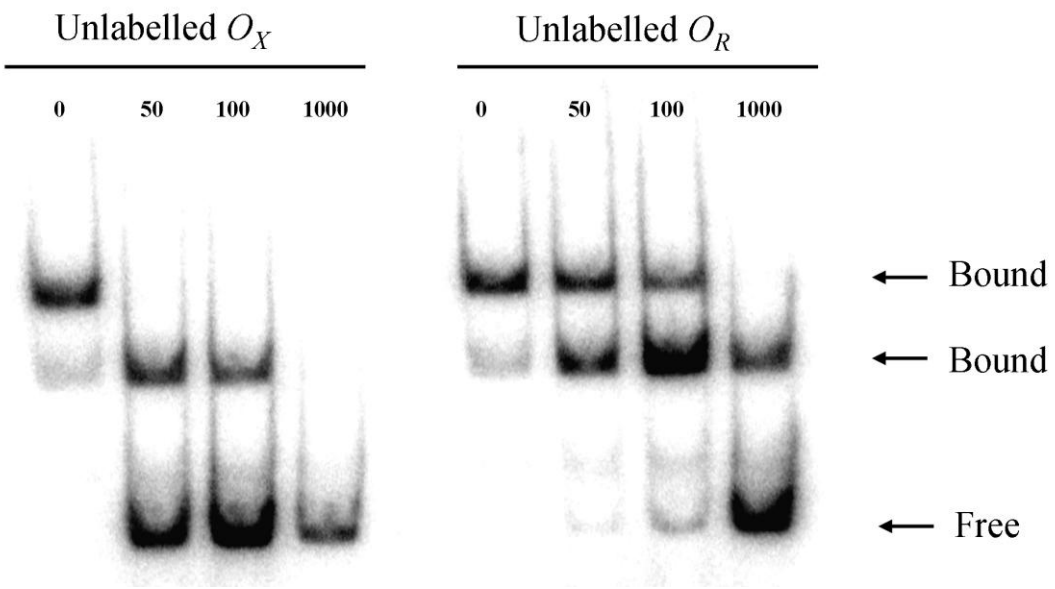


Fig. 4.10. Competitive electrophoretic mobility shift assay comparing the binding affinity of SimR to O_X and O_R. All lanes contain SimR at a final concentration of 20 nM and a constant amount of a radiolabelled *simR-simX* intergenic probe that carries both O_X and O_R. The molar excess (over the radiolabelled probe) of a competing unlabelled fragment containing either O_X or O_R is indicated above each lane.

4.7 Exogenous SD8 induces expression of the SimX efflux pump *in vivo*

In order to determine whether the *simR* and *simX* promoters respond to SD8 *in vivo*, I used the *S. lividans luxAB* reporter system to measure *simRp* and *simXp* activities in the presence of SimR and in response to exogenously added antibiotic. **Fig. 4.11A** shows the response curve of *simR* and *simX* promoter activity to increasing concentrations of SD8. Both promoters were induced by simocylinone D8, suggesting that the antibiotic can relieve SimR-mediated repression of both *simRp* and *simXp*. 0.6 μg/ml was the highest concentration tested because *S. lividans* is sensitive to simocylinone D8 (MIC = 2 μg/ml).

4.8 SD8 dissociates SimR from the *simR-simX* intergenic region

To determine if SimR responds directly to SD8, I examined the effect of the antibiotic on SimR-operator complex formation by EMSA (**Fig. 4.12A**). The SimR concentration was held constant and an increasing concentration of SD8 was introduced into the binding reaction. As the SD8 concentration increased, there was a progressive decrease in SimR-DNA complex formation and a concomitant liberation of free probe. 62.5 μM SD8 was sufficient to dissociate the SimR-DNA complexes almost completely. This effect was not due to DMSO, the SD8 solvent, as equivalent amounts of pure DMSO had no effect on SimR-DNA complex formation. Further, to test specificity, we examined the effect of SD8 on the DNA-binding activity of ActR, a TetR homologue that regulates actinorhodin export in *S. coelicolor* (see Discussion) and found that 100 μM SD8 had no effect on ActR binding to its cognate operator (N.B. this experiment was carried out by Sang Kyun Ahn, Department of Biochemistry and Biomedical Sciences, McMaster University, Canada). These results show that SD8 is able to specifically disrupt SimR-DNA complex formation.

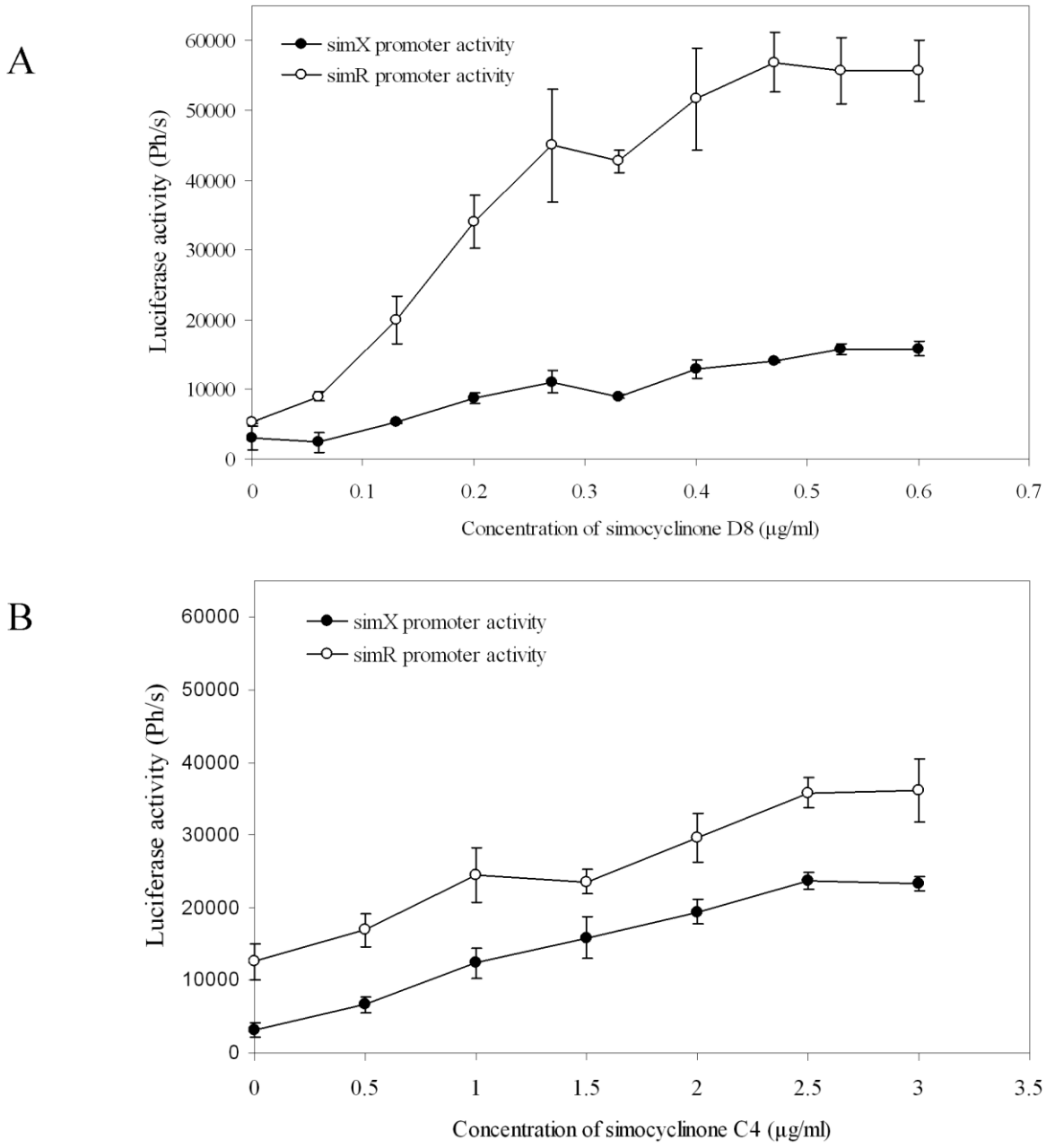


Fig. 4.11. Induction of the *simR* and *simX* promoters *in vivo* by (A) SD8 and (B) the biosynthetic intermediate SC4. *S. lividans* containing pIJ10469 (carrying *simR* under its native promoter) together with luciferase promoter-probe plasmids pIJ10465 (*simXp-luxAB*) or pIJ10466 (*simRp-luxAB*) were assayed. Values given correspond to the average of three biological replicates from three different spore stocks and standard errors are shown. For further details, see the legend to **Fig. 4.6.**

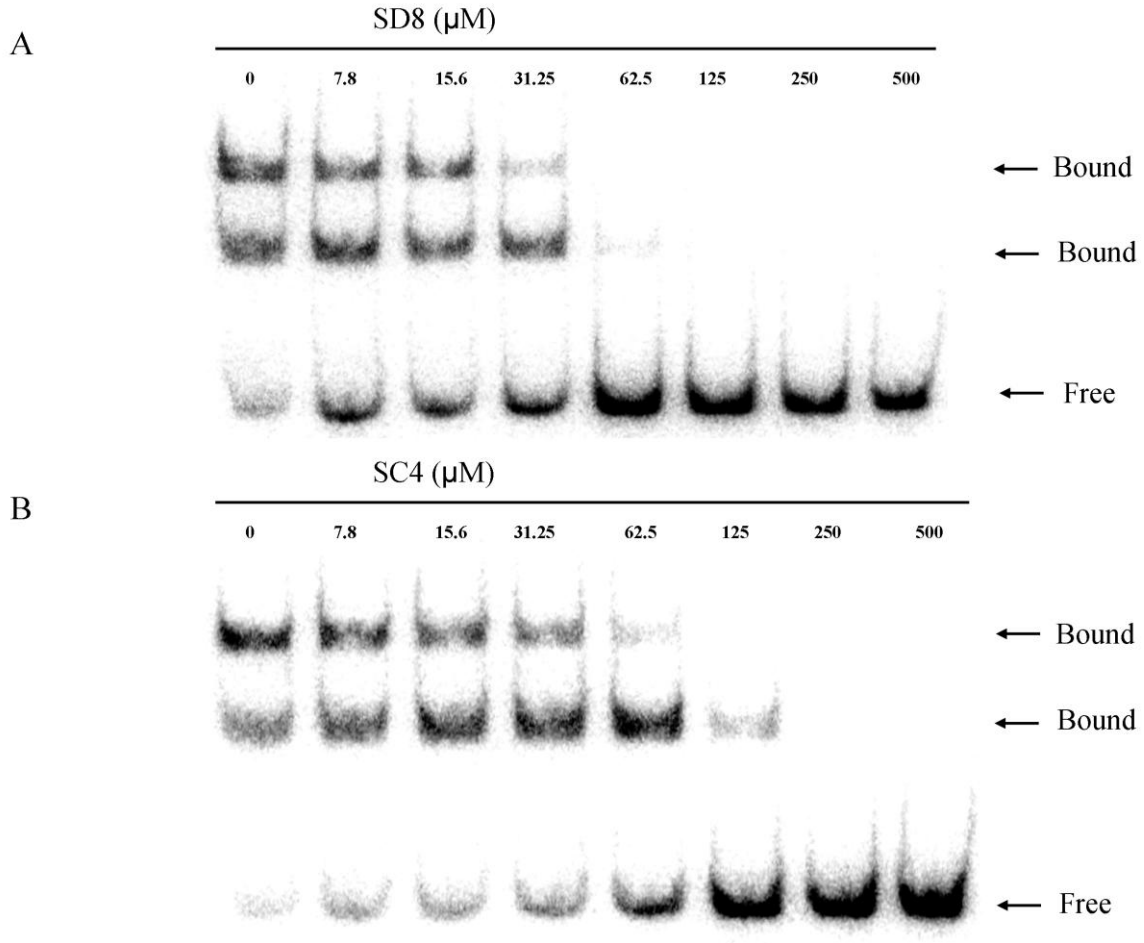


Fig. 4.12 Electrophoretic mobility shift assays showing that (A) SD8 and (B) its biosynthetic intermediate SC4 abolish SimR DNA-binding activity. All lanes contain a constant amount of SimR and radiolabelled *simR-simX* intergenic fragment containing both O_X and O_R. Bands corresponding to protein-DNA complexes (Bound) and free DNA (Free) are indicated. The final concentrations of SD8 or SC4 are indicated above each lane.

4.9 An intermediate in the simocyclinone biosynthetic pathway induces *simX in vivo* and dissociates SimR from its operators *in vitro*

In addition to SD8, I also tested the ability of SC4, a natural intermediate in the D8 biosynthetic pathway, to induce the *simR* and *simX* promoters *in vivo* and to dissociate SimR from its operators *in vitro*. SC4 lacks the aminocoumarin ring present in the mature antibiotic (**Fig. 4.1**) and is essentially inactive as a DNA gyrase inhibitor; the SD8 IC_{50} = 0.1 μM, whereas the SC4 IC_{50} > 100 μM (Edwards *et al*., 2009). SC4 induced the *simR* and *simX* promoters *in vivo*, although somewhat more weakly than the mature antibiotic (**Fig. 4.11B**). Because SC4 is not an active antibiotic, I was able to test higher concentrations than for SD8. Consistent with the *in vivo* inductions, SC4 also caused SimR to dissociate from its binding sites *in vitro*, although again somewhat more weakly than SD8; 250 μM SC4 was sufficient to abolish all complex formation (**Fig. 4.12B**).

4.10 Discussion

The work contained in this chapter shows that *simX* encodes a SD8 efflux pump and that the *simX* promoter is directly repressed by SimR, which binds to two operator sites in the *simR-simX* intergenic region. SD8 abolishes DNA binding by SimR, coupling the biosynthesis of simocyclinone to its export.

SD8 is a potent inhibitor of DNA gyrase supercoiling (IC_{50} = 0.1 μM) (Flatman *et al*., 2005; Oppegard *et al*., 2009; Edwards *et al*., 2009). The recent structure of the antibiotic bound to the DNA gyrase A subunit shows that the two moieties at the ends of SD8, the aminocoumarin ring and the angucyclic polyketide, bind to two separate, well-defined pockets within the GyrA DNA-binding saddle, linked by the intervening tetraene linker and D-olivose moieties (Edwards *et al*., 2009). Given the prominence of the aminocoumarin ring in the overall binding of the antibiotic to the GyrA subunit, it is not surprising that the biosynthetic intermediate SC4, which lacks the aminocoumarin ring (**Fig. 4.1**), is essentially inactive as a DNA gyrase inhibitor (IC_{50} >100 μM). Importantly, however, although SC4 is not a DNA gyrase inhibitor, it can efficiently derepress SimR *in vivo* and *in vitro*.

The most striking analogy to the research presented here is the work of Nodwell and colleagues on the regulation of efflux of actinorhodin (Tahlan *et al*., 2007, 2008; Ahn *et al*., 2007; Hopwood, 2007; Willems *et al.*, 2008). Actinorhodin is a blue-pigmented, six-ring polyketide antibiotic made by *S. coelicolor* (Bystrykh *et al*., 1996). Within the actinorhodin biosynthetic cluster are two co-transcribed genes, *actA* and *act*II-ORF3, encoding integral membrane proteins implicated in actinorhodin export (Caballero *et al*., 1991; Fernandez-Moreno *et al*., 1991). Expression of these two genes is regulated by a TetR-like protein, ActR, the product of the adjacent, divergently transcribed gene. Nodwell and colleagues have been characterising ligands that relieve repression by ActR. Importantly, they showed that, in addition to the mature six-ring antibiotic, three-ring intermediates from the biosynthetic pathway also relieve repression by ActR. From this, they suggested that the ability of actinorhodin intermediates to relieve repression by ActR might provide a 'feed-forward' mechanism that would ensure expression of the ActA efflux pump prior to the build-up of a toxic concentration of the mature antibiotic (Tahlan *et al*., 2007, 2008; Ahn *et al*., 2007; Hopwood, 2007; Willems *et al.*, 2008). Similarly, the ability of an inactive simocyclinone intermediate to relieve repression by SimR might also act as a biosynthetic checkpoint to ensure feed-forward regulation of simocyclinone export.

As applied to simocyclinone (or actinorhodin) export, the feed-forward hypothesis is speculative. I have shown that the pathway intermediate SC4 induces *simX* expression *in vivo* when applied exogenously. However, for a feed-forward mechanism to operate in the producing organism, SC4 or other SimR-binding pathway intermediates would have to accumulate in the cytoplasm to a concentration high enough to trigger *simX* expression, and the cytoplasmic concentrations of pathway intermediates are hard to determine experimentally. However, there are other examples in the literature where the TetR-like protein blocks expression not only of the exporter gene, but also of late biosynthetic genes, and in these cases it seems that an intermediate in the pathway must be responsible for inducing expression of these genes (i.e. feed-forward activation), since induction of the late biosynthetic enzymes is required to generate the mature antibiotic.

In the landomycin A producer, *Streptomyces cyanogenus*, a TetR-family regulator, LanK, represses expression of *lanJ*, encoding a putative landomycin A efflux pump (Ostash *et al*., 2008). LanK is derepressed by mature landomycin A, which carries a hexasaccharide chain, but also by intermediates in the pathway that carry only a pentasaccharide or a trisaccharide chain (Ostash *et al*., 2008). However, *lanJ* is co-transcribed with several downstream biosynthetic genes involved in late glycosylation steps. As a consequence, LanK couples production of intermediates not only to assembly of the export machinery, but also to expression of late biosynthetic enzymes that attach the final sugars to produce mature landomycin A (Ostash *et al*., 2008). A more complex example concerns the biosynthesis of the clinically important anti-cancer agents daunorubicin and doxorubicin made by *Streptomyces peucetius*. In this system, binding of daunorubicin-doxorubicin pathway intermediates like rhodomycin D appears to de-repress the TetR-like regulator DnrO, activating a cascade involving two further transcription factors, DnrN and DnrI, that leads to expression of the resistance genes and of late biosynthetic genes (Otten *et al*., 1995; Jiang and Hutchinson, 2006).

In addition to repressing the *simX* promoter, SimR also directly negatively regulates its own expression. In other systems, negative autoregulation has been shown to confer specific functions that are absent in systems that have simple regulation. Many of these studies on the design principles of genetic circuits that incorporate negative autoregulation have exploited TetR, the founding member of the family to which SimR belongs (Becskei and Serrano, 2000; Rosenfeld *et al*., 2002; D. Madar and U. Alon, pers. comm.). These studies show that negative autoregulatory feedback loops in gene circuits can provide stability, thereby limiting stochastic fluctuations in the system (Becskei and Serrano, 2000). In addition, negative autoregulation can speed the response time of the transcription network to a stimulus (Rosenfeld *et al*., 2002), and it can broaden the dynamic range of the input signal to which the downstream genes respond (D. Madar and U. Alon, pers. comm.).

Chapter 5

Structures of the TetR-like simocyclinone efflux pump repressor, SimR, and the mechanism of ligand-mediated derepression

5.1 Introduction

TFRs function as homodimers. Each monomer consists of two domains, an N-terminal DBD containing a HTH motif, and a C-terminal LBD (Ramos *et al*., 2005). The LBDs are extremely diverse in amino acid sequence and it is therefore not generally possible to model these domains based on existing structures with any confidence, nor to predict potential ligands (Ramos *et al*., 2005; Yu *et al*., 2010). In addition to TetR, the structures of several other TFRs have been determined with relevant ligands, and also with bound DNA, enabling their mechanisms of repression and derepression to be elucidated (Chapter 1). In those cases where derepression has been analysed by comparing DNA-bound and ligand-bound structures, it is brought about through the stabilisation, by the binding of the ligand, of a conformation in which the recognition helices are too far apart to simultaneously bind to adjacent major grooves of DNA. These mechanisms necessarily invoke allosteric communication between the LBDs and DBDs.

SD8 (**Fig. 4.1**) is a potent DNA gyrase inhibitor produced by *S. antibioticus*, with an IC_{50} ~ 0.1 μM for inhibition of DNA supercoiling (Schimana *et al*., 2000; Flatman *et al*., 2005; Edwards *et al*., 2009). SD8 is exported from the producing organism by a specific efflux pump, SimX, and the transcription of *simX* is repressed by SimR, a TFR that binds to two distinct operators in the intergenic region between the divergently transcribed *simR* and *simX* genes (Chapter 4 and Le *et al*., 2009). SD8 abolishes DNA-binding by SimR, inducing expression of the SimX efflux pump, and this provides an intimate mechanism that couples the biosynthesis of simocyclinone to its export (Chapter 4 and Le *et al*., 2009).

In order to gain insight into the molecular mechanism of ligand-mediated SimR derepression, I have solved the crystal structures of SimR, alone and in complexes with SD8 and its biosynthetic intermediate SC4, which lacks the AC moiety of the mature antibiotic (**Fig. 4.1**), but still derepresses SimR, albeit less efficiently (Chapter 4 and Le *et al*., 2009). The structures show an extensive ligand-binding pocket spanning both monomers in the functional dimeric unit and suggest a derepression mechanism that involves rigid-body motions of the subunits relative to

one another, coupled with a putative locking mechanism to restrict further conformational change.

5.2 The structure of ligand-free SimR

The structure of ligand-free SimR (SimR-apo) was determined at 1.95-Å resolution. X-ray data collection and refinement statistics are summarised in **Table 5.1**. There is a single copy of the SimR monomer in the asymmetric unit (ASU) and the final model comprises residues 7-244 of the 259-residue native sequence. A biologically-relevant homodimer is generated through the application of 2-fold crystallographic symmetry (**Fig. 5.1**). The secondary structure of SimR is almost entirely α-helical, being composed of 11 helices that span residues 29-43 (α1), 50-57 (α2), 61-67 (α3), 71-84 (α4), 96-113 (α5), 116-120 (α6), 130-144 (α7), 150-181 (α8), 185-202 (α9), 206-216 (α10) and 222-243 (α11) (**Figs. 5.1** and **5.2**). There are, in addition, two short β-strands near the N-terminus (residues 18-19 and 26-27), which form a β-hairpin. The SimR monomer can be divided into two domains, an N-terminal DNA-binding domain (DBD) and a C-terminal ligand-binding domain (LBD). The DBD consists of the β-hairpin and helices α1 to α3. Helices α2 and α3 form a helix-turn-helix (HTH) motif that is stabilized by packing against α1. By analogy with other TFRs, the HTH is predicted to interact with the major groove of DNA, with the recognition helix, α3, providing the majority of base-specific interactions. In addition, the N-terminal β-hairpin motif of SimR appears to be ideally placed to make minor groove interactions. Consistent with this proposal, there are a total of four Arg residues in the range 18-25 (**Fig. 5.2**), which could interact favourably with the negatively charged phosphate backbone of DNA. These residues contribute to the prominent electropositive surface patch that overlaps the recognition helix in SimR-apo (**Fig. 5.3**).

The last eight helices (α4-α11) make up the LBD responsible for small-molecule ligand binding and forming the interface between the two subunits of the dimer. Dimerization occurs largely through a pair of helices, α8 and α11, from each SimR monomer that form a four-helix bundle (**Fig. 5.1**). The dimerization interface is further enhanced by the α9-α10 segment, which forms an extended 'arm' that wraps around the LBD of the opposing monomer (**Fig. 5.1**) (Note that the distinctive kink

Data set[a]	Native apo (N-tag)	SeMet apo (N-tag)	Native apo (C-tag)	SD8 complex (C-tag)	SC4 complex (C-tag)
Data collection					
Space Group	H32	H32	H32	C2	$P6_522$
Cell parameters (Å/°)	a = b = 116.62, c = 110.58	a = b = 116.11, c = 110.35	a = b = 116.06, c = 109.38	a = 147.37, b = 83.91, c = 45.47, β = 90.94	a = b = 86.57, c = 349.86
Solvent content (%)	45	45	48	47	61
Beamline[b]	I03	I03	I02	I04	I02
Wavelength (Å)	0.9763	0.9763	0.9700	0.9400	0.9700
Resolution range[c] (Å)	48.50 - 2.30 (2.42 - 2.30)	48.37 - 3.40 (3.58 - 3.40)	48.04 - 1.95 (2.06 - 1.95)	42.39 - 2.30 (2.42 - 2.30)	63.06 - 2.30 (2.42 - 2.30)
Unique reflections[c]	12693 (1786)	4076 (585)	20761 (3005)	23951 (2988)	34449 (4983)
Completeness[c] (%)	98.0 (95.2)	99.9 (100.0)	99.8 (100.0)	97.0 (83.4)	97.3 (98.5)
Redundancy[c]	5.0 (3.8)	15.2 (15.7)	7.8 (7.9)	2.9 (2.3)	8.4 (8.5)
R_{merge}[c, d]	0.093 (0.523)	0.158 (0.422)	0.070 (0.570)	0.077 (0.229)	0.109 (0.372)
Mean I/σ(I)[c]	10.2 (2.3)	14.4 (6.8)	20.6 (3.8)	10.3 (3.3)	15.4 (5.4)
Wilson B value ($Å^2$)	57.9	78.4	31.8	30.7	28.8
Refinement					
Reflections: working/free[e]			19717/1044	22723/1227	32723/1726
R_{work}[f]	-	-	0.205	0.200	0.198
R_{free}[f]	-	-	0.234	0.270	0.246
Coordinate error[g] (Å)	-	-	0.149	0.265	0.210
Ramachandran favoured/allowed[h] (%)	-	-	97.2/98.8	98.2/99.8	99.0/100.0
Ramachandran outliers[h]	-	-	3	1	0
rmsd bond distances (Å)	-	-	0.018	0.017	0.017
rmsd bond angles (°)	-	-	1.57	1.46	1.58
Contents of model	-	-			
Protein residues in each chain (totals in brackets)	-	-	A: 7-244 (238)	A: 5-11, 27-246 (227) B: 8-10, 28-247 (223)	A: 6-247 (242) B: 5-15, 26-246 (230)

SD8/SC4 molecules	-	-	0	2	2
Polyethylene glycol molecules	-	-	0	0	2
Calcium ions	-	-	0	0	1
Chloride ions	-	-	0	6	8
Water molecules	-	-	121	164	277
Temperature factors (Å^2)					
Main chain atoms			36.9	24.4	19.3
Side chain atoms			39.9	25.6	21.6
SD8/SC4 molecules			-	23.8	23.7
Polyethylene glycol molecules			-	-	37.2
Calcium ions			-	-	24.3
Chloride ions			-	60.0	37.9
Water molecules			20.4	22.8	25.1
Overall			37.3	24.9	21.0
PDB accession code	-	-	2Y2Z	2Y30	2Y31

[a] Protein was either N-terminally (N-tag) or C-terminally (C-tag) His tagged.

[b] I02, I03, I04 = beamlines at the Diamond Light Source (Oxfordshire, UK).

[c] The figures in brackets indicate the values for outer resolution shell.

[d] $R_{merge} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ is the *i*th observation of reflection *hkl* and $\langle I(hkl)\rangle$ is the weighted average intensity for all observations *i* of reflection *hkl*.

[e] The data sets were split into "working" and "free" sets comprising 95% and 5% of the data, respectively. The free set was not used for refinement.

[f] The R-factors R_{work} and R_{free} are calculated as follows: $R = \sum(| F_{obs} - F_{calc} |)/\sum| F_{obs} | \times 100$, where F_{obs} and F_{calc} are the observed and calculated structure factor amplitudes, respectively.

[g] Estimate of the overall coordinate errors calculated in REFMAC5 based on R_{free} (Murshudov *et al*., 1997).

[h] As calculated using MOLPROBITY (Chen *et al*., 2010).

Table 5.1. Summary of SimR X-ray data and model parameters

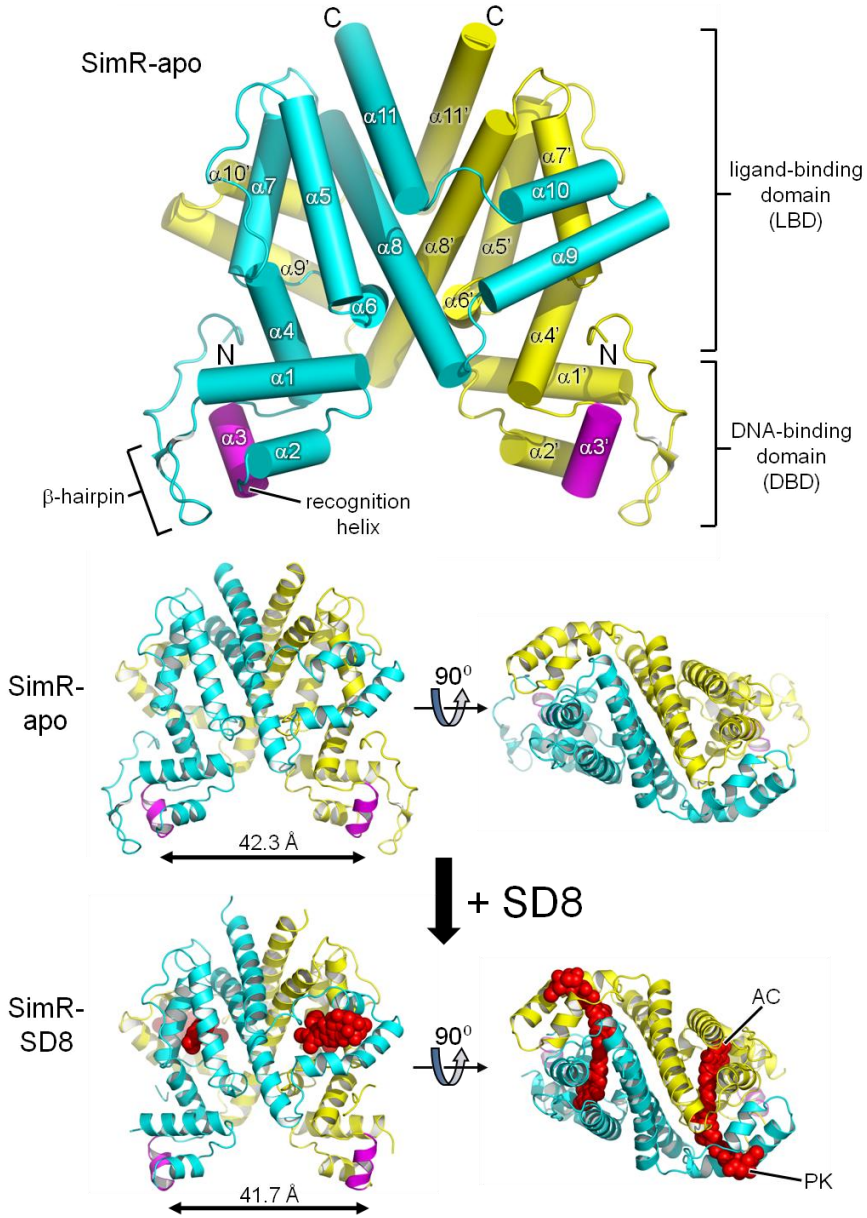


Fig. 5.1. Crystal structures of ligand-free (SimR-apo) and SD8 bound (SimR-SD8) forms of SimR. The top panel uses a cylindrical helix representation to highlight the secondary structure of SimR-apo with key features labelled. The two subunits of the homodimer are distinguished by cyan and yellow colouration, and the two recognition helices are shown in magenta. Where appropriate this colour scheme is adopted throughout the subsequent figures. Note the α9-α10 arm that engages with the opposing subunit. The lower panels compare SimR-apo and SimR-SD8 in cartoon representation viewed from the front (left) and from the top (right), with the SD8 molecules shown in red space-filling representation. The separations of the recognition helices are indicated. The labels AC and PK refer to the aminocoumarin and polyketide moieties of SD8, respectively.

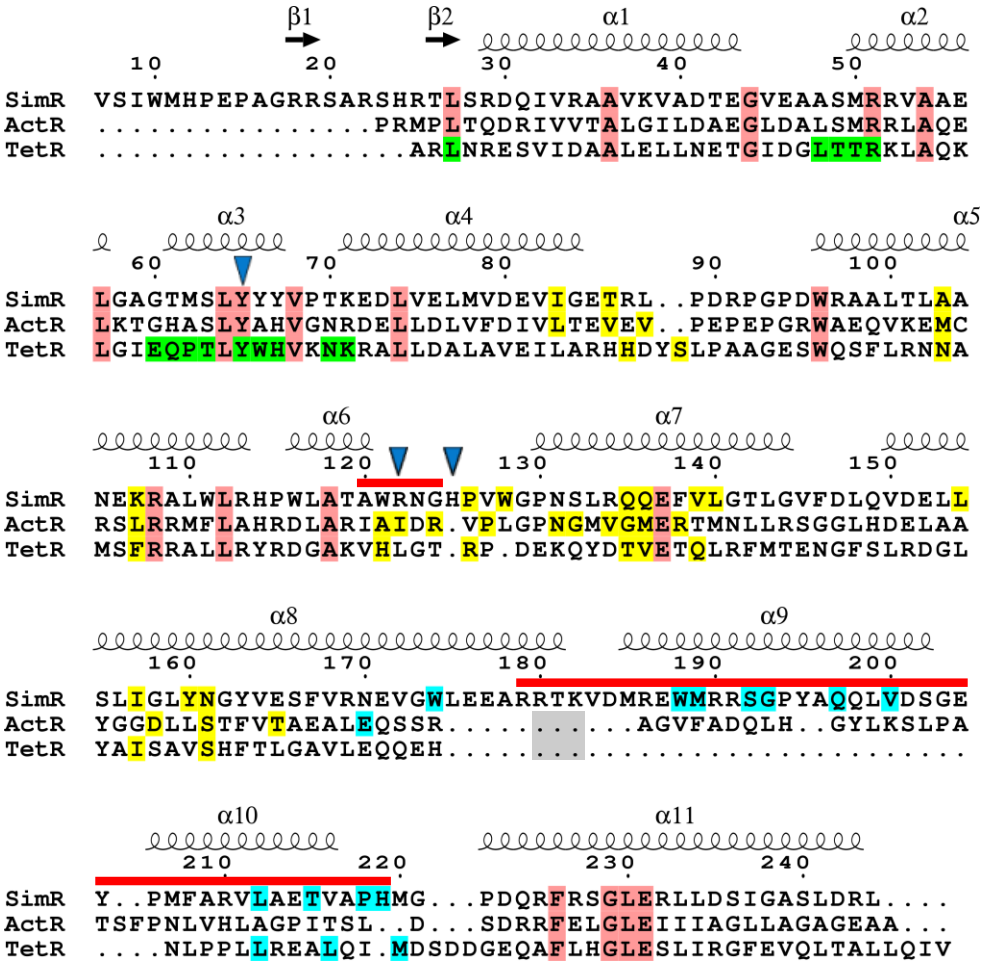


Fig. 5.2. Structure-based multiple sequence alignment of SimR with *S. coelicolor* ActR and *E. coli* TetR. The initial alignment was generated using the SSM server (Krissinel and Henrick, 2004) with the ligand-free structures (PDB: 2OPT and 1BJZ for ActR and TetR, respectively). This was subsequently adjusted manually with reference to the superposed structures, and then displayed using ESPript (Gouet *et al.*, 1999). Strictly conserved residues are highlighted with pink shading. Secondary structure elements for SimR are shown above the alignment, where α = α-helix, β = β-strand. Residues involved in binding SD8 (selected on the basis of being < 4.0 Å from the ligand) are highlighted with yellow and cyan shading to denote the subunit. Similarly, residues involved in actinorhodin binding by ActR (PDB: 3B6A) and tetracycline binding by TetR (PDB: 1BJY) are highlighted with the same colour scheme. Green shading indicates residues involved in DNA binding by TetR (PDB: 1QPI). Grey shading indicates regions in ActR (residues 178-187) and TetR (residues 152-164) that were not modelled in the ligand-free structures. The red bars denote regions of SimR that show the most significant conformational changes between the ligand-bound and ligand-free structures. The inverted blue triangles mark key residues referred to in the text.

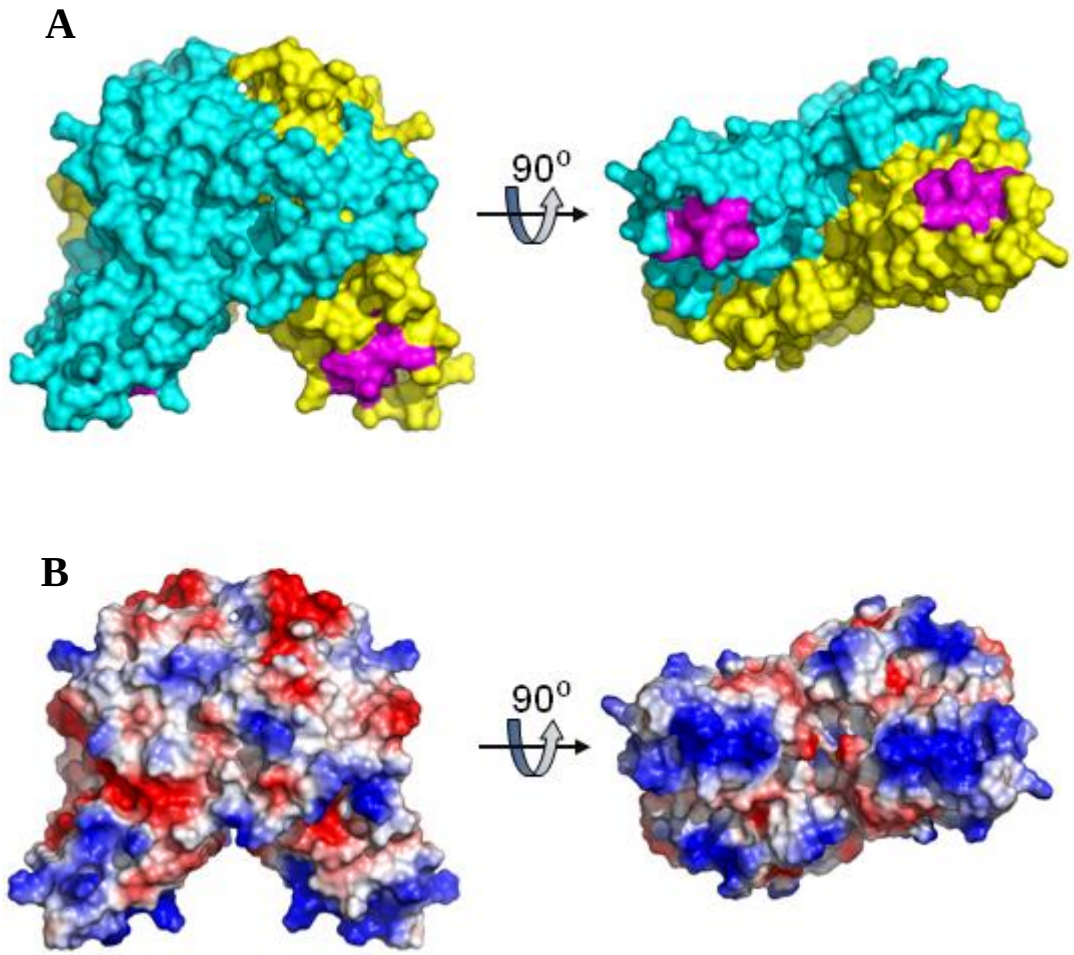


Fig. 5.3. Molecular surface representations of SimR-apo. (A) Molecular surface coloured as for **Fig. 5.1**, highlighting the arrangement of the subunits and the presentation of the recognition helices (magenta). (B) Molecular surface showing electrostatic surface potentials (red = electronegative, white = neutral, blue = electropositive). The left-hand panels show the front view and the right-hand panels show the bottom view. Note the overlap between the recognition helices and distinct electropositive regions. For this figure, all missing side chains were modelled in energetically favourable conformations (applies to Arg19, Arg25, Arg91 and Arg243 in both subunits).

in α9 is important for ligand binding; see below). In fact, all of helices α7-α11 from each subunit are involved to varying degrees in dimerization. Overall, this gives rise to an extensive dimer interface of 2475 $Å^2$ accounting for 17.5% of the solvent-accessible surface of an isolated monomer, as calculated by the Protein Interactions, Surfaces and Assemblies server (PISA; http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html) (Krissinel and Henrick, 2005). Within the core of each subunit, there is a substantial cavity (**Fig. 5.4A**). The two cavities of the dimer are linked via a narrow pore that passes through the centre of the dimer, and they each connect to bulk solvent through channels that pass between the two helices of the α9-α10 arm (**Fig. 5.4A**).

5.3 Comparison of the SimR structure with other TetR family members

The closest structural homologues of SimR were identified using the DALI server (http://ekhidna.biocenter.helsinki.fi/dali_server/) (Holm and Sander, 1995) (**Fig. 5.5**). This gave a total of 12 hits from the non-redundant database with Z scores of greater than 15.0. The top hit was *S. coelicolor* ActR with bound actinorhodin (PDB code 3B6A) (Willems *et al*., 2008) with a Z score of 18.7, a sequence identity of 25% and a root-mean-square deviation (rmsd) of 2.6 Å after superposition of single subunits. Hit number 6 was *E. coli* TetR with bound tetracycline (PDB code 1BJY; Z score 17.0; sequence identity 18%; rmsd 2.6 Å) (Orth *et al*., 1999).

In a recent systematic survey of TFR protein structures, TetR itself, together with ActR, were placed in a distinct subclass of structures (Yu *et al*., 2010). One of the characteristics of this subclass is a long insertion between α8 and the C-terminal α-helix. This corresponds to the α9-α10 arm of SimR. By contrast, CgmR from *Corynebacterium glutamicum* (PDB code 2ZOY; ligand-free form) is more typical of the majority of known TFR structures, having no such insertion (Itou *et al*., 2010) (**Fig. 5.5B**). As a consequence, the dimer interface of CgmR is less extensive than that of SimR-apo at 1819 $Å^2$ (as calculated by PISA). In addition, several TFRs have a significantly longer α1 helix, the N-terminus of which may make additional contacts to DNA as seen in QacR, CgmR and DesT (Grkovic *et al*., 2001; Itou *et al*., 2010; Miller *et al*., 2010; Schumacher *et al*., 2002).

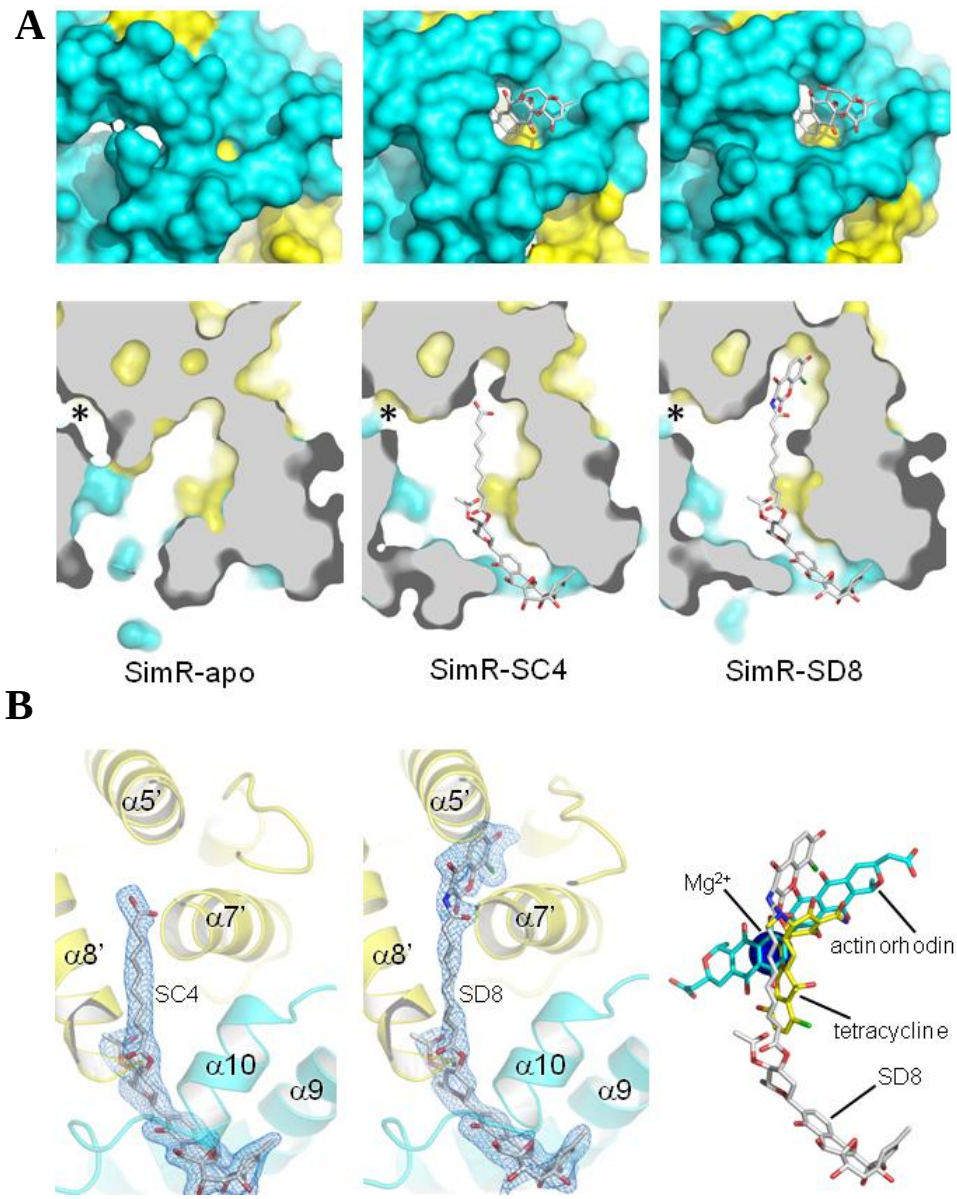


Fig. 5.4. The ligand-binding pocket of SimR. (A) Molecular surface representations of the ligand-free and ligand-bound structures of SimR focussing on the right-hand pocket shown in **Fig. 5.1**; ligands are shown with grey carbon atoms. The top panels show a front view and the lower panels show a top view as a thin slice through the molecule. Grey shading denotes regions within the molecular boundary. The asterisks indicate narrow channels leading to equivalent cavities in the other halves of the dimers. (B) Simulated annealing omit maps at 2.3 Å resolution for SimR-SC4 and SimR-SD8 (contoured at 3σ) are shown in blue superposed on their respective ligands; the protein is shown in semi-transparent cartoon representation. The right-hand panel illustrates the relative positions of the ligands for SimR, ActR and TetR after superposition of their respective structures.

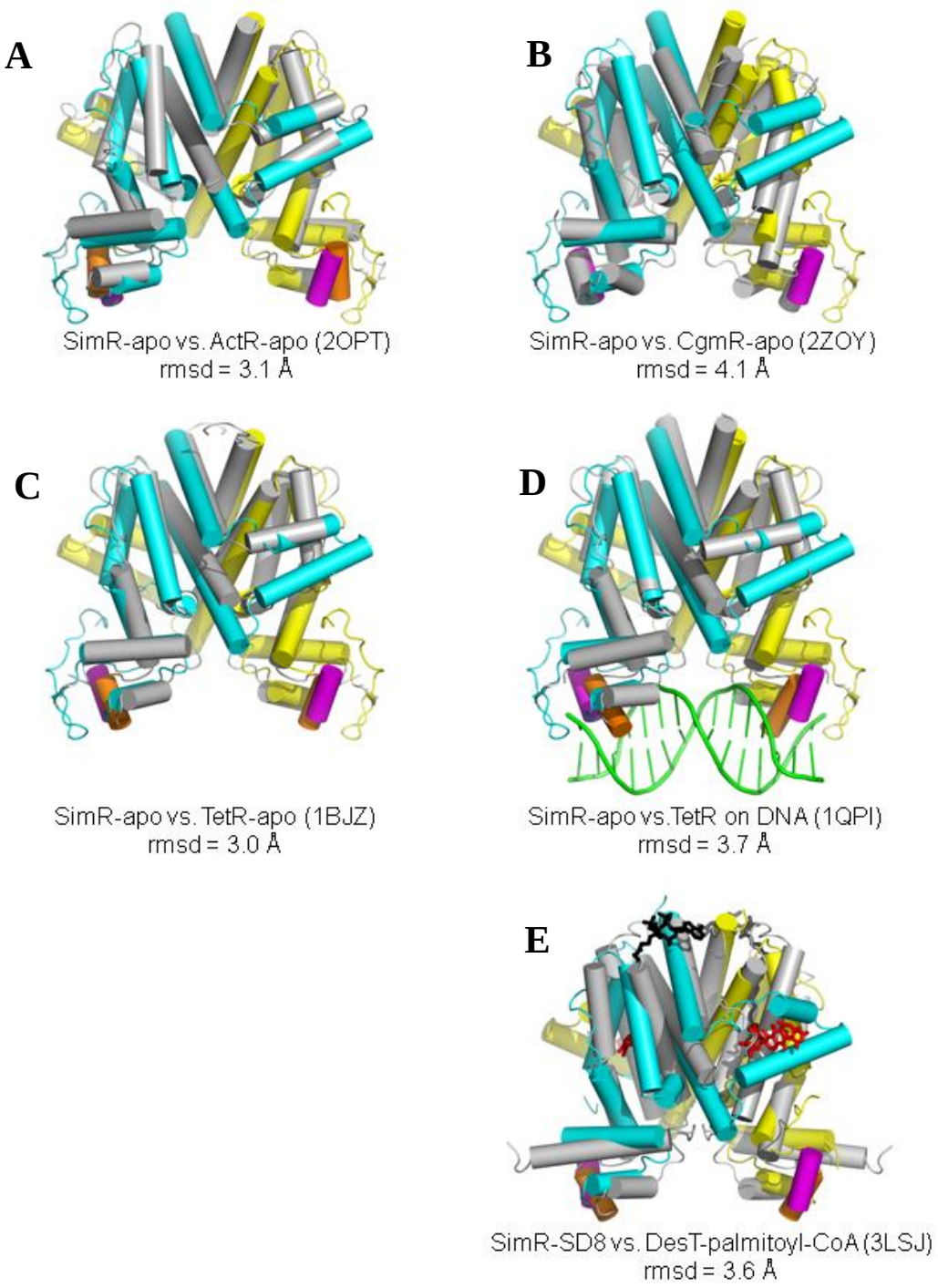


Fig. 5.5. Comparison of SimR structures to other TetR family transcriptional regulators. Throughout, cylindrical helix representations are used to simplify the comparisons, and the standard colour scheme is adopted for SimR; the second protein is shown in grey with the recognition helices in orange. For the comparison with DesT (E) the ligands are shown as sticks with SD8 in red and palmitoyl-CoA in black. Rmsd values are based on dimer-upon-dimer superpositions.

Tyr42 in TetR lies in the middle of the recognition helix (α3) and forms important interactions with DNA in the TetR-*tetO* complex (Orth *et al.*, 2000). This residue is reasonably well conserved in TFRs and the Cα atom serves as a reference point for measuring the separation of the recognition helices in a homodimer (Yu *et al*., 2010). In DNA-bound structures, this separation lies in the range 34.7 to 38.8 Å, roughly corresponding to the distance of 34 Å between consecutive major grooves in standard B-form DNA (Grkovic *et al*., 2001; Itou *et al*., 2010; Miller *et al*., 2010; Orth *et al*., 2000; Schumacher *et al*., 2002). By contrast, in apo-TFR structures the helix separation is highly variable, but is typically not compatible with DNA binding (Yu *et al*., 2010). The equivalent residue in SimR is Tyr65 and the Cα-Cα separation is 42.3 Å in SimR-apo (**Fig. 5.1**). Thus, we conclude that the conformation seen in SimR-apo is also not compatible with DNA binding (**Fig. 5.5D**).

5.4 Ligand-bound structures of SimR

SC4 is a biosynthetic intermediate of SD8 (**Fig. 4.1**) that is essentially inactive as a DNA gyrase inhibitor (Edwards *et al*., 2009). However, like the mature antibiotic, SC4 derepresses SimR; when applied exogenously, it induces expression of the SimX efflux pump *in vivo,* and it relieves DNA binding by SimR *in vitro* (Chapter 4 and Le *et al*., 2009). Structures of the complexes formed between SimR and the mature antibiotic SD8 (SimR-SD8), and with SC4 (SimR-SC4), were determined at 2.3 Å resolution. In both cases, crystals of the complexes were formed by co-crystallization with the respective ligands and gave rise to new crystal forms. Each structure was solved by molecular replacement using the SimR-apo structure as the template and, in both cases, there were two copies of the SimR subunit in the ASU, corresponding to the biological dimer. In contrast to the SimR-apo structure, most of the N-terminal polypeptide chain prior to α1 in the ligand-bound structures was poorly resolved; only in one monomer of SimR-SC4 was the β-hairpin visible. For both complexes, electron density for two copies of the corresponding ligand per SimR dimer was clearly present, occupying equivalent sites related by two-fold non-crystallographic symmetry (**Figs. 5.4B** and **5.6**).

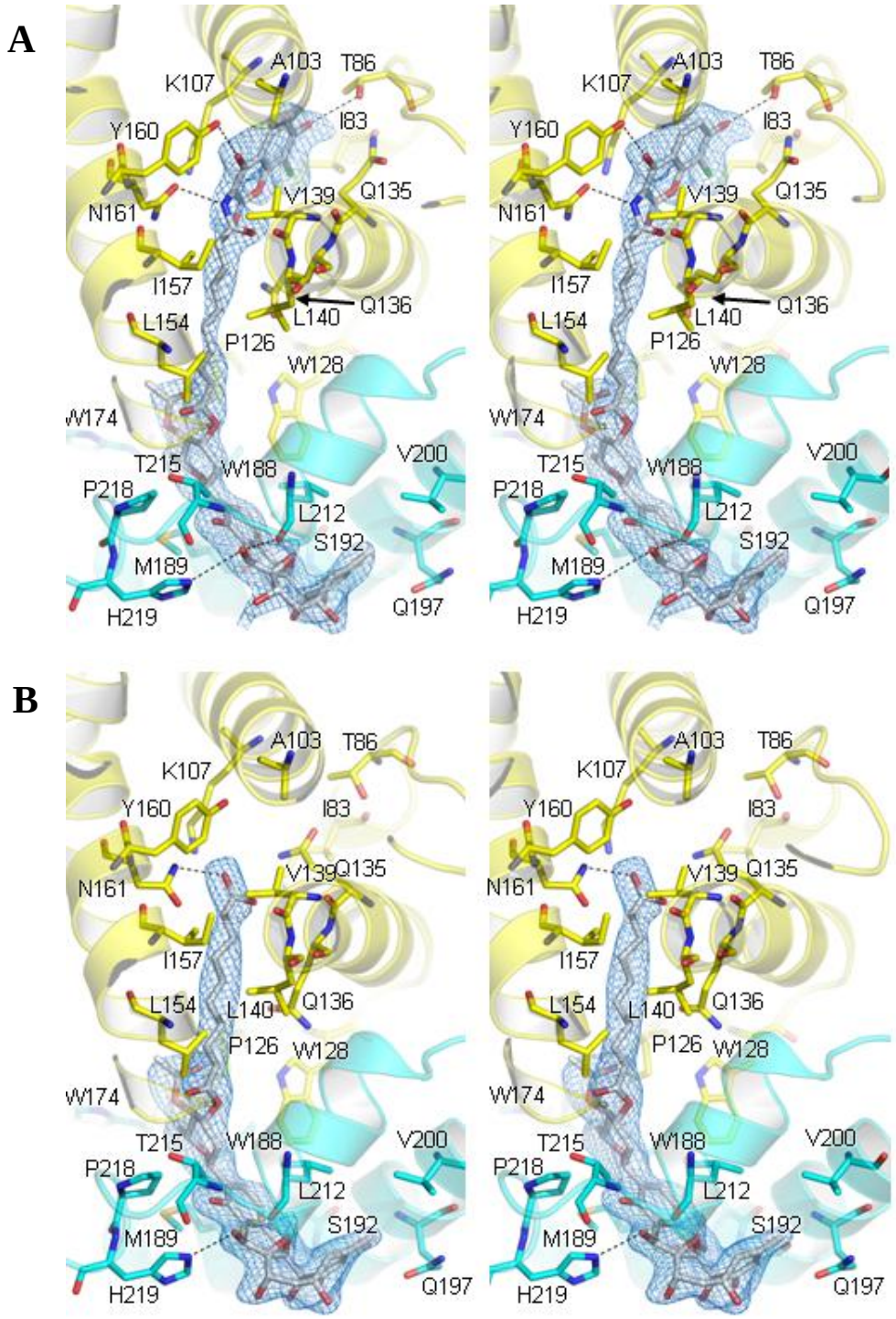


Fig. 5.6. The ligand-binding pocket of SimR. Stereoviews of the images shown in Fig. 4b with the addition of protein side chains, where (A) shows SimR-SD8 and (B) shows SimR-SC4. Simulated annealing omit maps at 2.3 Å resolution (contoured at 3σ) are shown in blue superposed on their respective ligands, which are shown with grey carbon atoms; the protein is shown in semi-transparent cartoon representation.

At the protein backbone level, the two ligand-bound structures of SimR were similar. In pairwise superpositions of individual subunits, the rmsd values were in the range 0.66 to 0.75 Å, and a dimer-upon-dimer superposition gave a value of 0.74 Å. In comparisons between the SimR-apo and the two ligand-bound structures at the subunit level, the most striking difference is the concerted upward shift of the α9-α10 arm upon ligand binding (**Fig. 5.7A**). Superpositions of individual ligand-bound subunits upon the SimR-apo subunit, based only on residues in the subunit core (i.e. excluding α9-α10, specifically residues 182-221), gave remarkably low rmsd values in the range 1.01 to 1.09 Å. Thus, the cores are extremely similar and, most importantly, the relative orientations of LBDs and DBDs within each subunit remain essentially unchanged upon ligand binding. At the dimer level, differences between the SimR-apo and the ligand-bound structures of SimR become more pronounced due to a twisting motion at the dimer interface. As a result, dimer-upon-dimer superpositions (using core residues) gave larger values of 2.13 and 2.04 Å, for SimR-SD8 vs. SimR-apo and SimR-SC4 vs. SimR-apo, respectively. This twisting motion is more noticeable in comparisons where only one subunit in each dimer is used for the superposition (**Fig. 5.8**). When viewed from the side of the dimer, one subunit rotates relative to the other by around 8°. The axis of rotation passes through equivalent points in each subunit roughly at the interface between the LBD and DBD just above the midpoint of α3. Because the pivot point is close to the DBD, this small rotation has very little effect on the separation of the recognition helices (**Fig. 5.8**); values of 41.7 and 42.6 Å were obtained for SimR-SD8 and SimR-SC4, respectively, being comparable to the value of 42.3 Å obtained for the SimR-apo structure (**Fig. 5.1**). The upward motion of the α9-α10 arm upon ligand binding is, in part, correlated with the movement of the adjacent subunit (**Figs. 5.7** and **5.8**). This is because each arm is closely associated with the opposing subunit and many of the interactions are retained, including several hydrogen bonds at the extremities of the arm (C-terminal end of α9 and N-terminal end of α10).

As a result of the subunit rotation, the C-termini move closer together. For example, the separation of the terminal residues of α11 (i.e. Arg243), decreases from 17.6 to 10.1 Å upon SD8 binding (**Fig. 5.8**); being remote from the pivot point, conformational differences are more exaggerated here. In addition, there is more subunit overlap, such that there is a modest increase in the dimer interface area from

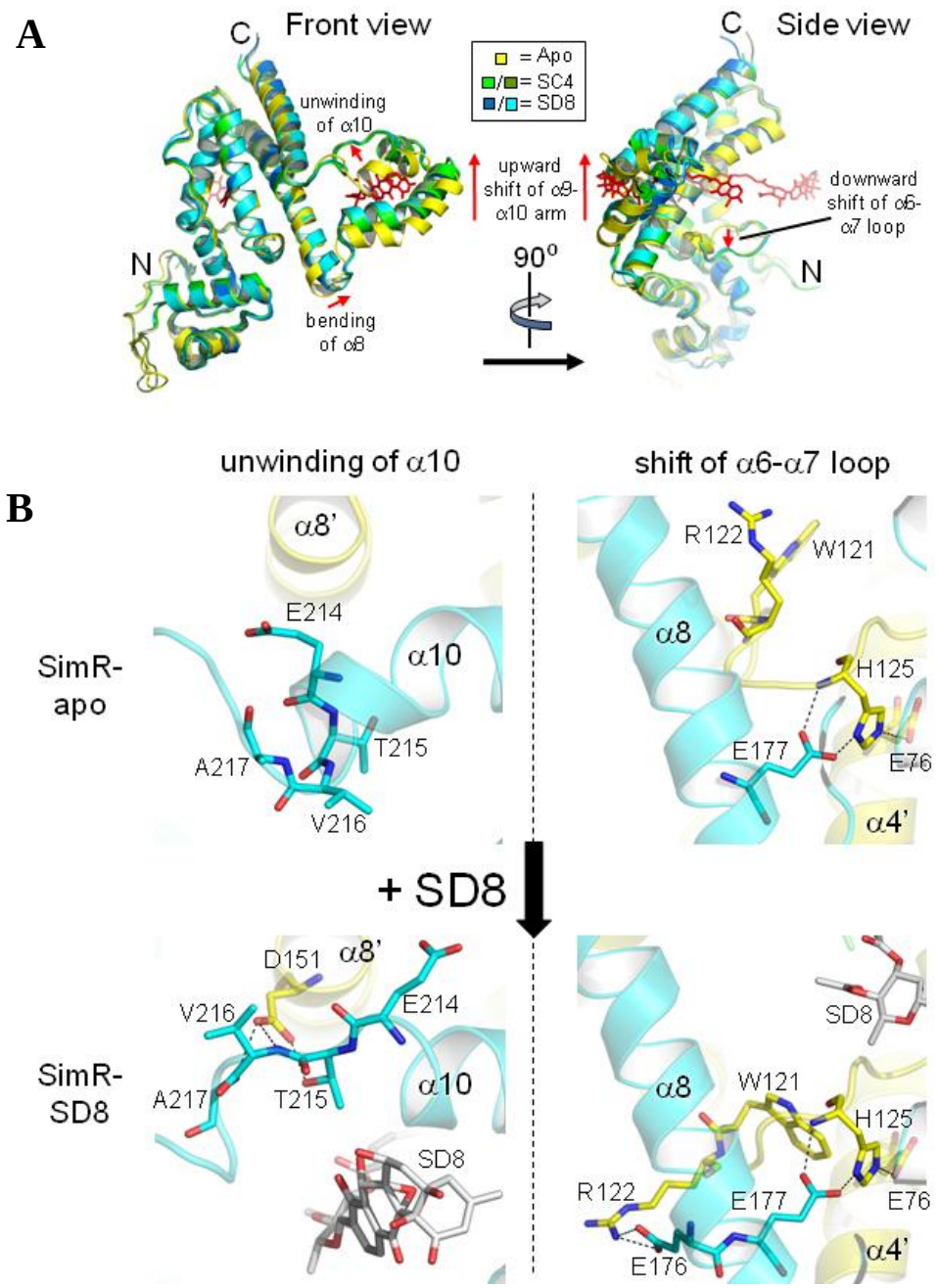


Fig. 5.7 Conformational differences between SimR-apo and SimR-SD8. (A) Superposition of all crystallographically independent subunits from all three SimR structures (five in total) based only on the core residues (i.e. not including the α9-α10 arms). Only the left-hand subunits from Fig. 2 are shown. (B) Close-up views illustrating the conformational changes in α10 and in the α6-α7 loop. Also shown in the right-hand panels is the Glu177-His125-Glu76 hydrogen-bonding network that is present in both structures. The SD8 molecules are shown in stick representation in (a) red or (b) grey carbons (see also **Fig. 5.10**).

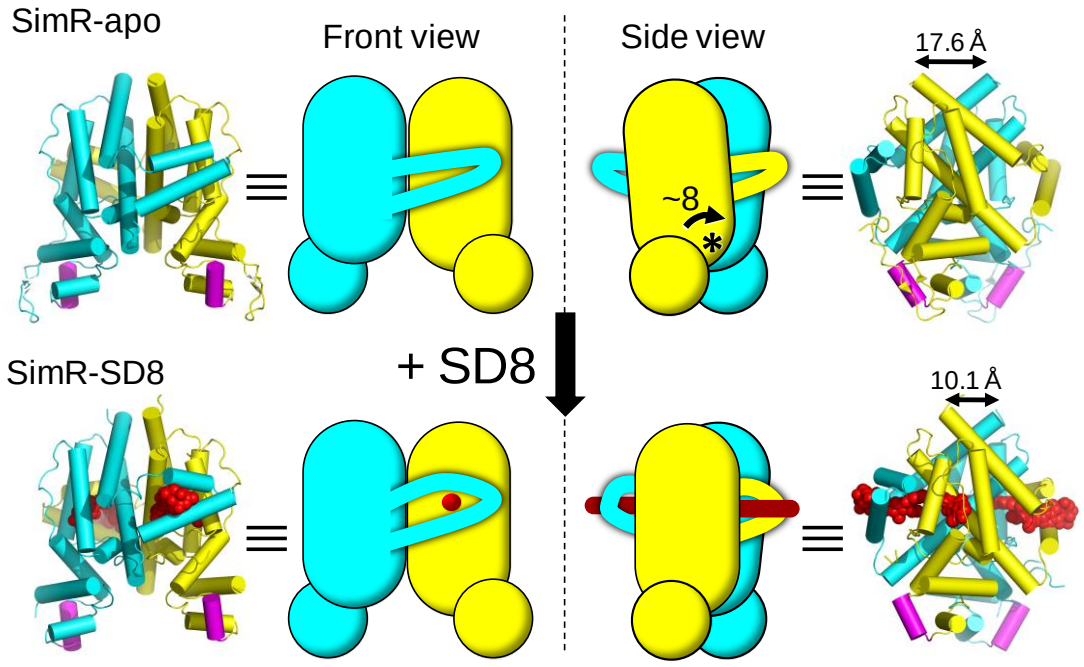


Fig. 5.8. Rotation of SimR subunits stabilised by ligand binding. Structures of SimR-apo and SimR-SD8 together with schematic representations illustrating the rotation of the subunits relative to one another. For this figure only, the 'front view' shows the dimer with the plane of the interface perpendicular to the paper (it differs from the 'front view' used elsewhere by a ~25° rotation around the vertical axis). In order to emphasize the subunit rotation, the cyan-coloured subunits are shown fixed in the same relative orientations. This can be clearly seen in the side view where the yellow subunit rotates by ~8° relative to the cyan subunit; the approximate pivot point is indicated by the asterisk. The distances separating the terminal residues of α11 (i.e. Arg243) in the two structures are marked.

2475 $Å^2$ to 2759 $Å^2$ upon SD8 binding. Non-bonded protein-protein interactions across the interfaces are extensive in both conformational states, although there is a slight increase in the number of pairwise interatomic distances under 4 Å, from 297 in SimR-apo to 336 in SimR-SD8. More significant is the increase in the number of protein-protein hydrogen bonds from 18 in SimR-apo to 30 in SimR-SD8 (values from PDBsum; http://www.ebi.ac.uk/pdbsum/) (Laskowski, 2001). Further analysis indicates that 12 of these are retained between the two structures. These comprise contacts between the α9-α10 arm of one subunit and the α6-α7 loop and α7 of the other, and between the α6-α7 loop of one subunit and α8 of the other.

5.5 The simocyclinone binding site in SimR

The SD8 and SC4 ligands bind to the same sites in their respective complexes with SimR, such that moieties common to both molecules occupy equivalent positions in the binding pocket (**Figs. 5.4** and **5.6**). Thus, only a detailed description of the protein-ligand interactions for SimR-SD8 is necessary. The SD8 molecule is bound in an extended conformation spanning almost 30 Å, making substantial contacts with both subunits of the homodimer (**Figs. 5.6** and **5.9**). The AC end is completely buried in the core of the SimR dimer, in a pocket formed by residues from α4, α5, α7 and α8 and several of the intervening loops from one SimR monomer. The α6-α7 loop adopts a different conformation in the ligand-bound structures, thereby avoiding a steric clash between Trp121 and the AC moiety. Otherwise, the AC pocket is essentially preformed at the protein backbone level in SimR-apo, only requiring a few side chains to reposition themselves, most notably, Gln135, Gln136 and Asn161. By contrast, the PK end is partially exposed to bulk solvent in a peripheral site that is formed almost exclusively by helices α9 and α10 (i.e. the extended arm) of the other monomer. The lower part of the PK binding site is present in SimR-apo and corresponds to the 'saddle' resulting from the kink in α9 (**Fig. 5.1**). The upper part of the site is completed in the ligand-bound form as the result of the conformational changes in the α9-α10 arm described below, principally the unwinding of α10. In between the AC and PK moieties, the cavity widens considerably (**Fig. 5.4**). This is where the linker and the olivose sugar are accommodated.

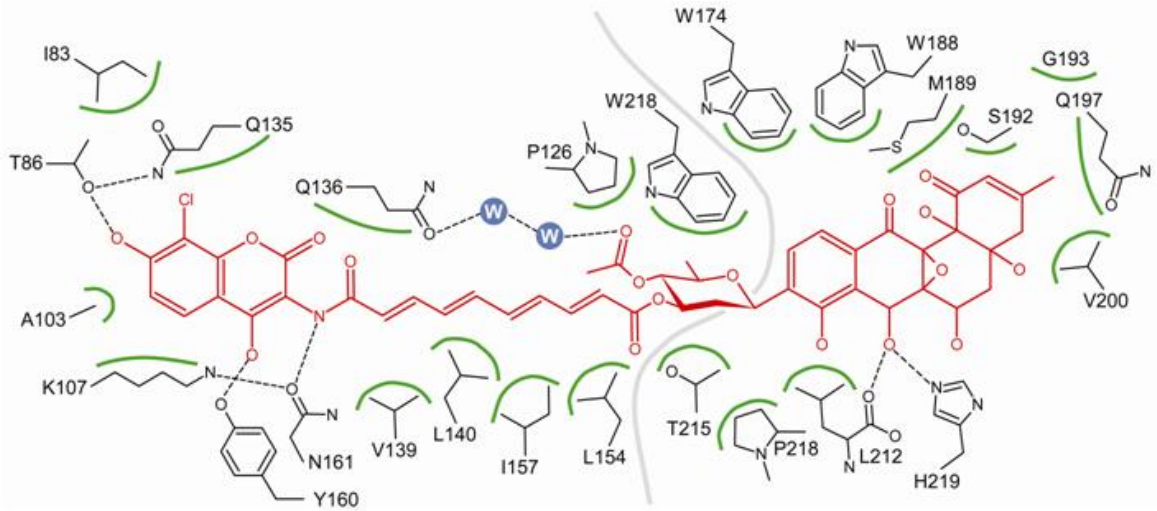


Fig. 5.9. Schematic representation of the SD8-binding pocket of SimR. The ligand is shown in red and all residues within 4 Å of it are shown in black. Dotted lines indicate hydrogen bonds and green arcs indicate non-bonded interactions. The two water molecules (labelled 'W') that link Gln136 to the olivose moiety of SD8 are shaded blue. The grey line delineates the boundary between the two subunits. For clarity all hydrogens have been omitted.

What is particularly striking about the binding of SD8 is that despite the ligand having 19 atoms that could either donate or accept a hydrogen bond, it makes a maximum of only five direct hydrogen bonds to SimR (**Figs. 5.6** and **5.9**). However, this is compensated by the extensive van der Waals contacts with the protein.

When comparing the SimR-SD8 and SimR-SC4 structures, we observed several minor differences due to the absence of the AC moiety in SC4. The side chains of Gln135 and Gln136, which are displaced in SimR-SD8 due to the presence of the AC moiety, in SimR-SC4 adopt conformations similar to those found in SimR-apo (**Fig. 5.6).** Similarly, the α4-α5 surface loop differs slightly in SimR-SC4 due to the lack of an interaction between Thr86 and the ligand. Since these small differences appear to influence neither the dimer interface nor the DBD, they are unlikely to explain why SC4 is less effective than SD8 at derepressing SimR *in vitro* (Chapter 4 and Le *et al*., 2009). Rather, we speculate that this is simply a consequence of the fewer favourable interactions that SC4 makes with the protein due to the lack of an AC moiety, which is likely to result in a reduced binding affinity of SC4 for SimR.

SD8 is a potent inhibitor of DNA gyrase, and functions by a novel mechanism, binding to the GyrA subunit and preventing DNA binding (Flatman *et al*., 2005). Recently, the crystal structure of the GyrA-SD8 complex was determined, revealing the structural basis of enzyme inhibition (Edwards *et al*., 2009). SD8 binds to GyrA in an extended conformation with the PK and AC moieties occupying two separate binding pockets. As might be expected, there is no correspondence between the ligand-binding pockets in GyrA and SimR.

In free solution, the tetraene linker of SD8 would be predicted to be linear, due to the conjugated double bond system, as is observed in GyrA-SD8 (Edwards *et al*., 2009) However, in SimR-SD8, the tetraene linker of SD8 is curved. This presumably occurs because the relative orientations of the AC and PK pockets are incompatible with the linker remaining linear. Even in SimR-SC4, the tetraene linker exhibits some curvature.

5.6 Comparison to other TetR family ligand-binding sites

The LBDs of TFRs are significantly more variable than their DBDs, reflecting the plethora of ligands that can be bound (Yu *et al*., 2010). Moreover, the manner in which ligands are bound is highly variable. Even in comparisons between SimR-SD8 and the ligand-bound forms of the structurally closely related ActR and TetR, it is clear that there is little correspondence when the structures are superposed (**Fig. 5.4B**). The long axis of actinorhodin crosses that of SD8 at an angle of roughly 70°, while the majority of tetracycline aligns loosely with the tetraene linker, rather than with one of the ring systems of SD8. Where structures of TFRs in complex with their cognate ligands have been determined, the cognate ligands are bound predominantly by one subunit of the biological dimer: in ActR there is only one contact with the second subunit, while in TetR there is none (**Fig. 5.2**). By contrast, simocyclinone makes substantial contacts with both subunits of the dimer (**Figs. 5.1, 5.2, 5.4, 5.6** and **5.9**), with the majority of the contacts to the second subunit involving the α9-α10 arm. The extent of the SD8 binding pocket is also unusual; the only other example showing a pocket of comparable length is the palmitoyl-CoA complex of DesT (PDB code 3LSJ) (Miller *et al*., 2010). However, in this case, the pocket entrance is at the top of the subunit and the ligand lies roughly perpendicular to SD8, interacting mainly with a single subunit (**Fig. 5.5E**).

5.7 Conformational changes in SimR captured upon ligand binding

Available evidence suggests that apo-TFRs sample a range of conformations and that ligand-binding captures one of these conformations, rather than inducing the conformational change (Reichheld *et al*., 2009; Yu *et al*., 2010). As a consequence, we have assumed here that the same is true of SimR.

In a superposition of the subunits from the various SimR structures, the upward shift of the α9-α10 arm in the ligand-bound form is the most striking difference (**Fig. 5.7A**). Presumably the α9-α10 arm is carried along with the neighbouring subunit when the two subunits rotate relative to one another as part of the normal dynamic behaviour of the system, and this conformational state is captured on ligand binding

(**Fig. 5.8**). However, the arm does not move entirely as a rigid body: although α9 is essentially unchanged, the C-terminal end of α10 partially unwinds (it becomes two residues shorter), and the now extended α10-α11 loop moves even further upward, causing the remainder of α10 to tilt upward. The loop is stabilized in its new conformation by additional inter-subunit hydrogen bonds involving Asp151 in α8 (**Figs. 5.7B** and **5.10**). The net effect of these motions is that the channel connecting the central cavity to bulk solvent in SimR-apo closes, and a new one opens adjacent to it, which is wide enough to accommodate the PK moieties of SD8 and SC4 (**Fig. 5.4**). A more subtle effect of the upward arm motion is that the C-terminal end of α8 bends towards the opposing subunit, leading to a number of minor conformational changes in the α8 - α9 loop.

In all three SimR structures, there are two symmetry-related hydrogen-bonding networks that lie on either side of the pivot and link Glu177 in α8 of one subunit to His125 in the α6-α7 loop of the other subunit, which in turn is hydrogen-bonded to Glu76 in α4 of the same subunit (**Figs. 5.7** and **5.10**). These networks may be important for tethering the subunits to one another and restraining inter-subunit rotations. The portion of the α6-α7 loop C-terminal to His125 remains unchanged, whereas the N-terminal portion moves downward in the ligand-bound structures. At the protein backbone level, the motion seems relatively minor. However, the side chains of two bulky residues, namely Trp121 and Arg122, flip downwards, with the tip of the latter moving by some 17 Å into the dimer interface, where it is salt-bridged to Glu176 in α8 of the opposing subunit (**Figs. 5.7** and **5.10**), thereby reciprocating the link between α8 and the α6-α7 loop, involving Glu177 and His125, described above.

In SimR-apo, the two monomer cores (i.e. excluding the α9-α10 arms) present relatively flat surfaces to one another, and thus could potentially slide past each other without steric restraint. However, in the ligand-bound form, the two copies of Arg122 project across the dimer interface into pockets in the surfaces of the opposing monomers (**Fig. 5.11**). Thus, they may act as retaining pins, effectively locking the subunits together.

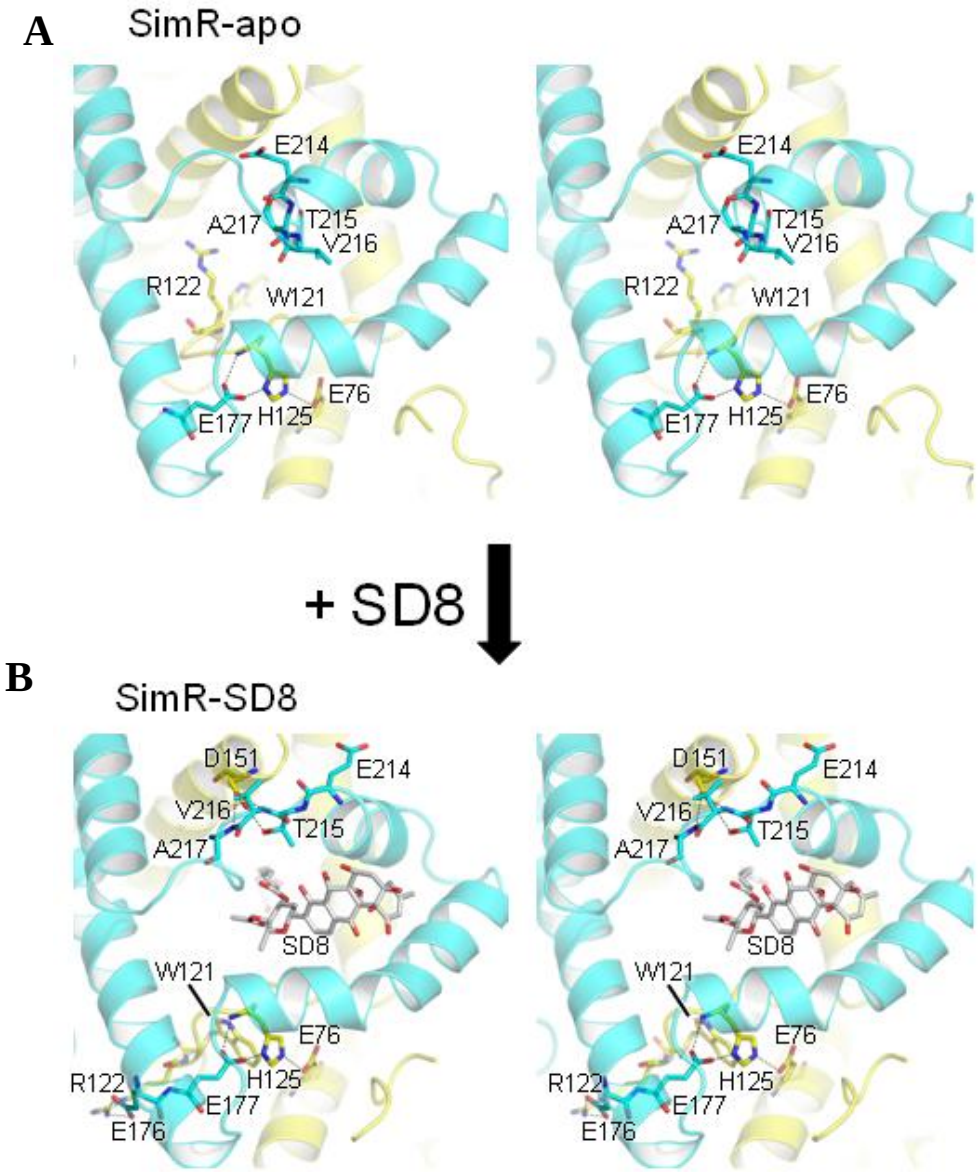


Fig. 5.10. Detail of differences between SimR-apo and SimR-SD8 structures. Stereoviews comparing the conformations of α10 and the α6-α7 loop in (A) SimR-apo and (B) SimR-SD8. Also shown is the Glu177-His125-Glu76 hydrogen bonding network that is present in both structures. The SD8 molecules are shown in stick representation with grey carbons for (B).

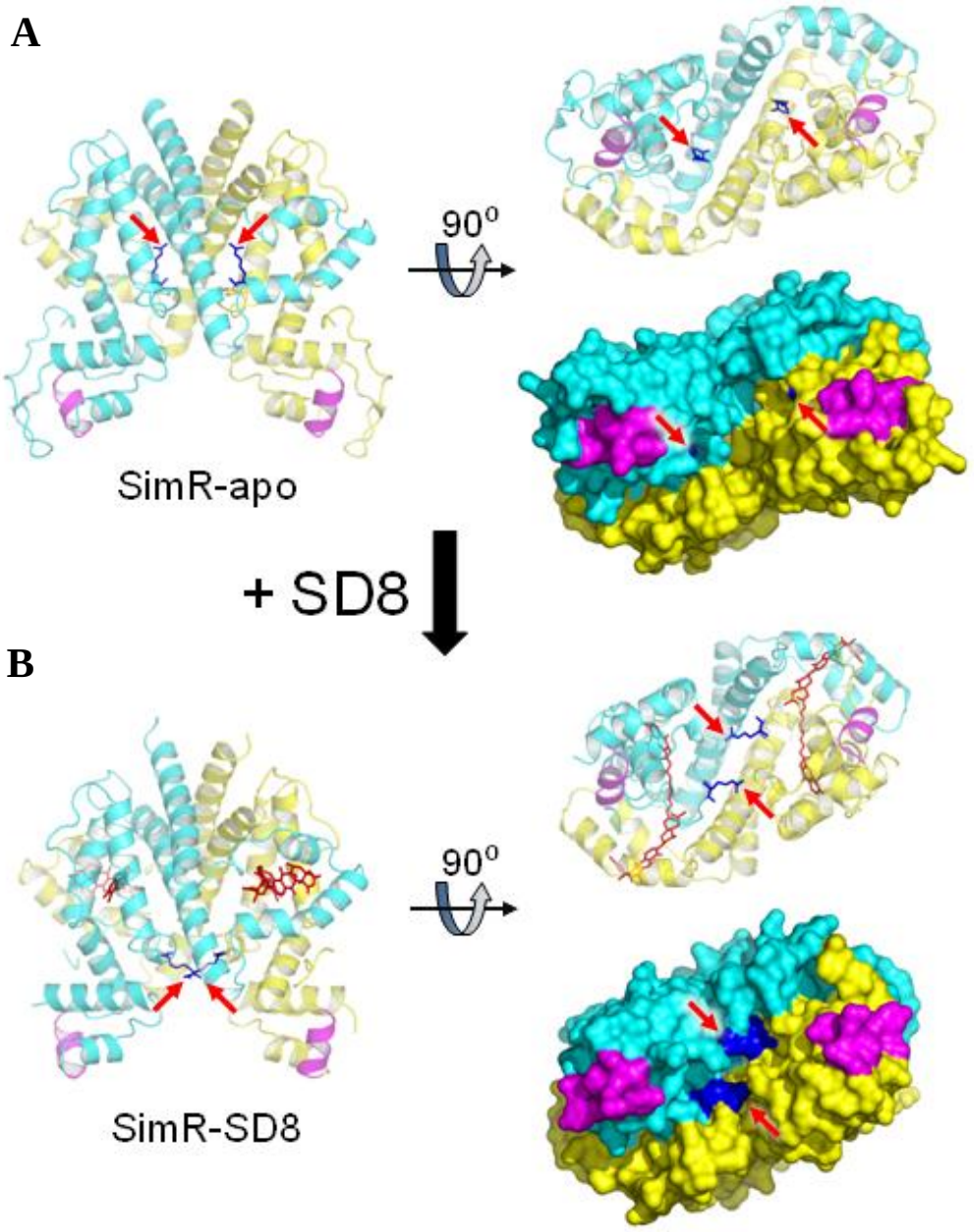


Fig. 5.11 Putative derepression mechanism of SimR. Semi-transparent cartoon and molecular surface representations of (A) SimR-apo and (B) SimR-SD8. The left-hand panels show the front view and the right-hand panels show the bottom view. The symmetry equivalent Arg122 residues are coloured blue (and shown in stick representation for the cartoons) and the ends of the side chains are indicated by the black arrows; SD8 molecules are shown as red sticks. In SimR-SD8 the inter-digitation of the Arg122 side chains is most apparent in the right-hand panels.

5.8 The mechanism of derepression in SimR

It is widely accepted that the derepression of TFRs involves allosteric mechanisms, since their ligand-binding sites are remote from their DBDs (Ramos *et al*., 2005; Yu *et al*., 2010). It may be significant that the majority of characterised TFR ligand-binding sites are contained almost entirely within individual subunits, and that for those TFRs for which ligand-bound and DNA-bound structures are available, the allosteric mechanism involves conformational changes transmitted largely within the same subunit (Itou *et al*., 2010; Miller *et al*., 2010; Orth *et al*., 1999; Orth *et al*., 2000; Schumacher *et al*., 2001). The binding of ligands to these sites captures a conformational state in which the DBD (the recognition helix in particular) is repositioned relative to the LBD such that the dimer is prevented from binding to cognate DNA. Nevertheless, inter-subunit communication appears to play a role in some systems, e.g. QacR (Schumacher *et al*., 2001) and CgmR (Itou *et al*., 2010), as the binding of ligands by one subunit can render the other subunit incapable of doing the same. In SimR, the binding of SD8 stabilises a number of conformational changes, the most significant of which occur at the dimer interface. These changes have the effect of repositioning the two subunits of the dimer relative to one another, and locking them in this new configuration. It seems likely that the side chain of Arg122 in the α6-α7 loop is important for this locking mechanism (**Fig. 5.11**). Moreover, it also seems likely that further rigidity is imparted on the system by the ligand, which threads through both subunits. Indeed, the induced curvature of the tetraene linker of bound SD8 may be indicative of the forces required to maintain the protein in the derepressed state.

In the derepression mechanism of TetR, the principal changes upon ligand binding are an upward movement of the α6-α7 loop (i.e. away from the DBD), a partial unwinding of α6, and a 'pendulum-like' swing of α4 around its C-terminal end that repositions the DBD relative to the LBD such that the two recognition helices move further apart to a configuration that would abrogate binding to DNA (Hinrichs *et al*., 1994; Kisker *et al*., 1995; Orth *et al*., 1998; Orth *et al*., 1999; Orth *et al*., 2000). By contrast, in SimR, although there are changes in the α6-α7 loop, it moves down instead of up, and there is no unwinding of α6. Furthermore, α4 does not swing at all, and thus the relative dispositions of the LBDs and DBDs within individual subunits

of SimR remain essentially unchanged. Therefore in both SimR-apo and SimR-SD8, the recognition helices are too far apart to allow SimR to bind DNA (**Fig. 5.1**). This is also true for the apo and ligand-bound states of TetR and ActR (Orth *et al*., 1998; Willems *et al*., 2008).

How does ligand binding prevent SimR from binding to DNA? It has been assumed that the ligand-free form of SimR is sufficiently dynamic in solution that it can readily sample different conformational states. Davidson and colleagues performed equilibrium unfolding experiments on wild-type TetR and proposed that the thermodynamic coupling of TetR domains, in particular, the rigidification of the DBDs upon ligand binding, underlies the allosteric response in TetR (Reichheld *et al*., 2009). For SimR, we speculate that a DNA-binding conformation could be achieved largely through rigid-body motions of the subunits about a pivot point towards the upper end of the LBD, without invoking significant repositioning of the DBDs relative to the LBDs. Such a conformational change would be disfavoured in the ligand-bound structures due to a combination of factors, including the rigidity imposed on the system by the ligand threading through both subunits and the increased number of favourable inter-subunit contacts, in particular, the inter-digitation of the Arg122 side chains (**Fig. 5.11)**, together preventing DNA binding by SimR-SD8.

5.9 Conclusion and Summary

SimR is a TetR-family transcriptional regulator that provides a mechanism that couples the biosynthesis of the antibiotic simocyclinone to its export in the producing organism *S. antibioticus*. I have determined the crystal structures of the protein in the absence of ligands and in complexes with the mature antibiotic, simocyclinone D8, and with the biosynthetic intermediate, SC4. SimR displays the α-helical fold typical of TFRs and exists as a homodimer. In addition to the usual nine helices, the SimR monomer has an insertion of two extra helices, prior to the C-terminal helix, which together form an arm that wraps around the opposing subunit. This feature is only found in a subfamily of TFRs that includes ActR and TetR, which are close structural homologues of SimR (Yu *et al*., 2010). Furthermore, SimR has an N-terminal, Arg-

rich β-hairpin motif that is not observed in other TFR structures, which appears to be ideally placed to engage with phosphate groups in the minor groove of DNA. This possibility is addressed in Chapter 6.

The ligand-bound structures of SimR reveal an extended binding pocket that, unusually for TFRs, spans both subunits in the functional dimer. Interactions with one subunit almost exclusively involve the α-helical arm, which is reconfigured in the ligand-bound form. Through comparisons of these structures, we are able to postulate a derepression mechanism for SimR that involves rigid-body motions of the subunits relative to one another, coupled with a locking mechanism that most likely prevents further conformational change, holding the DBDs too far apart for DNA binding. SimR vividly illustrates the versatility of TFRs in their ability to recognise and respond to diverse ligands.

Chapter 6

The crystal structure of the TetR family transcriptional repressor SimR bound to DNA and the role of a flexible N-terminal extension in minor groove binding

6.1 Introduction

To date, the structures of only four TFRs bound to cognate DNA have been determined (TetR, DesT, CgmR and QacR), and it is clear that the mode of operator recognition differs from one member of the TFR family to another (Itou *et al.*, 2010; Miller *et al.*, 2010; Orth *et al.*, 2000; Schumacher *et al.*, 2002). For example, the tetracycline efflux pump repressor, TetR, binds as a dimer to a 15-bp operator and deforms the binding site by 17^{o}, bending away from the protein in order to optimise the position of its HTH for specific base pair interaction (Orth *et al.*, 2000). By contrast, the multidrug efflux pump repressor from *Staphylococcus aureus*, QacR, binds its cognate DNA site as a dimer of dimers and bends its operator by just 3^{o}, but widens the major groove to create an optimal DNA environment for a second QacR dimer to bind cooperatively nearby (Schumacher *et al.*, 2002).

I have determined the structures of apo (unliganded) SimR and SimR in complex with either SD8 or its biosynthetic intermediate SC4 (Le *et al.*, 2011b and Chapter 5). These structures revealed the same overall domain architecture for SimR as for other TFRs, including a classical HTH motif. However, SimR possesses an additional arginine-rich N-terminal extension that precedes the core DBD, which is significantly longer than those present in the four TFRs for which protein-DNA crystal structures are available (TetR, DesT, CgmR and QacR) (**Fig. 6.1**). With the exception of three residues, this 28 amino acid residue extension is disordered in both subunits in the SimR-SD8 structure, and it is only partially ordered in one subunit in the SimR-SC4 structure (Le *et al.*, 2011b and Chapter 5). Consistent with this, the N-terminal extension of SimR is predicted to be disordered in solution.

Here, I show by deletion analysis that the flexible N-terminal extension of SimR plays an important role in DNA binding, and I present the crystal structure of SimR bound to its operator sequence, which shows that this extension binds in the minor groove adjacent to the major groove occupied by the classical HTH motif. Although the N-terminal extension is hypersensitive to proteolysis *in vitro*, it becomes protease-resistant upon binding cognate DNA. Together these data suggest that the N-terminal

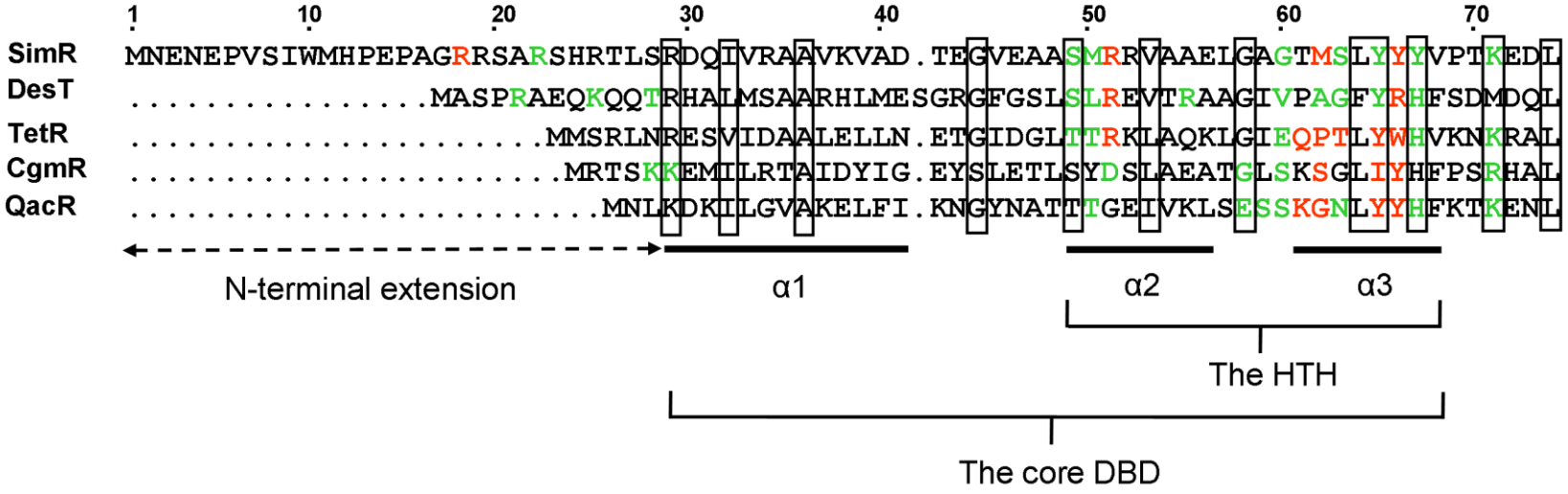


Fig. 6.1. Alignment of the amino acid sequence of SimR with the four other TFRs for which protein-DNA crystal structures are available (DesT, TetR, CgmR and QacR), showing the HTH motif, the core DBD and the N-terminal extension, herein termed the TFR arm present in SimR. For each TFR, amino acid residues that interact with the bases of the cognate DNA operator are highlighted in red, and those that contact the phosphate backbone are highlighted in green. Conserved residues are boxed.

extension transitions from a disordered to an ordered state upon DNA binding. Bioinformatic analysis of the entire TetR family shows that an N-terminal extension rich in positively charged residues is a feature of the majority of TFRs. Finally, comparison of the SimR-DNA and SimR-SD8 complexes reveals the conformational changes required to interchange between DNA- and ligand-bound states, which largely involve rigid-body motions of the subunits relative to one another.

6.2 N-terminally truncated SimR derivatives bind DNA with reduced affinity

SimR possesses a 28-residue N-terminal extension that precedes the core DBD, herein termed the TFR arm (**Fig. 6.1**), which carries four arginine residues at positions 18, 19, 22, and 25. This TFR arm is significantly longer than those in TetR, DesT, CgmR and QacR (**Fig. 6.1**), the four TFRs for which DNA-protein crystal structures are available (Itou *et al.*, 2010; Miller *et al.*, 2010; Orth *et al.*, 2000; Schumacher *et al.*, 2002). To determine if the TFR arm of SimR is involved in DNA binding, we made C-terminally His-tagged SimR derivatives with progressively shorter N-terminal extensions and tested them for binding to the *simR-simX* intergenic region by EMSA. Wild-type SimR and SimR derivatives with 10, 15, 22 or 25 amino acid residues deleted from the N-terminus were overexpressed and purified (**Fig. 6.2**). Increasing concentrations of protein were incubated with a DNA probe spanning the *simR-simX* intergenic region and the complexes were resolved on native polyacrylamide gels (**Fig. 6.3**). The *simR-simX* intergenic region contains two SimR operators: O_R closer to *simR*, and a higher affinity binding site, O_X, closer to *simX* (Le *et al.*, 2009 and Chapter 4). The lower and upper sets of shifted protein-DNA complexes seen in **Fig. 6.3** correspond, respectively, to single and double occupancy of these two SimR-binding sites (Le *et al.*, 2009). SimR DNA binding affinity was reduced ~30-fold when 10 or 15 amino acid residues were deleted from the N-terminus, and was reduced by at least 120-fold when 22 or 25 amino acid residues were removed (**Fig. 6.3**). These results suggested the TFR arm plays a role in DNA binding.

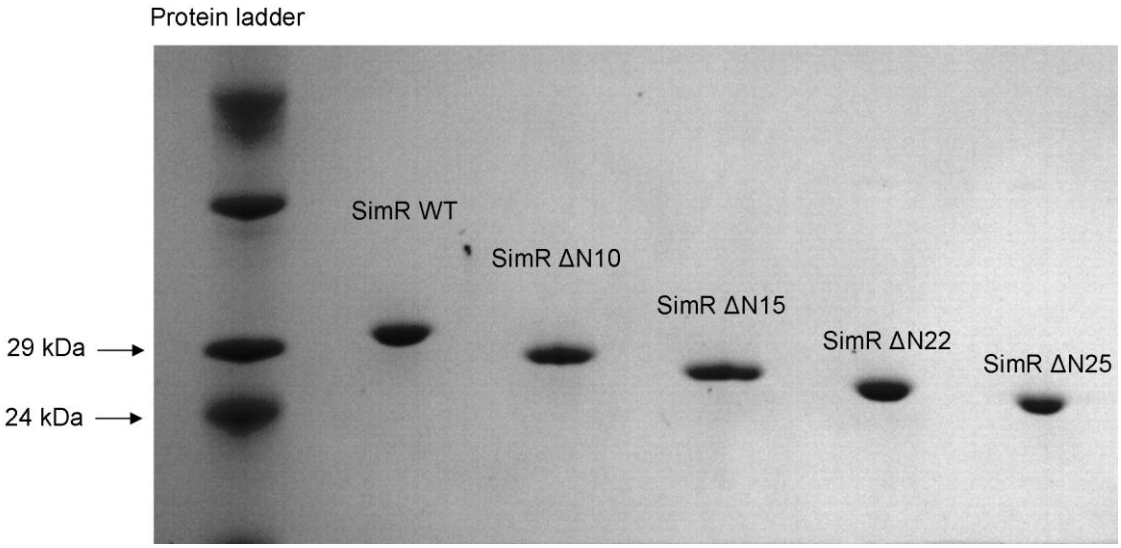


Fig. 6.2. Purified wild-type SimR and N-terminally truncated SimR derivatives. All SimR variants are C-terminally His tagged.

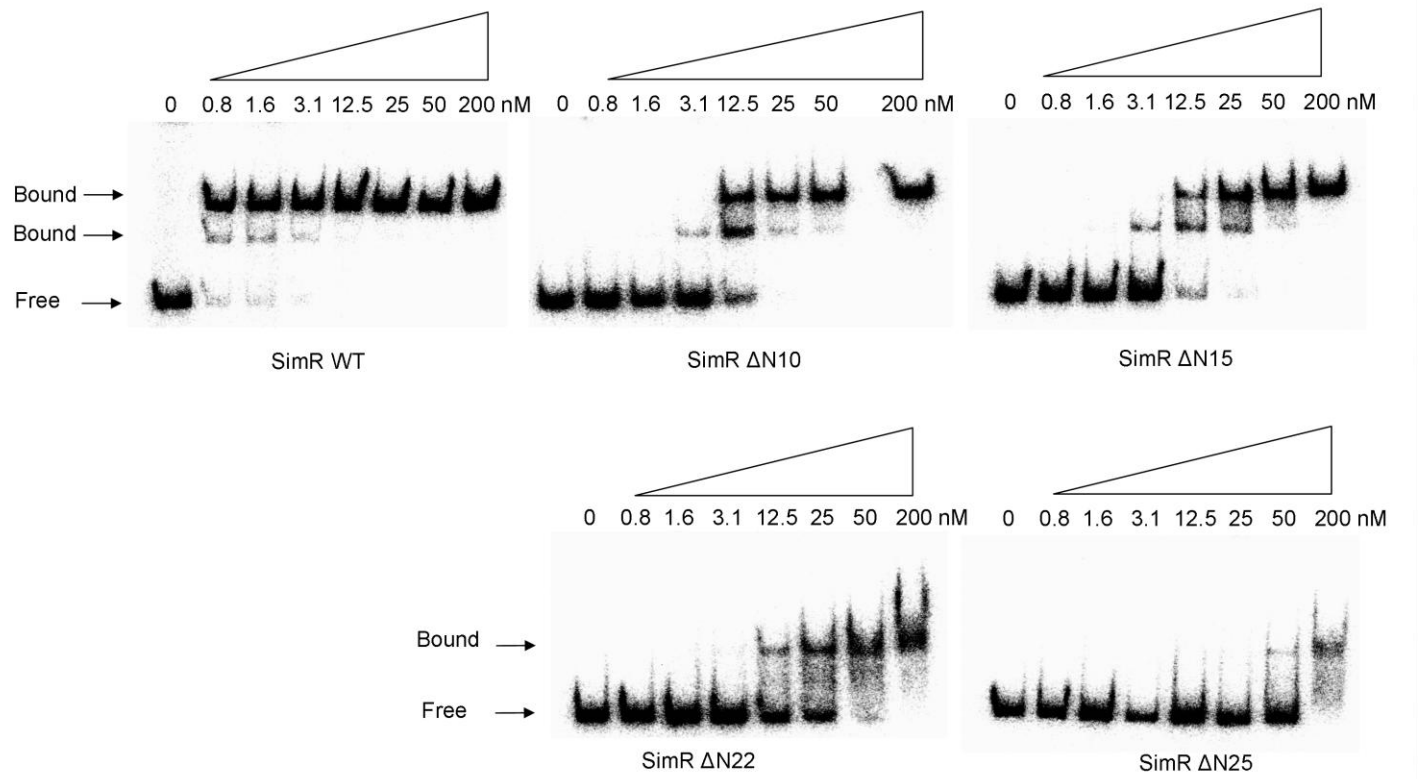


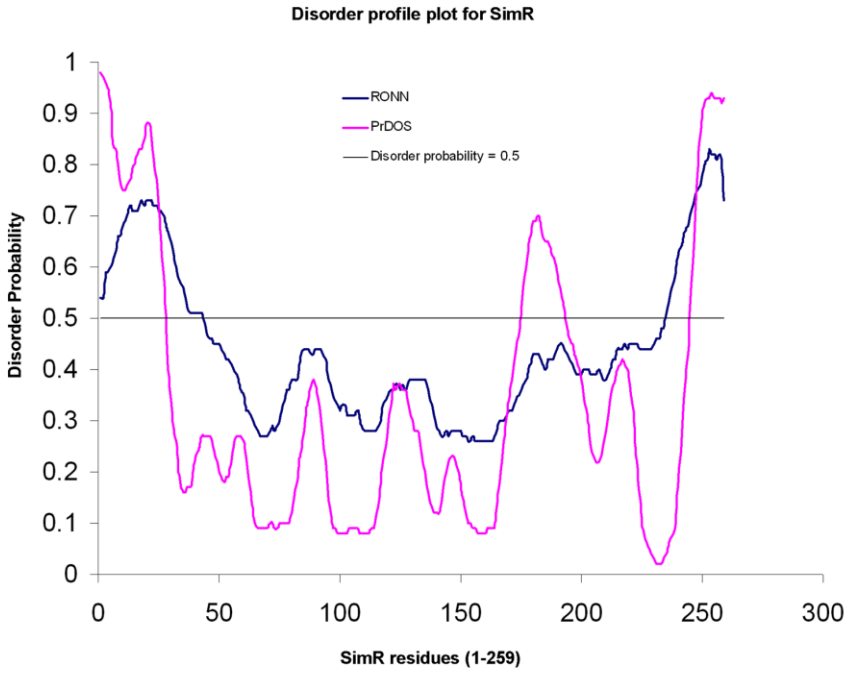
Fig. 6.3. Electrophoretic mobility shift assay (EMSA) showing the binding of purified wild-type and N-terminally truncated derivatives of SimR to the *simR-simX* intergenic region. Bands correspond to SimR-DNA complexes (Bound) and free DNA (Free) are indicated. The final concentration of SimR is indicated above each lane.

6.3 The TFR arm becomes protease-resistant upon DNA binding

The 28-amino acid TFR arm of SimR has a high proportion of disorder-promoting amino acids and is predicted by the Proteins Disorder Prediction System (PrDOS; http://prdos.hgc.jp/cgi-bin/top.cgi) and by the Regional Order Neural Network (RONN; http://www.strubi.ox.ac.uk/RONN) servers to be disordered in solution (**Fig. 6.4**). Additionally, with the exception of three residues (residues 8 to 10, here termed the anchor string) this extension is disordered in both monomers in the SimR-SD8 structure, and it is only partially ordered in one monomer in the SimR-SC4 structure (Le *et al.*, 2011b and Chapter 5). The TFR arm is ordered in the SimR-apo structure, but its structure is the likely result of crystal packing (**Fig. 6.5**).

Because disordered regions are often hypersensitive to proteolysis (Receveur-Brechot *et al.*, 2006), we examined the sensitivity of SimR to trypsin. The TFR arm was rapidly digested, leaving a much more stable product with an N-terminus at either residue Ser20 or Ser23 (**Fig. 6.6** and **Fig. 6.7)**. Taken together, these observations suggest that the TFR arm is solvent exposed and displays conformational flexibility in solution in the absence of cognate DNA.

Since many unstructured regions exhibit increased resistance to proteolysis on binding of a partner (Dyson and Wright, 2005; Receveur-Brechot *et al.*, 2006), we determined the effect of DNA binding on the sensitivity of the TFR arm to trypsin. Addition of 25- or 31-bp DNA duplexes spanning the O_X operator substantially decreased the rate of SimR proteolysis, suggesting that DNA binding renders the TFR arm more resistant to trypsin (**Fig. 6.6**). Consistent with this interpretation, proteolysis was not inhibited when a 15-bp O_X DNA duplex that is unable to bind to SimR was incubated with SimR (**Fig. 6.6** and **Fig. 6.8A).** In total, these experiments suggest that the TFR arm transitions from a disordered or conformationally flexible state to a more ordered, rigid state upon DNA binding.




```
1   MNENEPVSIW MHPEPAGRRS ARSHRTLSRD QIVRAAVKVA DTEGVEAASM 50
51  RRVAAELGAG TMSLYYYVPT KEDLVELMVD EVIGETRLPD RPGPDWRAAL 100
101 TLAANEKRAL WLRHPWLATA WRNGHPVWGP NSLRQQEFVL GTLGVFDLQV 150
151 DELLSLIGLY NGYVESFVRN EVGWLEEARR TKVDMREWMR RSGPYAQQLV 200
201 DSGEYPMFAR VLAETVAPHM GPDQRFRSGL ERLLDSIGAS LDRLSPPGRS 250
251 AASERPALA
```

Fig. 6.4. Disorder probability profile plot for SimR from the PrDOS and RONN server. PrDOS and RONN predict that at least 28 N-terminal amino acids (red) and the 12 C-terminal residues (red) are disordered.

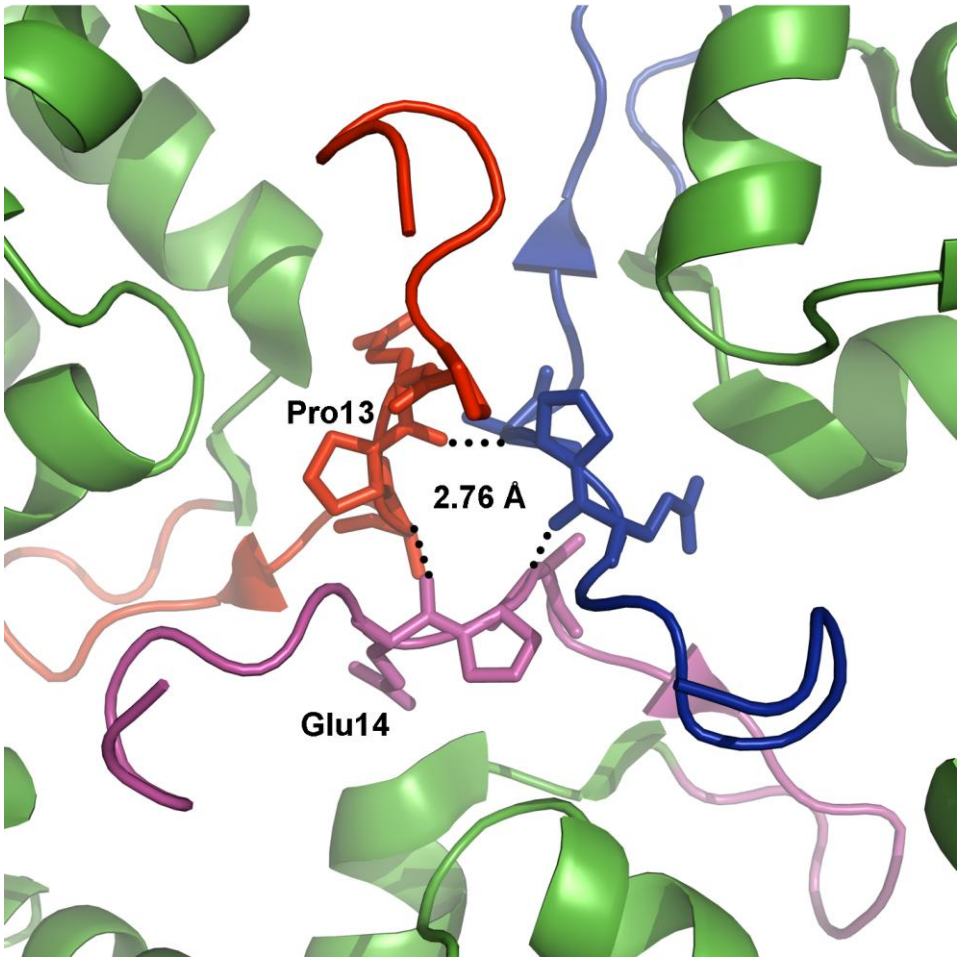


Fig. 6.5. In the SimR-apo structure the TFR arm is ordered because it is stabilised by interaction with two other symmetry-related arms. The three symmetry-related TFR arms are shown in red, blue and magenta. Inter-strand hydrogen bonds (dotted black lines) between the main chain O of Pro13 and the NH of Glu14 contribute to the stabilisation. There is only one SimR-apo monomer in the asymmetric unit.

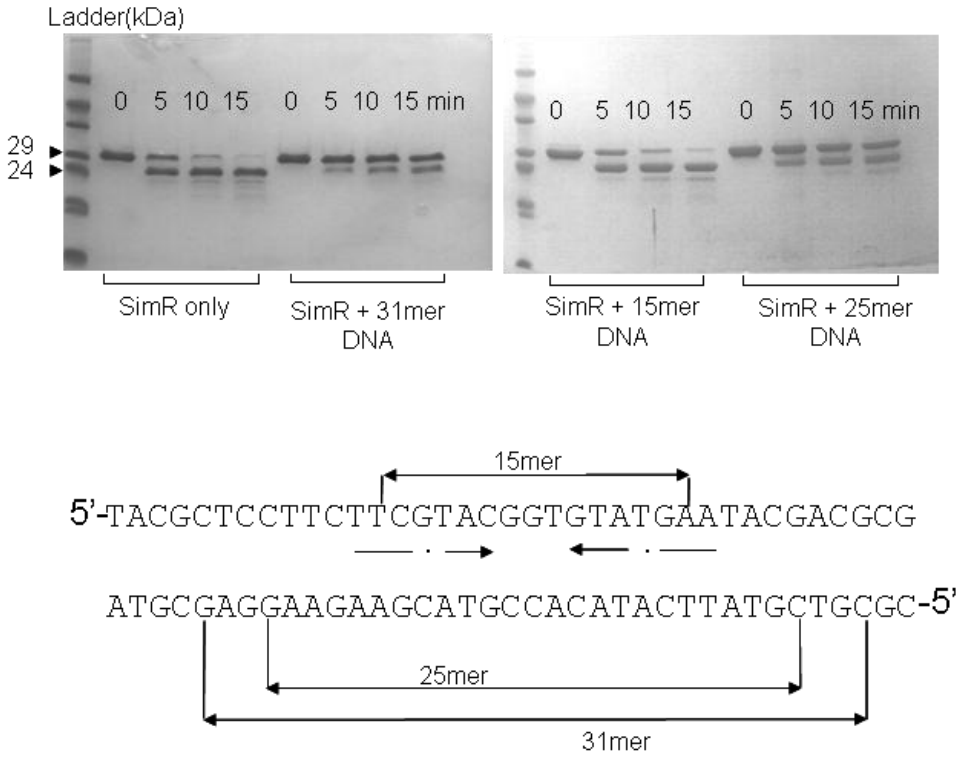


Fig. 6.6. Limited tryptic proteolysis of SimR in the presence or absence of DNA. SimR was incubated either alone or with the O_X operator DNA duplexes indicated, before the addition of trypsin. Note that the 15mer DNA duplex does not bind SimR (see **Fig. 6.8A**). After SDS-PAGE, SimR species were visualised by Coomassie blue staining. The major product of tryptic digestion (arrowed) was shown by Edman sequencing to have an N-terminus corresponding to Ser 20 or Ser23 of wild-type SimR.

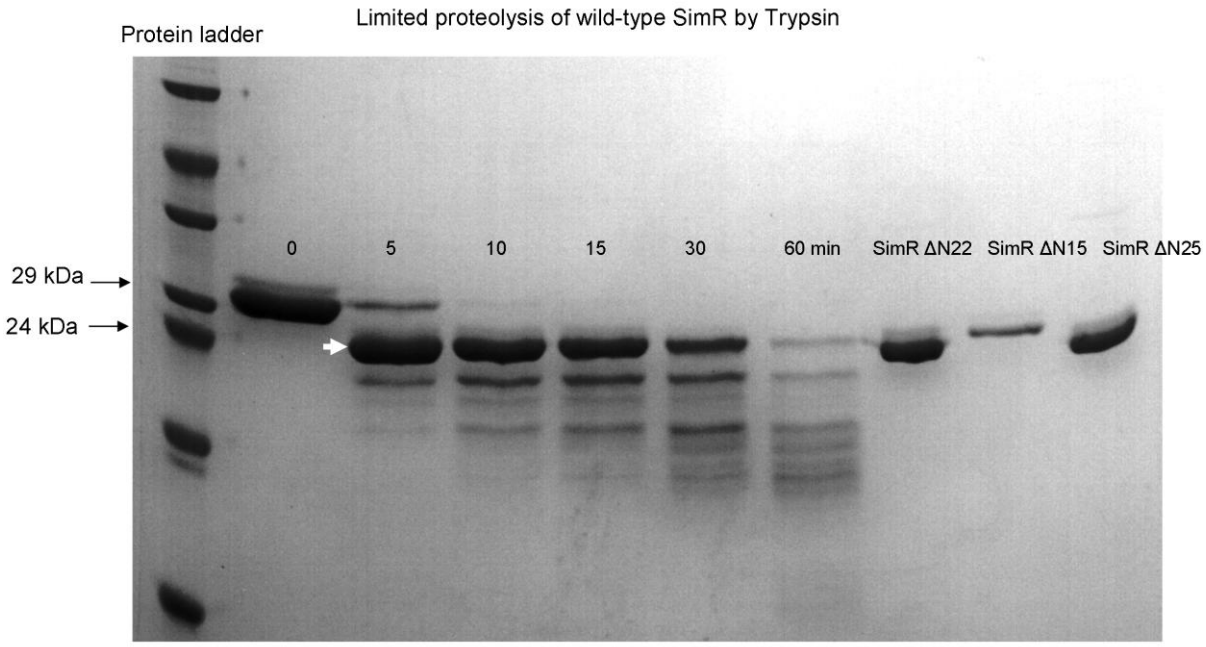


Fig. 6.7. A limited proteolysis time course shows that the TFR arm of SimR is highly susceptible to trypsin, consistent with it being conformationally flexible in solution. N-terminal sequencing of the most prominent band after 5 min digestion (arrowed) showed that the band is composed of two closely-migrated SimR fragments. One fragment is with the N-terminus at residue Ser23 (4 pmoles), while the other fragment starts with Ser20 (2-3 pmoles).

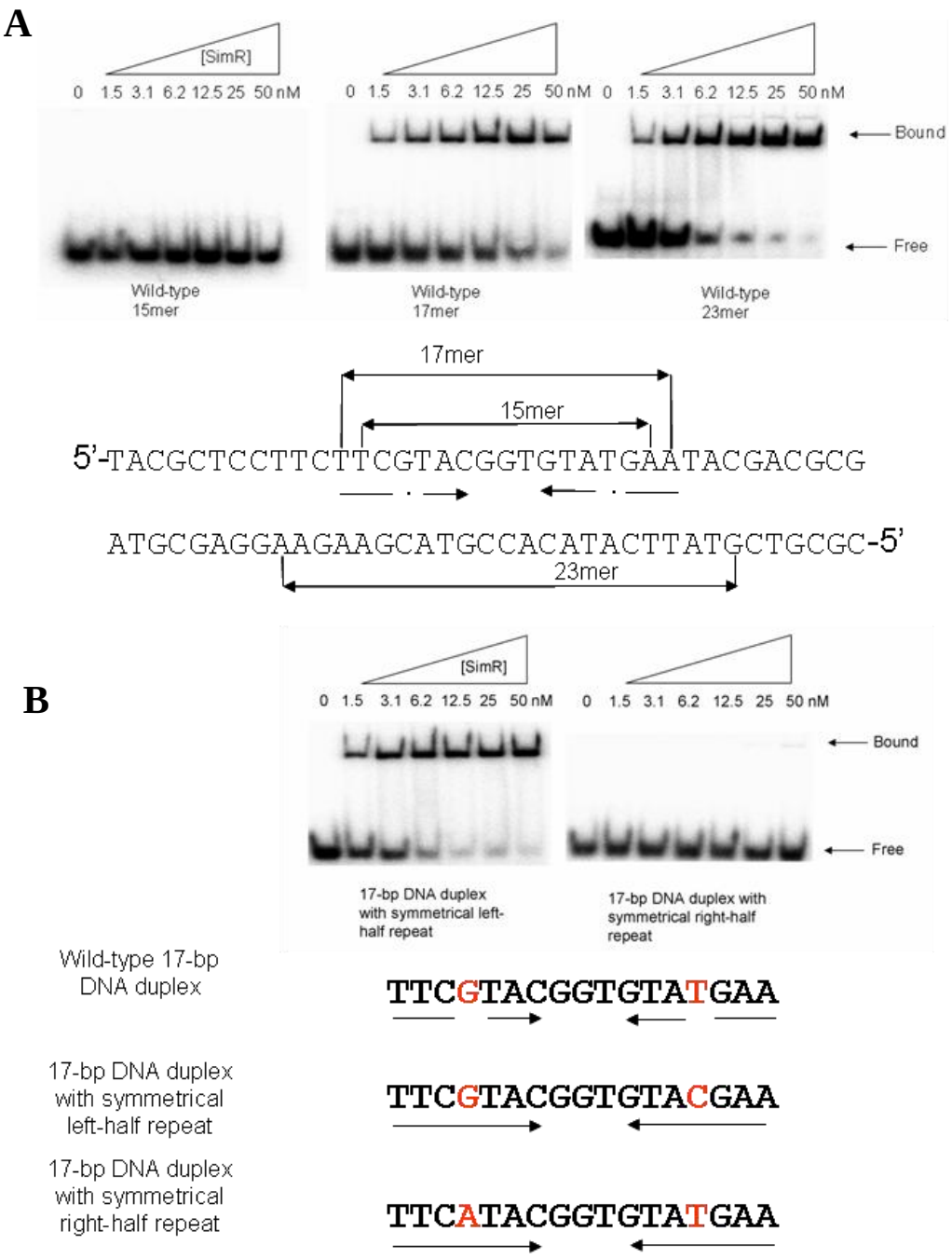


Fig. 6.8. Electrophoretic mobility shift assays (A) using wild-type SimR and radiolabelled DNA duplexes of different length encompassing the wild-type *OX* operator. (B) using wild-type SimR and radiolabelled DNA duplexes encompassing the symmetrical operators based on either the right-half repeat or the left-half repeat. Binding reaction mixtures were resolved on 8% (w/v) native PAGE gels.

6.4 The structure of SimR bound to its DNA operator

To understand how SimR binds to its operator sequence and to shed light on the role of the TFR arm in DNA binding, we crystallized SimR in complex with DNA. We tested DNA duplexes from 17 to 21 bp in length and found that only the minimal, blunt-ended 17-bp duplex crystallized in complex with SimR. The 17-bp DNA duplex used was the O_X operator (TTCGTACGGTGTATGAA), but carrying two base-pair changes to generate a near perfect inverted repeat (5'-TTCGTACGGCGTACGAA-3'), which bound SimR at least as tightly as the wild-type 17-bp DNA duplex **(Fig. 6.8B)**. We solved the structure of full-length SimR (residues 1-259) in complex with this 17-bp DNA duplex to 2.99 Å resolution (**Fig. 6.9A**). X-ray data collection and refinement statistics are summarised in **Table 6.1**. The asymmetric unit (ASU) contained two SimR dimers, each bound to a 17-bp DNA duplex. The two SimR dimer-DNA complexes are essentially identical [root mean square deviation (RMSD) between complexes for the Cα backbone = 0.15 Å], and thus only one complex is discussed throughout (**Fig. 6.9A**). The conformation of bound DNA is mostly regular B-form but is bent away from the SimR dimer by ~15^{o} (see below and **Fig. 6.18A**). The bases at the end of adjacent DNA duplexes stack and interact to form a pseudo-continuous double-helical DNA filament running through the crystal (**Fig. 6.9B** and **Fig. 6.10**).

6.5 Interactions between the HTH motif and the major groove

The core DBD is composed of helices α1 to α3 (residues 29-67). Helix α2 (residues 49-58) and the recognition helix α3 (residues 61-67) form the HTH motif which packs against α1 for stabilisation (**Fig. 6.9A**). Surprisingly, the recognition helix makes no canonical hydrogen bonds with the bases. However, the side chain of Met62 makes a series of contacts to three different bases including van der Waals to C3 (Cβ to C^5), and an uncommon electrostatic interaction between the Sδ atom and the face of the base of T12, which is analogous to Sδ stacking over the aromatic side chains of tryptophan,

Data set	SimR-DNA (17mer)

Data collection	
Space Group	$P2_12_12_1$
Cell parameters (Å/°)	a = 85.8, b = 112.6, c = 163.7
Solvent content (%)	62.5
Wavelength (Å)	1.11
Resolution range[a] (Å)	92.78-2.99 (3.15-2.99)
Unique reflections[a] (#)	31030 (4547)
Completeness[a] (%)	95.2 (96.5)
Redundancy[a]	3.1 (3.1)
R_{merge}[b] (%)	10.0 (59.1)
$\langle I \rangle / \langle \sigma I \rangle$	8.3 (1.8)
Wilson B value ($Å^2$)	53.4
Refinement	
R_{cryst}[c] (based on 95% of data)	20.9
R_{free}[c] (based on 5% of data)	25.1
Coordinate error[d] (Å)	0.420
Ramachandran favoured[e] (%)	98.0
Ramachandran outliers[e] (%)	0.22
rmsd bond distances (Å)	0.010
rmsd bond angles (°)	1.22
Mean B-value for protein ($Å^2$)	57.6
Mean B-value for the DNA ($Å^2$)	54.6
Contents of model	
Protein residues in each chain (totals in brackets)	A: 7-241 B: 7-15 and 26-242 C: 7-242 D: 7-15 and 26-242
DNA nucleotides	E and F, G and H : 1-17
PDB accession code	3ZQL

[a] The figures in brackets indicate the values for highest resolution shell.

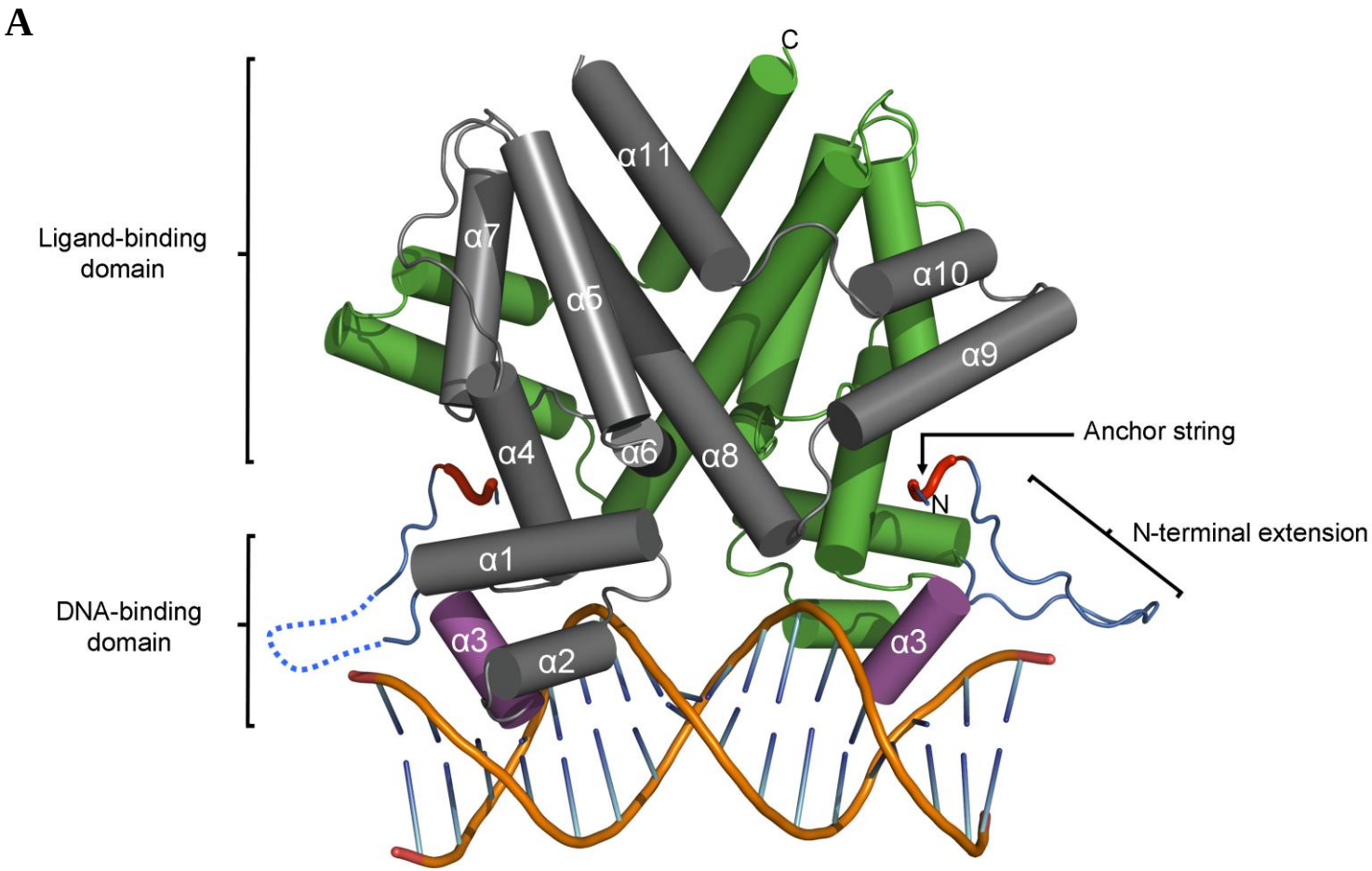
[b] $R_{merge} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ is the ith observation of reflection hkl and $\langle I(hkl) \rangle$ is the weighted average intensity for all observations i of reflection hkl.

[c] The R-factors R_{cryst} and R_{free} are calculated as follows: $R = \sum(|F_{obs} - F_{calc}|) / \sum|F_{obs}| \times 100$, where F_{obs} and F_{calc} are the observed and calculated structure factor amplitudes, respectively.

[d] Estimate of the overall coordinate errors calculated in REFMAC5 based on R_{free} (Murshudov *et al.*, 1997)

[e] As calculated using MOLPROBITY (Chen *et al.*, 2010)

Table 6.1. Summary of SimR-DNA X-ray data and model parameters


A
C
α11
α7
α5
α10
α9
Ligand-binding domain
Anchor string
α4
α6
α8
N
N-terminal extension
α1
DNA-binding domain
α3
α2
α3

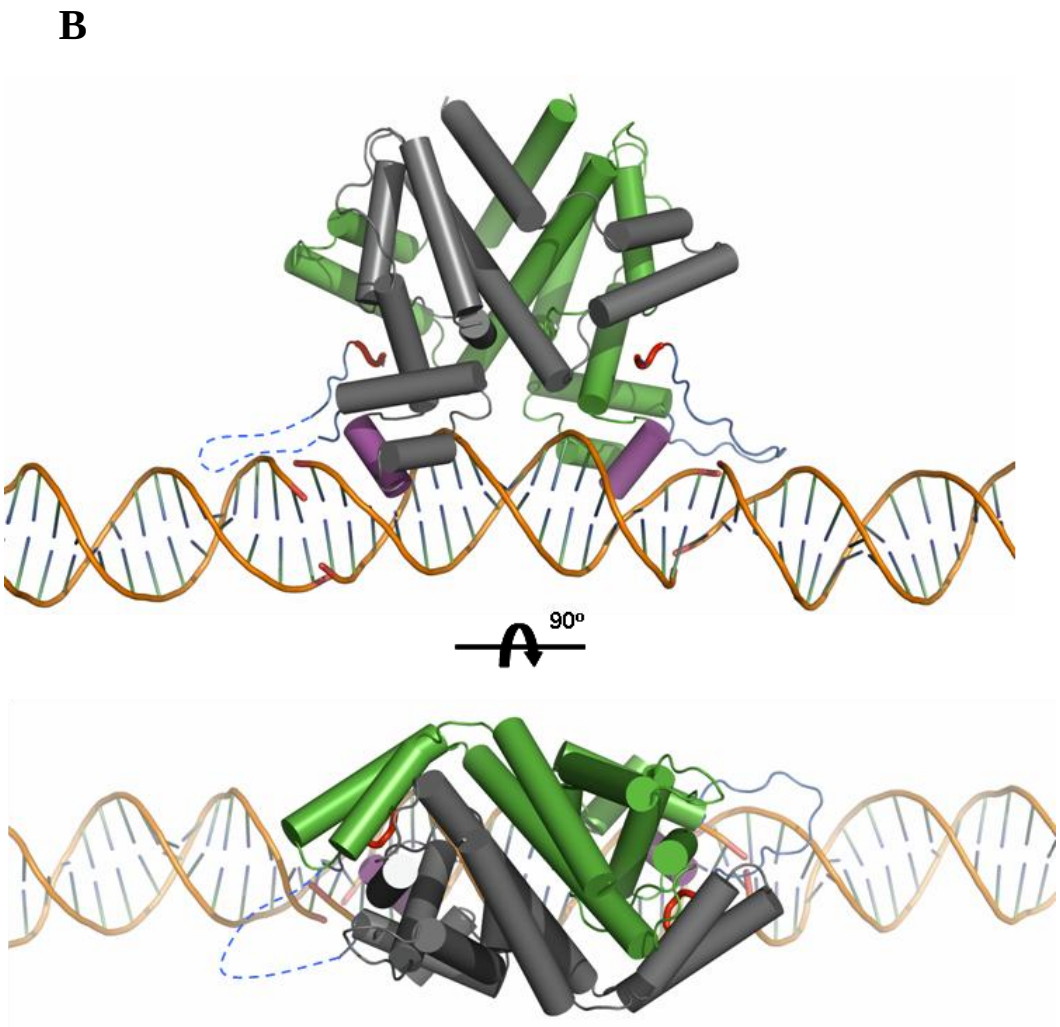


Fig. 6.9. Structure of the SimR-17mer complex (A) in isolation or (B) showing the adjacent DNA duplexes in the crystal. A cylindrical helix representation is used to highlight the secondary structure of SimR with key features labelled in (A). One subunit of the biological-relevant dimer is shown in grey and one in green. The recognition helix α3 is shown in magenta, the TFR arm is shown in blue, and the N- and C-termini are labelled. The anchor string of the TFR arm (residues 8-11) is shown as a red tube cartoon. The dotted blue line represents the disordered TFR arm in the left-hand SimR subunit. In (B) only the DNA components of the adjacent symmetry complexes are shown in order to highlight the pseudo-continuous DNA filament running through the crystal (See also **Fig. 6.10** and **Fig. 6.14**).

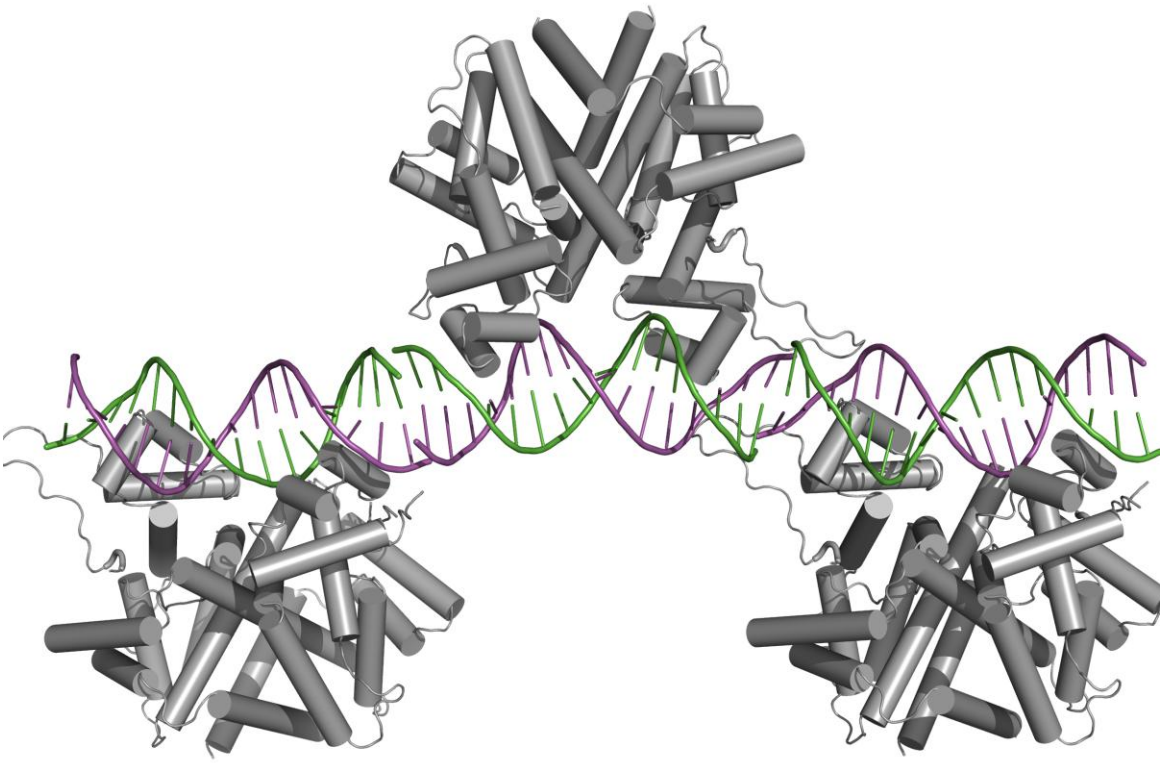


Fig. 6.10. Structure of the SimR-17mer duplex in the context of the crystal. Complementary DNA strands are shown in magenta and green. Neighbouring DNA duplexes stack and interact to form pseudo-continuous double-helical DNA filaments running through the crystal.

histidine and phenylalanine (Pal and Chakrabarti, 2001) (**Fig. 6.11**). This interaction is buttressed by van der Waals contacts to the C^7 methyl group of T12. The SD atom of Met62 also accepts a hydrogen bond from the N^6 hydrogen bond donor of A13. Another key interaction involved in the DNA sequence recognition mechanism of SimR is the stacking of the side chain of residue Tyr66 with the C7 exocyclic methyl groups of T1 and T2. This interaction explains in great part why SimR has a higher affinity for the O_X operator, which has this pair of thymines, than for O_R, which has a pair of guanines at these positions (Le *et al*., 2009 and Chapter 4). The dominant recognition helix interactions are with the phosphate backbone. For each operator half-site, there are hydrogen bonds between the hydroxyl group of Ser63 and the phosphate group of C3, between the hydroxyl group of Tyr65 and the phosphate group of T12, and between Tyr67 and the phosphate group of T2 (**Fig. 6.11**). Just outside helix α3, the backbone NH group of Gly60 hydrogen bonds with the phosphate group of C3. On binding DNA, the recognition helix adopts a 3_{10} helical conformation, in contrast to the canonical α-helical conformation seen in the structures of SimR-apo and SimR-simocyclinone complexes (Le *et al.*, 2009 and Chapter 4). This conformational alteration in the recognition helix on DNA binding is also observed in TetR, and is believed to facilitate intimate interaction with the DNA (Orth *et al.*, 2000).

Three residues in helix α2 contribute to DNA binding, with the side chain hydroxyl group of Ser49 forming a hydrogen bond with the phosphate backbone of C10 and the backbone NH group of Met50 forming a hydrogen bond with the phosphate backbone of G11 (**Fig. 6.11).** The guanidinium group of Arg51 is involved in direct base recognition by bifurcated hydrogen bonds from NH2 to the O^6 and N^7 acceptors of G11. Other interactions between SimR and the major groove are hydrogen bonds between the amino group of Lys71 and the phosphate group of G11, and between the backbone NH group of Lys71 and the phosphate group of T12. Lys71 lies at the N-terminus of helix α4 at the very beginning of the LBD, just outside the core HTH motif of the DBD. This residue is highly conserved among TFRs and the equivalent lysine in TetR also forms a hydrogen bond with the phosphate backbone (Orth *et al.*, 2000).

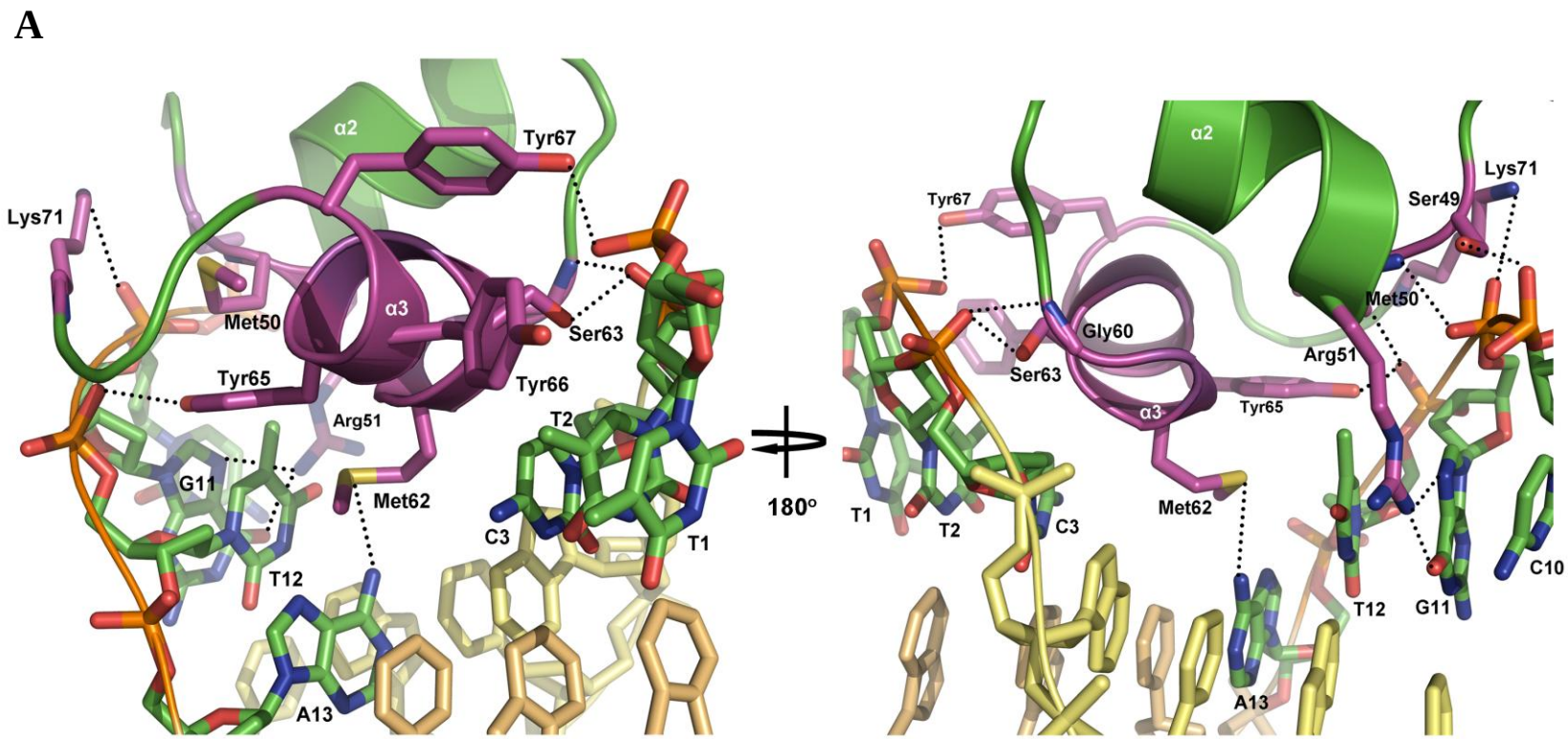


Fig. 6.11. (A) Interactions between the HTH motif and the major groove. Stick representations of the interacting residues are shown in magenta. The Cα backbone of recognition helix α3 is shown in magenta and that of helix α2 is shown in green. Hydrogen bonds are represented by dotted black lines. The interacting bases are labelled and only the ring frames are shown for non-interacting bases.

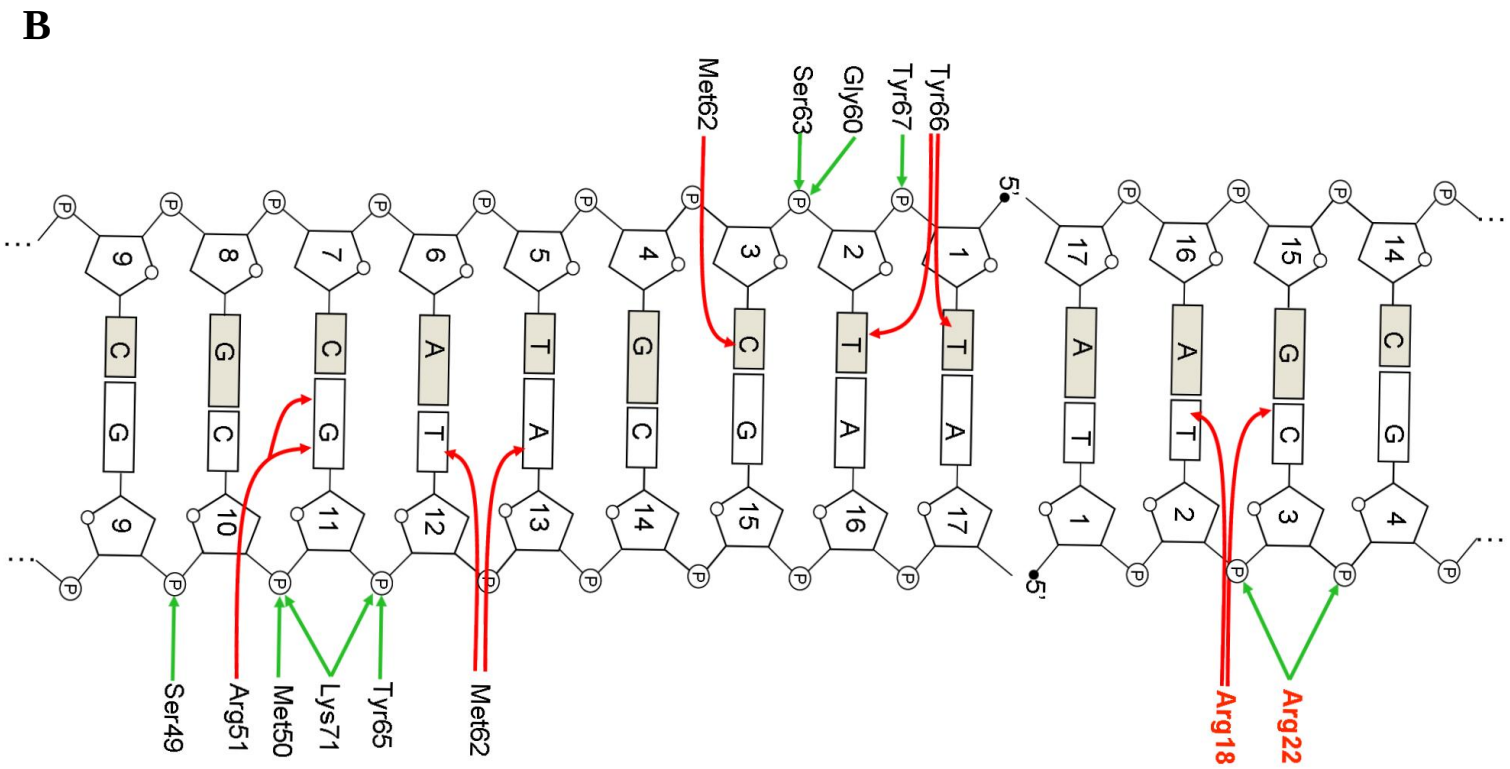


Fig. 6.11. (B) Schematic representation of SimR-DNA contacts. For simplicity, only a recognition half-site and the first four base pairs of an adjacent duplex are shown. Interactions between amino acid residues and the bases of the cognate DNA operator are indicated by red arrows, and those between amino acid residues and the phosphate backbone are represented by green arrows. Amino acid residues belonging to TFR arm are shown in red.

TFRs frequently rely on phosphate backbone contacts to mediate interaction with the DNA. In an extreme case, the DesT-DNA interface involves 11 phosphate backbone contacts but only two specific interactions with a pair of guanine bases within each half site (Miller *et al*., 2010). In contrast, TetR and QacR make extensive direct hydrogen-bond contacts with the bases (Orth *et al*., 2000; Schumacher *et al*., 2002). In this sense, SimR is perhaps more similar to DesT than to TetR or QacR in its DNA sequence recognition mechanism. Thus, although the overall structure of the DBD in TFRs is conserved, it is clear that the mode of operator recognition differs from one member of the TFR family to another (Itou *et al*., 2010; Miller *et al*., 2010; Orth *et al*., 2000; Schumacher *et al*., 2002). TFRs recycle conserved residues and inventively employ non-conserved ones within the DBD for either base-specific hydrogen bond formation or for phosphate backbone contacts (**Fig. 6.1**). It seems that there is no deterministic set of rules for TFR-DNA recognition.

6.6 Interactions between the TFR arm and the minor groove

If the structure of a single SimR-DNA complex is viewed in isolation, it can be seen that the TFR arm does not make contact with the cognate DNA duplex (**Fig. 6.9A**). Instead, the TFR arm binds the minor groove of the adjacent DNA duplex in the pseudo-filament (**Fig. 6.9B**). This binding to the minor groove is mediated through arginine residues that sit at the tip of the TFR arm (**Figs. 6.11B**, **6.12** and **6.13B**). Specifically, N2 of the guanidinium group of Arg18 forms a hydrogen bond with O^2 of C3, whilst its N1 interacts with the O^2 of T2, the guanidinium group of Arg22 forms two salt bridges to the phosphate backbone of C3 and G4 (**Figs. 6.11B** and **6.12**). The electropositive side chain of Arg18 is deeply buried in this minor groove (**Fig. 6.12**), where the electronegative potential of the phosphate backbone is focused (Rohs *et al.*, 2009; Rohs *et al.*, 2010). This helps anchor the tip of the TFR arm in the minor groove. A third arginine in the flexible TFR arm, Arg19, does not contact DNA in the structure reported here (**Fig. 6.12** and **6.13**). However, given the noncovalent nature of the DNA pseudo-filament, we considered the possibility that Arg19 might be involved in DNA binding in

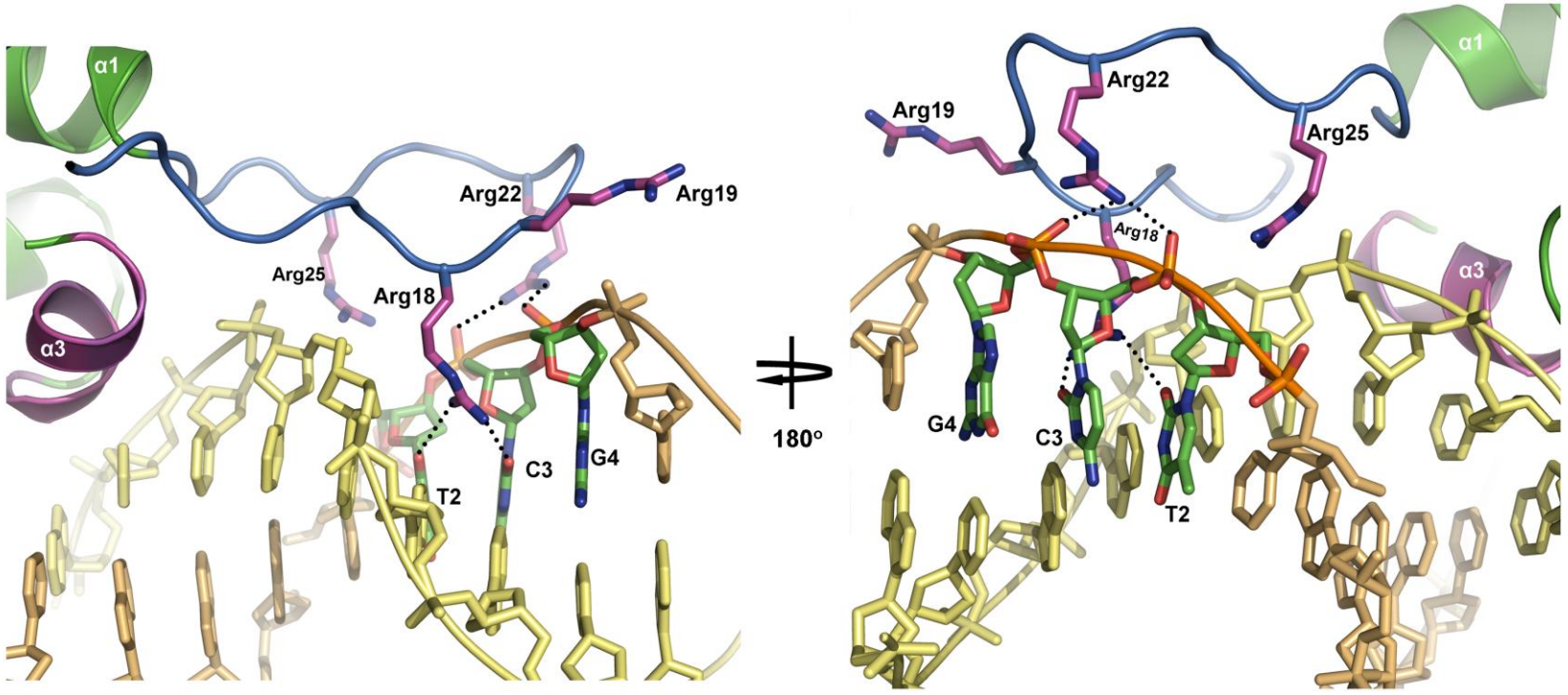


Fig. 6.12. Interactions between the TFR arm and the minor groove. The Cα backbone of the TFR arm is shown in blue and stick representations of arginine residues Arg18, Arg19, Arg22 and Arg25 are shown in magenta. Hydrogen bonds are represented by dotted black lines. The interacting bases are labelled and only the ring frames are shown for non-interacting bases.

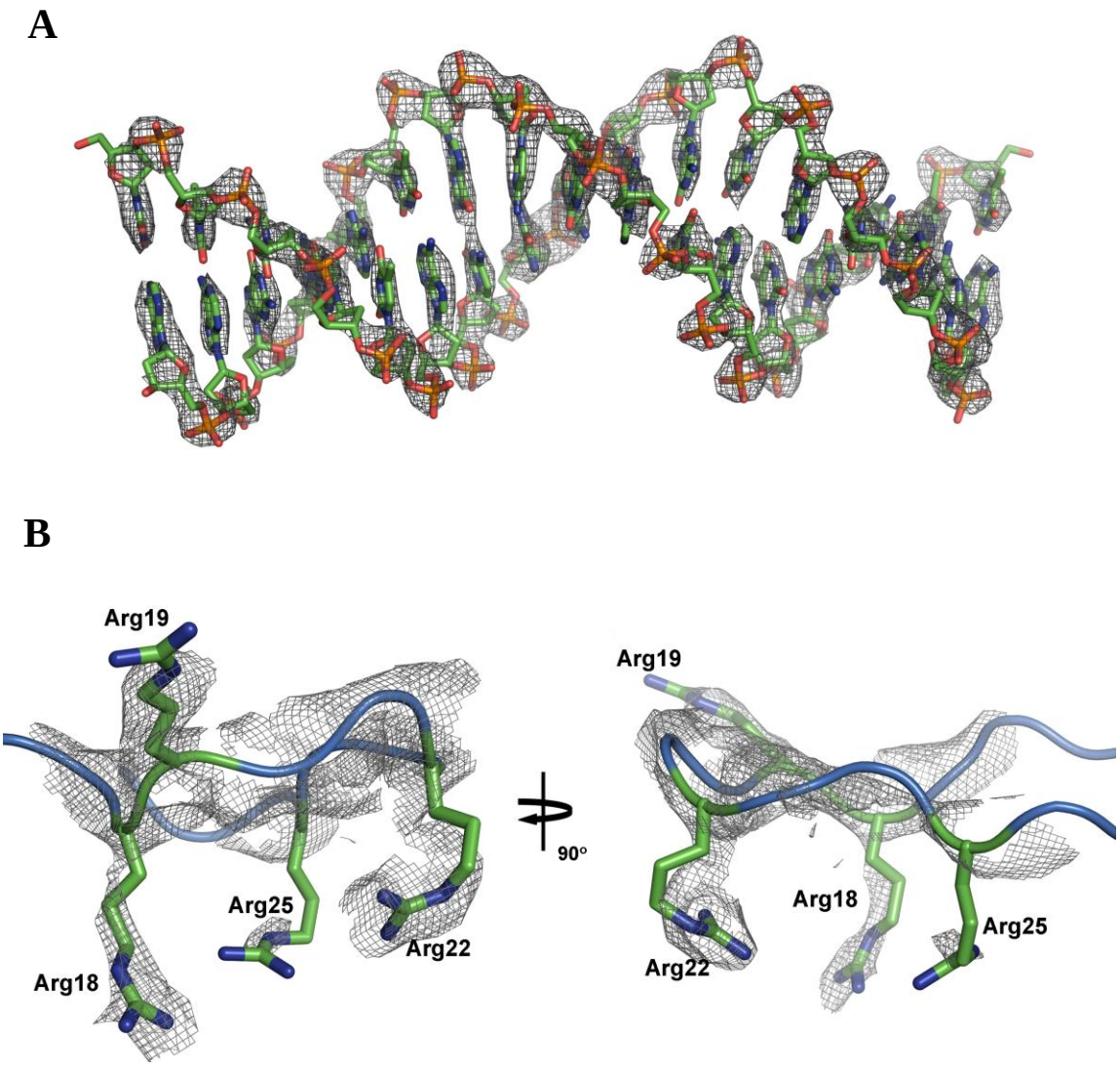


Fig. 6.13. Simulated annealing omit maps at 2.99Å resolution, contoured at 2.5 σ for (A) the DNA, and (B) residues 18-25 of the TFR arm. The arm is shown in blue, and residues Arg18, Arg19, Arg22 and Arg25 in green.

truly continuous double-stranded DNA. To examine this possibility, we mutagenised Arg19 to alanine and assayed the resulting protein for its ability to bind to the *simR-simX* intergenic region by EMSA. SimR R19A bound DNA with an affinity equal to that of wild-type SimR (**Fig. 6.14**), suggesting Arg19 does not contribute to DNA binding. In contrast, when we constructed SimR R18A and SimR R22A variants, we found that each exhibited an approximate 15-fold reduction in binding affinity (**Fig. 6.14**), consistent with roles for R18 and R22 in DNA binding, as suggested by the structure of the SimR-DNA complex.

Initially, it was difficult to understand why SimR variants lacking just 10 or 15 amino acids N-terminal residues should have reduced DNA binding affinity, given that they retain the interacting arginine residues. In the previously solved structures of apo-SimR and SimR-ligand complexes, although the TFR arm is mostly disordered, residues 8 to 10, herein termed the anchor string, are always visible in electron density maps (Le *et al.*, 2011b and Chapter 5), probably because this string of amino acid residues is stabilised by van der Waals interactions with the cleft between the LBD and the DBD. It therefore seems likely that this short segment, highlighted in red in **Fig. 6.9**, serves as an anchoring point for the TFR arm to loop back onto the body of SimR. This arrangement may be important for restricting the flexibility of the TFR arm, so that it is poised appropriately to interact with the minor groove. Deleting 10 or 15 amino acids from the N-terminus would remove this anchor point, destabilising loop formation and reducing DNA binding affinity. The more severe deletions, removing 22 or 25 amino acids, further reduce binding affinity because they remove the interacting arginine residues themselves.

In the crystal structure of the SimR-DNA complex, the TFR arm is seen in one SimR subunit but is disordered in the other subunit (**Fig. 6.9**). From an inspection of the end-to-end base stacking between adjacent DNA duplexes within the crystal, it is clear that the two ends are not equivalent. The stacking at the right-hand end (as viewed in **Fig. 6.15**) allows the neighbouring DNA strands to transit smoothly across the gap, producing a relatively normal minor groove. However, on the left-hand end the strands

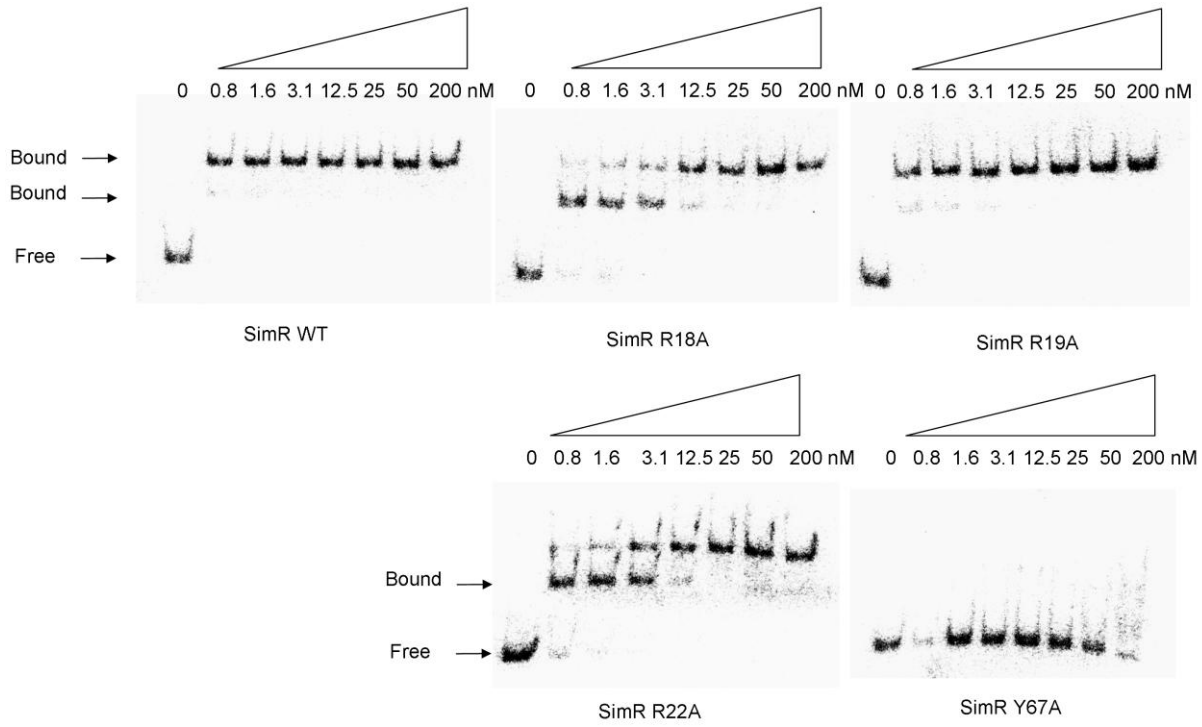


Fig. 6.14. Electrophoretic mobility shift assay showing the binding of the purified wild-type or R18A, R19A, R22A and Y67A derivatives of SimR to the *simR-simX* intergenic region. Bands correspond to SimR-DNA complexes (Bound) and free DNA (Free) are indicated. Final concentration of SimR is indicated above each lane.

Point mutations to substitute specific positions in the HTH motif (Y67) or the TFR arm (R18, R19, R22) of SimR were introduced using a modified QuikChange site-directed mutagenesis procedure (Stratagene). The mutation Y67A, which is located outside the TFR arm and makes a DNA phosphate backbone contact, serves as a control for this experiment.

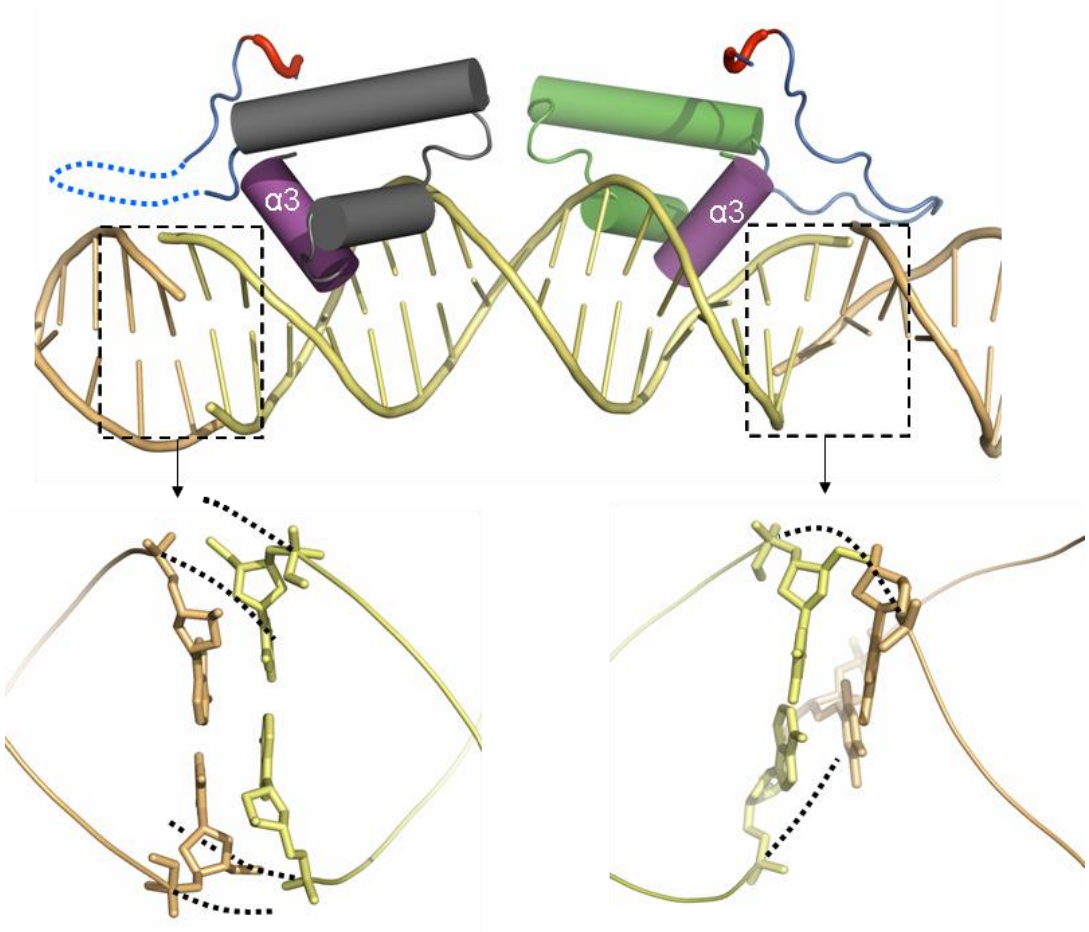


Fig. 6.15. Non-equivalent stacking between adjacent DNA duplexes in the crystal pseudo-filament creates two different minor grooves. Only the DBD of SimR is shown. At the right-hand end of the central DNA duplex the base stacking allows the DNA phosphate backbone to transit smoothly (dotted lines) between adjacent duplexes, creating a relatively normal minor groove. At the left-hand end of the central DNA duplex the base stacking causes the phosphate backbone to veer away to avoid a steric clash (dotted lines), producing an abnormal minor groove. Adjacent DNA duplexes are shown in contrasting colours.

veer away to avoid a steric clash while maintaining base pair stacking, producing a much wider minor groove (**Fig. 6.15**). It seems likely that the TFR arm is unable to interact with this "abnormal" minor groove and is therefore disordered in the crystal. In the structure of the SimR-17mer duplex, apart from the interaction of the anchor string with the body of SimR, the only contacts made by the TFR arm are with the minor groove of DNA (**Fig. 6.9** and **Fig. 6.10**). Based on the crystal structure of the SimR-DNA complex and the results of the proteolysis-protection assays, we propose that the TFR arm transitions from a disordered or conformationally flexible state to a more ordered state upon binding to its cognate DNA.

6.7 N-terminally truncated SimR derivatives have a smaller footprint on DNA than wild-type SimR

We used DNaseI protection to compare the footprints of wild-type SimR and the N-terminally truncated SimRs on the O_X and O_R operators in the *simR-simX* intergenic region (**Fig. 6.16A**). In each case, saturating amounts of SimR protein were used to ensure complete protection of the binding sites. The footprint for wild-type SimR was comparable with that reported previously (Le *et al.*, 2009 and Chapter 4). In contrast, in the footprints generated using the N-terminally truncated SimR proteins, the edge of the protected region retracted at both ends of the footprint when compared to the footprint of full-length SimR (**Fig. 6.16**). Specifically, when N-terminally truncated proteins were used, on the upper DNA strand the O_R footprint retracted by two base pairs at the left edge and by one base pair at the right edge (**Fig. 6.16)**. No retraction of the O_R footprint was apparent on the lower DNA strand. When N-terminally truncated proteins were used, on the upper DNA strand the left edge of the O_X footprint retracted by one base pair, while no retraction was apparent at the right edge (**Fig. 6.16)**. On the lower DNA strand, the O_X footprint receded by one base pair at both ends. These observations indicates that the TFR arm sterically hinders DNaseI, protecting one additional

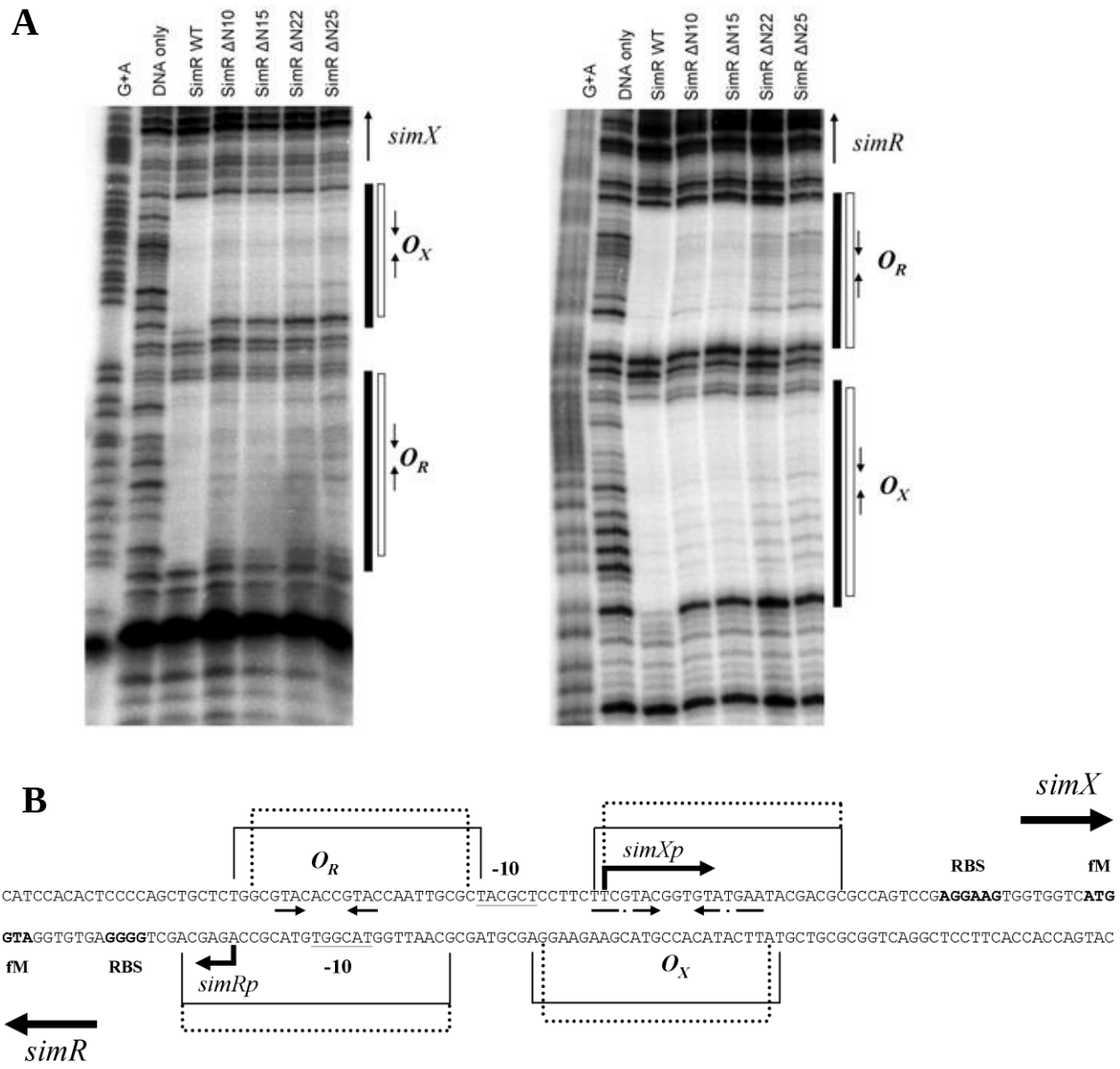


Fig. 6.16. (A) DNaseI footprinting analysis of the binding of wild-type and N-terminally truncated derivatives of SimR to the *simR-simX* intergenic region. A DNA fragment containing the *simR-simX* intergenic region, 5'-end labelled on either the upper strand (left panel) or the lower strand (right panel), was exposed to DNaseI in the presence of saturating concentrations of SimR protein (200 nM for wild-type SimR, SimRΔN10 and

(cont next page)

SimRΔN15; 400 nM for SimRΔN22 and SimRΔN25). The sequencing ladders were generated by subjecting the probes to Maxam-Gilbert G+A chemical sequencing. Regions protected from DNaseI cleavage (operators O_X and O_R) by wild-type SimR are indicated by solid vertical bars, and those protected by the N-terminally truncated SimR derivatives are indicated by open bars. Inverted repeats within the DNaseI protected regions are indicated by convergent arrows. (B) Sequence of the *simR-simX* intergenic region summarising the DNaseI footprinting data. Regions protected by wild-type SimR are indicated by solid lines, and those protected by the N-terminally truncated SimR derivatives are indicated by dotted lines. Also indicated are the *simRp* and *simXp* transcription start points and putative -10 sequences, the *simR* and *simX* ribosome binding sites (RBS), and the imperfect inverted repeats within the footprints.

phosphodiester bond from cleavage by the nuclease. Each SimR mutant protein produced the same footprint, regardless of whether 10, 15, 22 or 25 amino acids had been deleted from the N-terminus, consistent with the idea that residues 8 to 10, (i.e. the anchor string), are needed for the TFR arm to be fully functional, as discussed above. Note that the retraction of the footprint occurs at both ends of the operator, suggesting that the TFR arms of both monomers in the SimR dimer function in solution. We also performed a complementary experiment to determine the binding affinity of wild-type SimR to three DNA duplexes of different lengths (15, 17 and 23 bp) spanning the O_X inverted repeat sequence. The 23-bp duplex bound SimR more strongly than the minimal 17-bp duplex, showing that DNA flanking the core 17-bp inverted repeat contributes to SimR binding (**Fig. 6.8A**). The 15-bp duplex failed to bind SimR (**Fig. 6.8A**). In addition, although the minimal 17-bp duplex binds to SimR relatively well (**Fig. 6.8A**), it is unable to protect the TFR arm of SimR from tryptic digestion, while a 23mer reduced the rate of proteolysis considerably **Fig. 6.17**). Taken together, these observations suggest that, in solution, the TFR arm interacts with DNA outside the core 17-bp O_X operator, consistent with the SimR-DNA structure, which shows dimer-DNA interactions spanning 21 base pairs.

Among the five TFRs for which protein-DNA crystal structures are available (TetR, DesT, CgmR, QacR and SimR; **Fig. 6.1**), only SimR possesses a flexible TFR arm that undergoes a transition to an ordered state upon DNA binding. DesT has a 12-amino acid residue N-terminal extension (**Fig. 6.1)** but it is not disordered, instead forming part of an extended helix α1. Residues Arg5 and Lys9 of this short N-terminal extension in DesT nevertheless contribute to DNA binding (Miller *et al.*, 2010), which is unusual because the main role of helix α1 is in stabilising the HTH motif (α2 to α3). Residues N-terminal to the core DBD in two other TFRs, *Neisseria gonorrhoeae* MtrR (11 amino acids) and *Streptomyces coelicolor* ActR (32 amino acids) have also been suggested to be involved in DNA binding (Hoffmann *et al.*, 2005; Willems *et al.*, 2008), implying a possible common role for TFR N-terminal extensions when present (see also the global

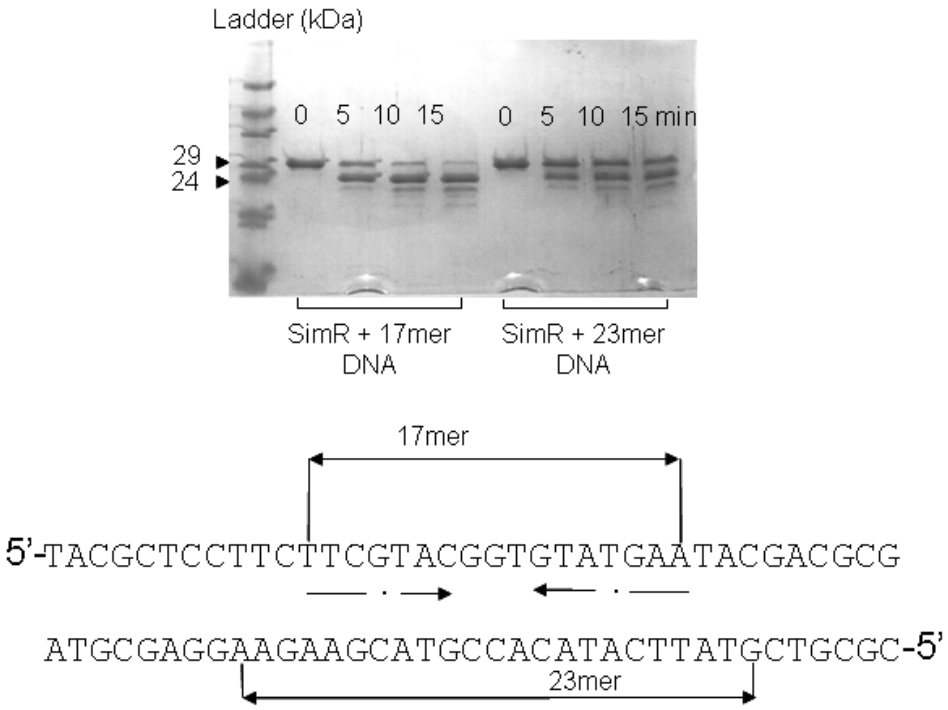


Fig. 6.17. Limited trypsin proteolysis of SimR with binding DNA. SimR was incubated with either a 17mer or a 23mer DNA duplexes encompassing the inverted repeat of *OX* operator before trypsin digestion. While a 17mer duplex binds to SimR (See also **Fig. 6.8**), it is unable to protect the N-terminal extension from trypsin digestion as well as the 23mer duplex.

TFR bioinformatic analysis presented below). Similar kinds of extensions have been identified in at least two other families of DNA-binding proteins. For example, members of the eukaryotic Hox family recognise nearly identical major groove sequences through the recognition helix of their homeodomain but use an extended arm to insert into the minor groove to enhance binding specificity (Joshi *et al.*, 2007). A related example is phage lambda repressor, which has a conventional HTH motif and an additional N-terminal extension that promotes DNA binding, in this case by interacting with the major groove (Beamer and Pabo, 1992). A comprehensive analysis of all available protein-DNA structures has shown that the binding of arginine residues to narrow minor grooves is a widely used mechanism in protein-DNA recognition. This readout mechanism exploits the fact that narrow minor grooves, often associated with A-tracts, strongly enhance the negative electrostatic potential of the DNA (Rohs *et al.*, 2009; Rohs *et al.*, 2010). However, it should be noted that the minor groove bound by the TFR arm of SimR is not associated with an A-tract, which likely contributes to its slightly enlarged groove width (**Fig. 6.19**).

6.8 The arginine- and lysine-rich TFR arm is likely to be a common feature of TetR family members

We searched the PFAM database (http://pfam.sanger.ac.uk/) for proteins that match the Hidden Markov Model profile PF00440, identifying 12,715 non-redundant TFRs (see Materials and Methods for further details). The amino acid sequences of these TFRs were then aligned using MUSCLE (Edgar, 2004) to identify the core DBD and any N-terminal extension. 28% had N-terminal extensions of less than 10 amino acids, 44% had N-terminal extensions of 11-20 amino acids, 17% had N-terminal extensions of 21-30 amino acids, and 11% had N-terminal extensions longer than 31 amino acids. Further, the fraction of Arg and Lys residues in these N-terminal extensions (mean value = 20.5%) was almost double the frequency found in the globular body of the TFRs (mean value = 11.4%) (**Fig. 6.18A**). Finally, the RONN server predicts that the majority

of these N-terminal extensions are likely to be disordered in solution (**Fig. 6.18B**). It therefore seems likely that a conformationally malleable DNA-binding N-terminal extension is a common feature of TFRs.

6.9 DNA bending induced by SimR binding

DNA helical parameters were analysed using the Curves+ programme (Lavery *et al.*, 2009). The overall conformation of the 17-bp duplex is B-DNA, with an average helical twist of 33.7^{o} (compared to a helical twist value of 36.0^{o} for an idealised B-form DNA). It should be noted that individual steps might show significant deviations from the average value. The global bending of DNA is ~15^{o} (**Fig. 6.19A**). Since bending is most affected by the base step roll and twist angles (Dickerson, 1998), we plotted the roll and twist angles against the base steps to pinpoint the source of bending (**Fig. 6.19B**). There are two significant positive rolls (10 to 10.7^{o}) centred around base steps 6-7-8 in the operator half-site and symmetrically around steps 9-10-11 of the opposite half-site (**Fig. 6.19B**). The increase in roll angle coincides with the decrease in twist angle (26.7^{o} to 26.9^{o}) (**Fig. 6.19B**). The average global roll and twist angles are 2.9^{o} and 33.4^{o}, respectively. Thus local kinks around those base steps produce a global bend in the DNA, rather than a smooth bending. Moreover, there is a significant increase in the width of the minor groove from base step 6 through to base step 12, while the major groove width is just below the value for an idealised B-form DNA **(Fig. 6.19C**). Since the average distance between the two recognition helices in the SimR-DNA complex is 36.8 Å (assessed as the distance between the Cα atom of Tyr65 in each subunit (Yu *et al.*, 2010), greater than the distance between two consecutive major grooves in idealised B-DNA (34 Å), it is likely that the bending and the unwinding of the central DNA steps might be necessary for optimal positioning of the HTH motifs in adjacent major grooves. Lastly, although the sequence of the 17-bp duplex used in this study is a perfect inverted repeat with the exception of the central GC base-pair, the groove width and roll parameters are not symmetrical across this central base-pair. This reflects the

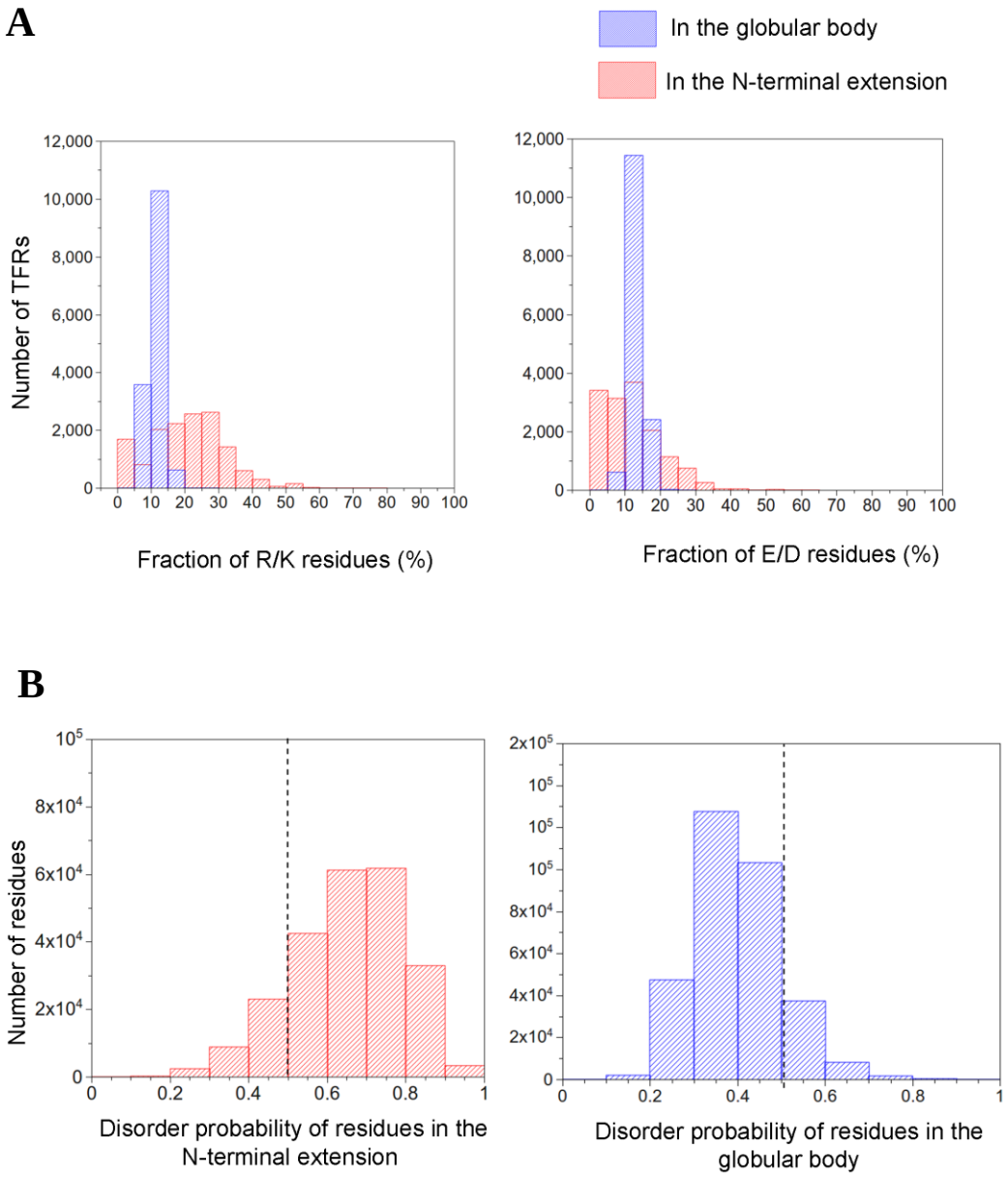


Fig. 6.18. Bioinformatic analysis of (A) the fraction of R+K or D+E residues in the N-terminal extension or in the globular body of 12,715 non-redundant TFRs (B) the disorder probability of residues in either the N-terminal extension or in the globular body. Residues with probability greater than 0.5 (dashed line) are predicted to be disordered.

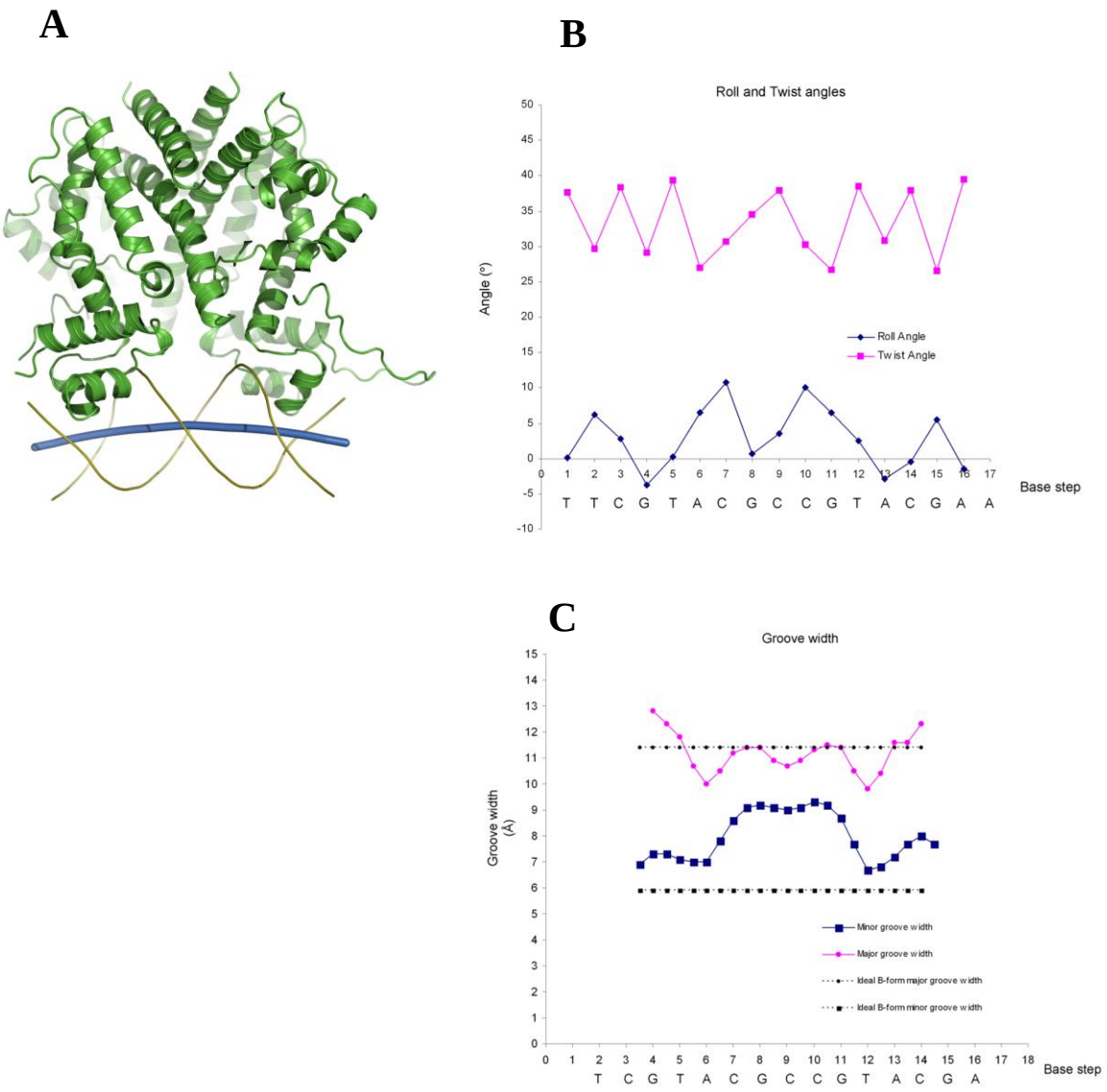


Fig. 6.19. DNA parameters as calculated by the Curves+ programme (Lavery *et al.*, 2009). (A) The bound DNA in SimR-DNA complex is globally bent by ~15o. The curved blue rod illustrates the helical axis of the bound DNA. The phosphate backbone of the DNA is shown in yellow and the SimR dimer is shown as green cartoon. (B) The roll and twist angle for each base pair step of the bound DNA. (C) The major and minor groove width of the bound DNA. The corresponding values for an ideal B-form DNA are also included for comparison.

non-equivalent end-to-end interactions between neighbouring DNA duplexes described above (**Fig. 6.15**).

6.10 Comparison of the SimR-DNA and SimR-simocyclinone complexes suggests the mechanism of derepression

In chapter 5, I speculated about the mechanism of simocyclinone-mediated derepression, based on a comparison of the structures of SimR-apo and the SimR-SD8 complex (Le *et al.*, 2011b and Chapter 5). However, it was apparent that SimR-apo had not crystallised in its DNA-binding form, since the distance between the recognition helices in SimR-apo was 42.3 Å (measured from the separation of the Cα atoms of the central Tyr65 residues), a spacing incompatible with binding to two consecutive major grooves (Le *et al.*, 2011b and Chapter 5). Moreover, this spacing was comparable to the corresponding value of 41.7 Å obtained for the SimR-SD8 complex. Indeed, TFR apo-proteins in general do not crystallize in their DNA-binding form (Yu *et al.*, 2010). The helix separation obtained for SimR-DNA was significantly shorter at 36.8 Å (averaged over the two complexes in the ASU), this value lying within the range of 34.7 to 38.8 Å observed in other TFR-DNA complexes (Miller *et al.*, 2010; Yu *et al.*, 2010). The major structural differences between the repressed, DNA-bound conformation of SimR and the derepressed, SD8-bound conformation, result from a 16° rotation of the subunits relative to one another roughly about the centre of the dimer interface (**Fig. 6.20** and **Fig. 6.21**). This re-defines many of the inter-subunit contacts, although the interface areas remain similar at 2795 $Å^2$ and 2640 $Å^2$ on average for SimR-SD8 and SimR-DNA, respectively [as calculated by the Protein Interactions, Surfaces and Assemblies server (PISA, http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html) (Krissinel and Henrick, 2005)]. However, five reciprocated inter-subunit hydrogen bonds (i.e. ten in total) are preserved between the two conformational states. These link the C-terminal end of α8 and the α9-α10 wrapping arm to the LBD of the adjacent subunit. As a consequence, when the

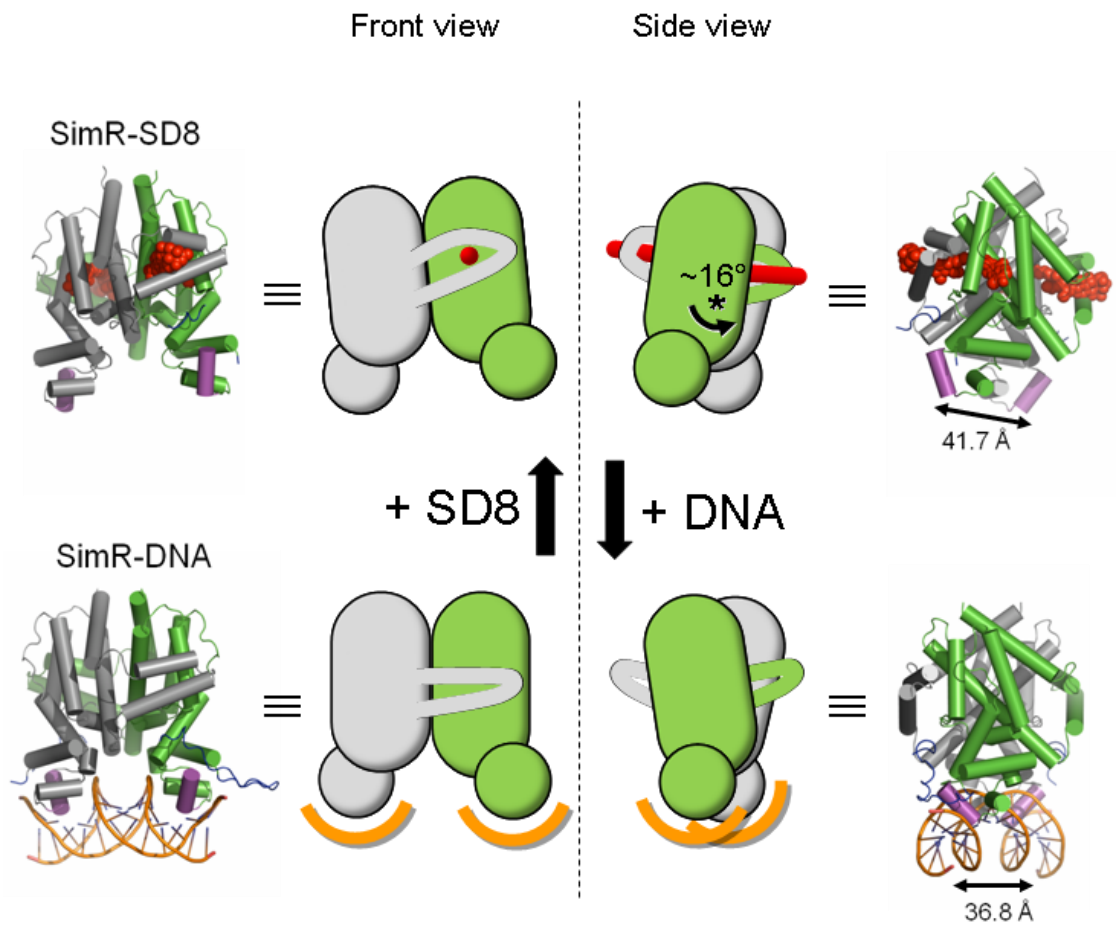


Fig. 6.20. Structures of SimR-simocyclinone and SimR-DNA together with schematic representations illustrating the rigid-body rotation of the subunits relative to one another. In order to emphasize the subunit rotation, the grey coloured subunits are shown fixed in the same relative orientations. This can be clearly seen in the side view where the green subunit rotates by ~16° relative to the grey subunit; the approximate pivot point is indicated by the asterisk (See also **Fig. 6.21** and **Fig. 6.22**). The distances separating the recognition helices α3 and α3' in the two structures are indicated.

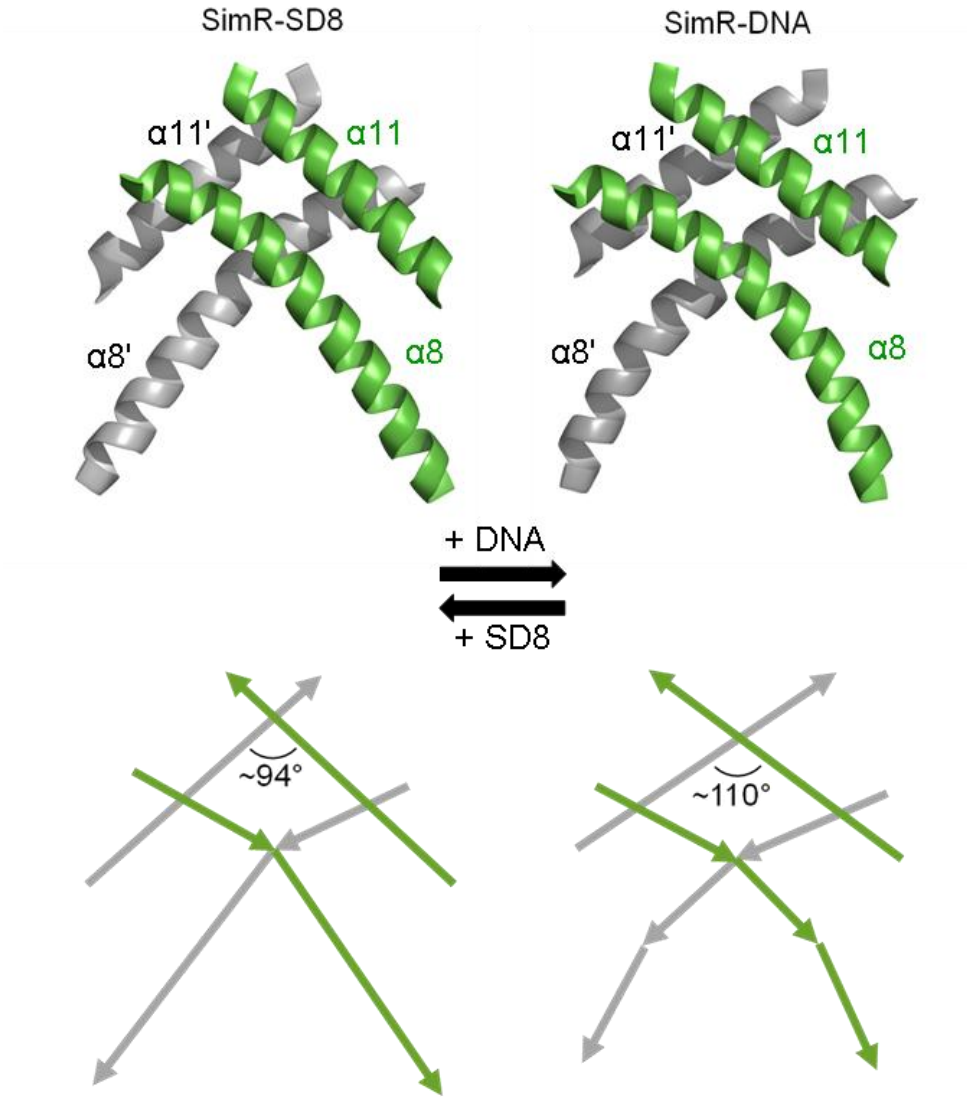


Fig. 6.21. Comparison of the crossing angles of helices α8 and α11 with α8' and α11' at the dimer interface in the SimR-SD8 (left panels) and SimR-DNA complexes (right panels). The view corresponds to the side view in **Fig. 6.20.** Overall, one subunit rotates relative to the other by ~16° which correlates with the change in the crossing angle of helices α11 and α11'. In both structures, helices α8 and α8' are slightly bent roughly where they cross; there is a further bend in each helix towards its C-terminus in the DNA-bound form. This bending is coupled to the movement of the α9-α10 wrapping arm (see also **Fig. 6.20** and **Fig. 6.22**).

subunits rotate, the α9-α10 wrapping arm moves with the adjacent subunit and the C-terminal end of α8 bends (**Fig. 6.21)**. Pairwise superpositions of individual subunits taken from the SimR-SD8 and SimR-DNA structures based on the subunit cores (i.e., inclusive of residues 29-168 plus 222-247 and exclusive of the TFR arm, the C-terminal end of α8 and the α9-α10 wrapping arm) gave RMSD values in the range 0.85 to 0.96 Å, indicating that the cores move essentially as rigid bodies at the protein backbone level and, importantly, there is no significant re-orientation of the DBD with respect to the LBD, in contrast to the "pendulum-like" motion seen in TetR (**Fig. 6.22**) (Orth *et al.*, 2000; Ramos *et al.*, 2005).

Nevertheless, the crystal structures do not convey the dynamic behaviour of the system and, as has been illustrated for other TFRs (Reichheld *et al.*, 2009; Yu *et al.*, 2010), in the absence of ligands or DNA, the protein is generally highly flexible and capable of sampling a variety of conformations, presumably including those akin to both the ligand-bound and DNA-bound states. The binding of SD8, a relatively hydrophobic molecule, contributes to the hydrophobic core of the SimR dimer; this will have a stabilising effect on the overall structure, locking it into a relatively rigid, low-energy state. Moreover, the combination of the threading of the ligand through both subunits and the projection of the side chain of Arg122 into the opposing subunit contribute to the rigidification of the system (Le *et al.*, 2011b and Chapter 5). The flexibility of the apo form is important to enable the TFR arms and the recognition helices to engage optimally with the DNA. The resulting favourable protein-DNA interactions will have a stabilising effect on this conformation of SimR. Moreover, in the DNA binding conformation the repositioning of the C-terminal end of helix α8 appropriately places it to make salt bridges to the DBD of the same subunit and to that of the opposing subunit, especially between Arg179 and Glu46, and between Arg180 and Glu72, respectively. These interactions, not present in the SD8-bound form, will further stabilise the DNA-bound conformation of SimR.

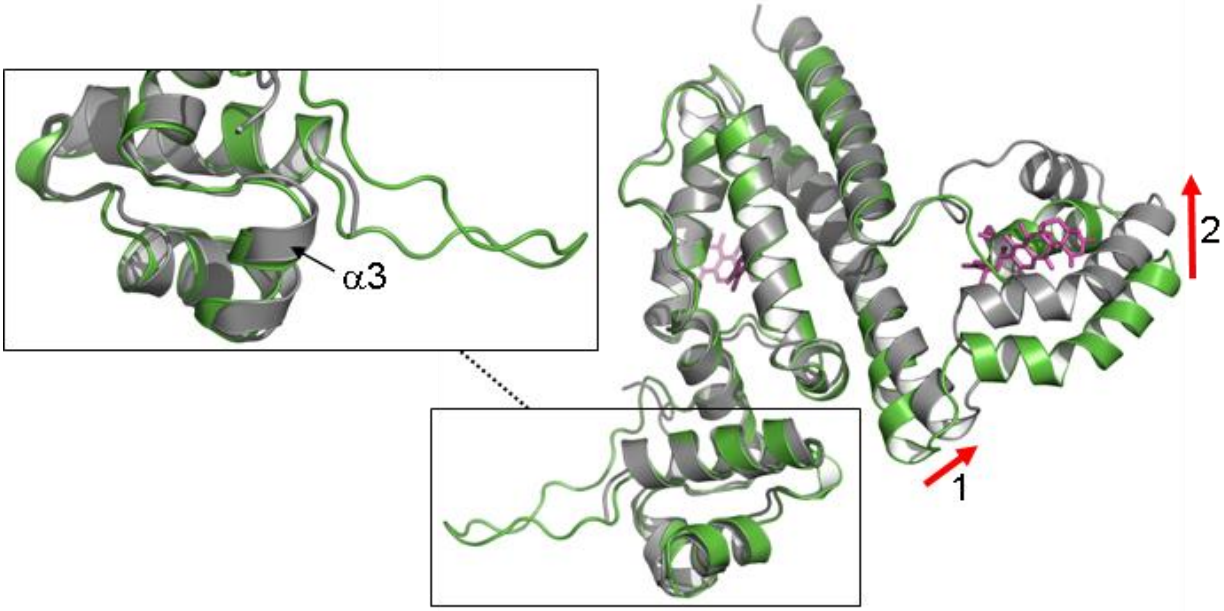


Fig. 6.22. Superposition of single subunits from the SimR-SD8 complex (grey) and the SimR-DNA complex (green) based on residues 29-168 plus 222-247 inclusive. The red arrows denote the direction of the conformational shifts upon transit from the DNA-bound to the SD8-bound form i.e. the bending of the C-terminal end of helix α8 (1) and the upward movement of the wrapping arm comprised of helices α9 and α10 (2). The inset shows the reverse view of the DBDs which illustrates their correspondence as a result of the superposition.

Chapter 7
General Discussion

7.1 Attempts to test the 'feed-forward' hypothesis

The simocyclinone biosynthetic intermediate SC4 (**Fig. 4.1**) lacks the aminocoumarin moiety and is essentially inactive as a DNA gyrase inhibitor (Edwards *et al*., 2009). However, we have shown that SC4 induces *simX* expression *in vivo* when added exogenously and efficiently relieves DNA binding by SimR *in vitro* (Chapter 4; Le *et al*., 2009). Similar observations have previously been made by Nodwell and colleagues working on the control of efflux of actinorhodin, the blue-pigmented polyketide antibiotic made by *S. coelicolor* (Tahlan *et al*., 2007 and Hopwood, 2007). Expression of the actinorhodin efflux pump is regulated by the repressor ActR. The Nodwell laboratory showed that, in addition to the mature six-ring antibiotic, three-ring intermediates from the biosynthetic pathway also relieve repression of the actinorhodin efflux pump by ActR (Tahlan *et al*., 2007). These observations in the actinorhodin and simocyclinone systems raise the exciting possibility of a feed-forward mechanism, in which bio-inactive intermediates ensure expression of the efflux pump prior to the build-up of a toxic concentration of the potentially lethal mature antibiotic (Tahlan *et al*., 2007; Hopwood 2007; Le *et al*., 2009).

As applied to simocyclinone (or actinorhodin) export, the feed-forward hypothesis is speculative. For a feed-forward mechanism to operate in the producing organism, SC4 or other SimR-binding pathway intermediates would have to accumulate in the cytoplasm of *S. antibioticus* to a concentration high enough to trigger *simX* expression, and it is possible that pathway intermediates are channelled between biosynthetic enzymes such that their cytoplasmic concentration is zero and SimR is never exposed to them. However, the crystal structures of SimR, alone and in complex with SD8 or SC4 (Chapter 5; Le *et al*., 2010) suggest a way to test the feed-forward hypothesis directly. Our structure of the repressor-ligand complex shows that the polyketide end of SD8 lies at the mouth of the drug-binding cavity whereas the aminocoumarin end lies at the far end of the cavity buried deep within the protein (**Fig. 5.4A**). It might therefore be possible to mutagenize SimR to occlude the aminocoumarin-binding pocket, creating SimR* proteins that can bind the intermediate SC4 (which lacks the aminocoumarin

ring) but cannot accept the mature antibiotic. These intermediate-specific SimR* proteins could then be used as biosensors *in vivo* to determine if SimR experiences pathway intermediates in the producing organism and hence to test the feed-forward hypothesis directly, and to answer a key question in antibiotic research - are biosynthetic intermediates released into the cytoplasm during antibiotic production?

From the crystal structures of the SimR-SD8 and SimR-SC4 complexes, I identified the amino acids that delineate the aminocoumarin end of the simocyclinone binding pocket (**Fig. 5.9**). In an attempt to create SimR* proteins, I generated and purified the 19 mutant SimR proteins described in **Table 7.1** and tested their ability to bind DNA, and to respond to SD8 and to the intermediate SC4. These amino acid substitutions caused a wide range of changes in the behaviour of SimR (**Table 7.1**). However, none of the mutant proteins showed the desired properties of a SimR* protein, i.e. the ability to bind cognate DNA with the same affinity as wild-type SimR, and the ability to respond to the intermediate SC4 but not to the mature antibiotic SD8. Therefore, I have not yet been able to test the "feed-forward" hypothesis in *S. antibioticus*.

If suitable SimR* proteins are identified in the future, the next step would be to confirm the behavior of these mutant proteins *in vivo*. I would introduce the mutant *simR* alleles (on the integrative plasmid pMS82) into *S. lividans* carrying either a *simRp-lux* or *simXp-lux* reporter fusion promoter and then expose these strains to exogenously added intermediate C4 and to simocyclinone, as described for the wild-type *simR* allele in Chapter 4 and Le *et al*. (2009). Assuming the *simR** alleles behaved appropriately in *S. lividans*, they could then be used to test the feed-forward hypothesis directly and to determine if the pathway intermediates exist freely in the cytoplasm of the producer, *S. antibioticus*, using the SimR* proteins as biosensors. To achieve this, PCR-targeting would be used to exactly replace the wild-type *simR* allele in *S. antibioticus* with the *simR** allele that responds to SC4 but not SD8. The existing *simRp-lux* or *simXp-lux* reporter fusions would be introduced into these strains and the activity of the reporters would be monitored through a time-course during growth in production medium. In parallel, wild-type *S. antibioticus* containing the *lux* reporter fusions (and the wild-type

Mutations	DNA binding affinity	Derepression by SD8	Derepression by SC4
K107A	WT	++	++
I83W	WT	++	+
T86W	– –	ND	ND
L88W	WT	++	+
H219A	WT	++	++
V139W	WT	+	+
Q135A/K/R	WT	WT	WT
Q136A	WT	WT	WT
Q136W	–	ND	ND
L212A	– –	ND	ND
M189A	– –	ND	ND
A103F	– –	ND	ND
W128A	– – –	ND	ND
L140A	– –	ND	ND
L154A	– –	ND	ND
Y160W	WT	ND	ND
Y160A	unable to purify	ND	ND
K107R Q135E E165Q	unable to purify	ND	ND
K107R Q135D E165Q	unable to purify	ND	ND

+ Greater sensitivity to SD8 or SC4 than wild-type SimR
WT Wild-type behaviour
– Decreased DNA-binding affinity relative to wild-type SimR
ND Not determined

Table 7.1. Mutations of amino acids that delinate the simocyclinone binding pocket of SimR and their effect on DNA binding affinity, and on derepression by SD8 and SC4.

simR allele) would be monitored as a control. If the *lux* reporters are induced (derepressed) only in the strain carrying wild-type SimR, it would show that only mature simocyclinone accumulates in the cytoplasm to a concentration sufficient to trigger expression of the efflux pump. In contrast, if the reporters are also induced in the presence of the intermediate-specific SimR proteins, it would show that C4 (or earlier intermediates) are released into the cytoplasm and that a feed-forward mechanism does occur in the producing organism.

7.2 Is *simX* the simocyclinone resistance determinant in *S. antibioticus*?

A remaining unresolved question is whether the efflux pump SimX is the resistance determinant to SD8 in the producing organism, *S. antibioticus*. In the streptomycetes that produce the aminocoumarins novobiocin (*Streptomyces sphaeroides*), clorobiocin (*Streptomyces roseochromogenes*), and coumermycin A_1 (*Streptomyces rishiriensis*), expression of an aminocoumarin-resistant $GyrB^R$ gyrase subunit encoded within the biosynthetic gene cluster is turned on during antibiotic production, hence conferring resistance (Thiara and Cundliffe, 1988, 1989, 1993; Schmutz *et al.*, 2003). The absence of an equivalent *gyrB* resistance gene in the simocyclinone biosynthetic gene cluster was initially a surprise when the *sim* locus was sequenced, until it was subsequently and unexpectedly shown that simocyclinone has a completely different mode of action, binding instead to the GyrA subunit of the enzyme (Flatman *et al*., 2005; Edwards *et al*., 2009). However, no *gyrA* resistance gene is present within the *sim* cluster either (Galm *et al*., 2002; Trefzer *et al*., 2002). Expressing *simX* from the strong, constitutive promoter, *ermEp**, conferred simocyclinone resistance on the heterologous host *S. lividans* (MIC 65 μg/ml instead of 2 μg/ml for *S. lividans* alone), showing that SimX can act as an effective resistance determinant when expressed at appropriate levels. However, the MIC was only 4 μg/ml when *simX* was expressed from its own promoter in the absence of SimR. This could reflect additional levels of *simX* regulation that have yet to be uncovered. For example, in addition to its repression by SimR, it is possible that *simX* expression in *S. antibioticus* might also be activated by a simocyclinone pathway-specific activator that is absent from *S. lividans* (e.g. SimR2, see below).

Based on precedent, resistance genes seem invariably to be associated with the biosynthetic gene cluster in antibiotic-producing streptomycetes. However, one other possibility I considered was that the DNA gyrase of *S. antibioticus* might be innately insensitive to SD8, such that the antibiotic has no target in the producing organism. To begin to address this possibility, I have isolated the *S. antibioticus gyr* genes. The genome of *S. antibioticus* has not been sequenced, and so I employed ligation-mediated PCR to clone the native *gyrA* and *gyrB* genes from *S. antibioticus*, and then determine their sequence. The cloning of these genes will allow me to use genetic and biochemical methods to determine whether *S. antibioticus* DNA gyrase is sensitive to SD8.

7.3 Other transcriptional regulators encoded in the simocyclinone biosynthetic cluster, SimR2 and SimR3

Encoded within the simocyclinone biosynthetic cluster are two additional transcriptional regulators, SimR2 and SimR3 (**Fig. 1.8**). Investigating the roles of SimR2 and SimR3 would reveal whether they also regulate *simX*, and thus have a potential role in SD8 resistance. Moreover, such research can shed light on the interplay between the three transcription regulators within the *sim* cluster in integrating and coordinating the biosynthesis of a complex antibiotic.

simR2 encodes a so-called atypical response regulator (ARR). In bacteria, some response regulators lack the conserved residues important for the phosphorylation that causes typical response regulators to switch their output response, suggesting their activity may be controlled by alternative regulatory mechanisms (Aínsa *et al*., 1999; Guthrie *et al*., 1998; Wang *et al*., 2009). ARRs are found in a wide range of bacteria and are involved in the regulation of bacterial growth and development, secondary metabolite biosynthesis, iron transport, cell movement, and virulence (Raghavan and Groisman, 2010). Several ARRs have been shown to play key roles in *Streptomyces* biology, for example WhiI is required for differentiation in *S. coelicolor* (Aínsa *et al*., 1999), RedZ is required for production of the antibiotic undecylprodigionine in *S. coelicolor* (Guthrie *et*

al., 1998) and JadR1 is required for production of the antibiotic jadomycin in *S. venezuelae* (Wang *et al*., 2009). All three of these proteins have degenerate phosphorylation pockets and RedZ actually carries a valine in place of the aspartate residue that is normally the site of phosphorylation (Guthrie *et al*., 1998). Finally, all these ARRs are 'orphan' RRs – they are not genetically linked to a sensor kinase gene, as is most usually the case with typical response regulators. These data suggest that ARRs are likely to respond to some other input, but until recently the nature of that input had not been demonstrated. Wang *et al*. (2009) recently showed that JadR1 is a transcriptional activator of certain promoters in the jadomycin biosynthetic cluster and that the binding of jadomycin to the N-terminal domain (the domain normally subject to phosphorylation in conventional RRs) causes JadR1 to dissociate from its target promoters. Similarly, they showed that the ARR RedZ directly activates transcription of at least one promoter (*redDp*) in the undecylprodigionine biosynthetic gene cluster and that binding of undecylprodigionine to RedZ causes it to dissociate from its target (Wang *et al*., 2009). These results provide a mechanism for autoregulation of antibiotic biosynthesis by binding of the end product to an ARR. Strikingly, the ARR SimR2 is the closest homologue of JadR1 in the databases (42% sequence identity), raising the likelihood that its activity will be controlled by simocyclinone and/or its biosynthetic intermediates.

The third transcription factor encoded within the *sim* cluster, SimR3, is a member of the MarR family of ligand-responsive transcriptional regulators (Wilkinson and Grove, 2006). Most members of the MarR family function as repressors that lose their affinity for their DNA binding sites in response to their cognate ligand (Wilkinson and Grove, 2006; Perera *et al*., 2009), but at least some members of the family mediate transcriptional activation in response to ligand (Zhao *et al.*, 2008). The presence of a gene encoding a MarR family member within the *sim* cluster suggests that it will control expression of one or more promoters in the cluster and that its activity will in turn be regulated by a small molecule effector ligand. Most MarR proteins bind negatively charged phenolic compounds (Wilkinson and Grove, 2006). If this were true of SimR3, then a likely candidate for its cognate effector ligand would be one or more of the

anionic phenolic intermediates in the pathway that generates the aminocoumarin moiety of simocyclinone.

I have constructed null mutations in each of *simR*, *simR2* and *simR3* in *S. antibioticus.* We also have a mini-array, suitable for both transcriptional profiling and ChIP-on-chip, covering the *sim* biosynthetic cluster and sequenced flanking DNA (~84 kb in total). We aim to use this mini-array to examine expression of each gene in the *sim* biosynthetic cluster in the wild type and in the constructed *simR2* and *simR3* null mutants in a time course during growth in production medium. In addition, we also have polyclonal antibodies against both regulators and aim to use these antisera to perform time-resolved ChIP-chip experiments on *S. antibioticus* grown in production medium, taking samples before, during and after antibiotic production phase, to define which promoters within the cluster are bound *in vivo* by each regulator and when. These experiments will reveal whether *simX* is also regulated by SimR2 or SimR3.

An additional experiment to address whether *simX* functions as the simocyclinone resistance determinant would be to introduce *simR2* (and/or *simR3*) into the *S. lividans* strain already carrying *simX-simR*. If *simX* is subject to activation by SimR2 (ARR family), as well as repression by SimR, then such a strain should exhibit high-level resistance to SD8.

I have already overexpressed SimR2 and SimR3 in *E. coli* in soluble forms. Following on from the ChIP-chip and transcriptional profiling experiments, we will use the purified proteins for *in vitro* studies. We will PCR amplify target promoters identified in the *sim* cluster, radiolabel them, and use EMSA and DNaseI footprinting to confirm which promoters within the *sim* cluster are direct regulatory targets for SimR2 and SimR3. Once these *in vitro* assays are established, EMSA experiments will be performed in the presence of simocyclinone and various simocyclinone intermediates to determine if the DNA-binding activities of SimR2 (ARR family) and SimR3 (ligand-responsive MarR family) are controlled by the mature antibiotic and/or pathway intermediates. Most MarR proteins bind anionic phenolic compounds (Wilkinson and Grove, 2006). As a

consequence, we will initially test anionic phenolic intermediates from the aminocoumarin pathway as effector ligands for SimR3, before testing intermediates from other parts of the simocyclinone pathway.

Such research has the potential to examine how metabolite signalling within the simocyclinone pathway coordinates and integrates expression of the ~50 proteins involved in the biosynthesis of the mature antibiotic. Since there is now strong evidence that these metabolite signalling mechanisms are likely to be widespread within streptomycete antibiotic biosynthetic pathways, these results could even prove relevant to knowledge-based improvements in the yield of commercially important antibiotics.

7.4 Engineering the simocyclinone biosynthetic cluster to create novel antibiotics

Although the SD8-producing organism, *S. antibioticus*, can be manipulated genetically (Galm *et al*., 2002; Trefzer *et al*., 2002), it is very unattractive as a genetic system. As a consequence, although the ~84-kb gene cluster responsible for the biosynthesis of simocyclinone has been sequenced, it has not been realistic to engineer the cluster in the producing organism. Therefore, in order to efficiently investigate the roles of the biosynthetic genes within the *sim* cluster, we need to heterologously express the *sim* cluster in an experimentally convenient model *Streptomyces* species, such as *S. lividans*, *S. coelicolor* or *Streptomyces venezuelae*.

The *sim* cluster was originally isolated on a series of overlapping cosmid clones (Galm *et al*., 2002; Trefzer *et al*., 2002), but it would be technically extremely difficult to reconstruct the entire *sim* cluster as a single contiguous clone from these original cosmids. To circumvent this problem, we commissioned BioS+T (Montreal, Canada) to construct a library of *S. antibioticus* DNA in the *E. coli-Streptomyces* shuttle PAC vector, pESAC-13 (a derivative of pPAC-S1; Sosio *et al*., 2000), which accepts insets of up to ~150 kb. From this library we have identified 8 clones that carry the entire *sim* cluster. These PAC constructs will be introduced into e.g. *S. lividans* by conjugation and the resulting strains will be assayed for their ability to make simocyclinones. If the *sim*

cluster is expressed in a heterologous host, it will mean that we can now genetically manipulate the cluster at will in *E. coli* (using, for example, Lambda-Red PCR-targeting; see Chapter 2 for more detail on the method) and then conjugate it into e.g. *S. lividans* to see what products are made.

It is interesting to uncover novel chemistry and to be able to rationally engineer existing antibiotics to improve their potency or to overcome resistance. The biosynthesis of simocyclinone implies the existence of novel chemistry. For example, the modular polyketide synthase (PKS) responsible for assembly of the tetraene dicarboxylic linker for the aminocoumarin and angucycline "warheads" that bind synergistically to gyrase is unusual in that several domains expected to be required for tetraene assembly appear to be absent. Specifically, dehydratase (DH) domains are missing from modules 2 and 4 and a ketroreductase (KR) domain is missing from module 2 of the PKS (**Fig. 1.9**). It has been proposed by Trefzer *et al*. (2002) that these missing catalytic activities are supplied to the PKS *in trans* by the SimC6 and SimC7 proteins (**Fig. 1.9**). The role of these proteins in tetraene assembly could be examined by creating in-frame deletions in the genes encoding them (*simC6* and *simC7*) and investigating their effect on SD8 production. Depending on the specificity of downstream enzymes in the biosynthetic pathway, the mutants may produce analogues of SD8 with an altered dicarboxylic acid linker or simocyclinone production may be abolished and the isolated aminocoumarin and angucycline moieties may accumulate instead.

The crystal structure of DNA gyrase-SD8 showed the antibiotic crosslinking the AC and PK pockets of separate GyrA subunits (Edwards *et al*., 2009; **Fig. 1.10**). However, modelling and analytical ultracentrifugation data presented by Edwards *et al*. (2009) suggest that a single simocyclinone molecule in a ‘strained’ or ‘bent’ conformation can bridge between the AC and PK pockets within the same GyrA subunit. This ‘bent’ conformation might occur more readily if the conjugated system of double bonds in the tetraene linker was disrupted, thereby making simocyclinone more flexible and, potentially at least, a more potent DNA gyrase inhibitor. Therefore, in addition to the *simC6* and *simC7* mutagenesis experiments described above, it would also be interesting

to prevent tetraene (2,4,6,8-Decatetraenoyl) chain assembly altogether by deleting the *simC1ABC* Type I polyketide synthase genes, and then trying to make novel derivatives of simocyclinone by 'mutasynthesis.' In this approach, chemically synthesised tetraene substrate analogues, e.g. 2,4,6,8-Decatetraenoyl and 2,4,6-Octatrienoyl N-acetylcysteamine thioesters, would be fed to the *simC1ABC* mutant. Again, depending on the specificity of downstream enzymes in the biosynthetic pathway, this mutasynthetic approach may produce analogues of SD8 with an altered dicarboxylic acid linker.

References

1. Adams, P.D., Grosse-Kunstleve, R.W., Hung, L.W., Ioerger, T.R., McCoy, A.J., Moriarty, N.W., Read, R.J., Sacchettini, J.C., Sauter, N.K. and Terwilliger, T.C. (2002) PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr*, **58**, 1948-1954.

2. Ahn, S.K., Tahlan, K., Yu, Z. and Nodwell, J. (2007) Investigation of transcription repression and small-molecule responsiveness by TetR-like transcription factors using a heterologous Escherichia coli-based assay. *J Bacteriol*, **189**, 6655-6664.

3. Aigle, B., Wietzorrek, A., Takano, E. and Bibb, M.J. (2000) A single amino acid substitution in region 1.2 of the principal sigma factor of Streptomyces coelicolor A3(2) results in pleiotropic loss of antibiotic production. *Mol Microbiol*, **37**, 995-1004.

4. Ainsa, J.A., Parry, H.D. and Chater, K.F. (1999) A response regulator-like protein that functions at an intermediate stage of sporulation in Streptomyces coelicolor A3(2). *Mol Microbiol*, **34**, 607-619.

5. Alguel, Y., Meng, C., Teran, W., Krell, T., Ramos, J.L., Gallegos, M.T. and Zhang, X. (2007) Crystal structures of multidrug binding protein TtgR in complex with antibiotics and plant antimicrobials. *J Mol Biol*, **369**, 829-840.

6. Ball, P.R., Shales, S.W. and Chopra, I. (1980) Plasmid-mediated tetracycline resistance in Escherichia coli involves increased efflux of the antibiotic. *Biochem Biophys Res Commun*, **93**, 74-81.

7. Bates, A.D. and Maxwell, A. (2005) DNA Topology. *Oxford University Press*.

8. Baulard, A.R., Betts, J.C., Engohang-Ndong, J., Quan, S., McAdam, R.A., Brennan, P.J., Locht, C. and Besra, G.S. (2000) Activation of the pro-drug ethionamide is regulated in mycobacteria. *J Biol Chem*, **275**, 28326-28331.

9. Beamer, L.J. and Pabo, C.O. (1992) Refined 1.8 A crystal structure of the lambda repressor-operator complex. *J Mol Biol*, **227**, 177-196.

10. Becskei, A. and Serrano, L. (2000) Engineering stability in gene networks by autoregulation. *Nature*, **405**, 590-593.

11. Berdy, J. (2005) Bioactive microbial metabolites. *J Antibiot (Tokyo)*, **58**, 1-26.

12. Berger, J.M. (1998) Structure of DNA topoisomerases. *Biochim Biophys Acta*, **1400**, 3-18.

13. Berger, J.M. and Batcho, A.D. (1978) Coumarin-glycoside antibiotics. *J. Chromatogr. Libr.*, **15**, 101-158.

14. Bertrand, K.P., Postle, K., Wray, L.V., Jr. and Reznikoff, W.S. (1983) Overlapping divergent promoters control expression of Tn10 tetracycline resistance. *Gene*, **23**, 149-156.

15. Bierman, M., Logan, R., O'Brien, K., Seno, E.T., Rao, R.N. and Schoner, B.E. (1992) Plasmid cloning vectors for the conjugal transfer of DNA from Escherichia coli to Streptomyces spp. *Gene*, **116**, 43-49.

16. Brino, L., Urzhumtsev, A., Mousli, M., Bronner, C., Mitschler, A., Oudet, P. and Moras, D. (2000) Dimerization of Escherichia coli DNA-gyrase B provides a structural mechanism for activating the ATPase catalytic center. *J Biol Chem*, **275**, 9468-9475.

17. Brown, M.H. and Skurray, R.A. (2001) Staphylococcal multidrug efflux protein QacA. *J Mol Microbiol Biotechnol*, **3**, 163-170.

18. Buttner, M.J., Fearnley, I.M. and Bibb, M.J. (1987) The agarase gene (dagA) of Streptomyces coelicolor A3(2): nucleotide sequence and transcriptional analysis. *Mol Gen Genet*, **209**, 101-109.

19. Buttner, M.J., Hong, H.J. and Hutchings, M.I. (2008) Vancomycin resistance VanS/VanR two-component systems. *Bacterial Signal Transduction: Networks and Drug Targets*, **631**, 200-213.

20. Buttner, M.J., Smith, A.M. and Bibb, M.J. (1988) At least three different RNA polymerase holoenzymes direct transcription of the agarase gene (dagA) of Streptomyces coelicolor A3(2). *Cell*, **52**, 599-607.

21. Bystrykh, L.V., Fernandez-Moreno, M.A., Herrema, J.K., Malpartida, F., Hopwood, D.A. and Dijkhuizen, L. (1996) Production of actinorhodin-related "blue pigments" by Streptomyces coelicolor A3(2). *J Bacteriol*, **178**, 2238-2244.

22. Caballero, J.L., Malpartida, F. and Hopwood, D.A. (1991) Transcriptional organization and regulation of an antibiotic export complex in the producing Streptomyces culture. *Mol Gen Genet*, **228**, 372-380.

23. Caffrey, P., Lynch, S., Flood, E., Finnan, S. and Oliynyk, M. (2001) Amphotericin biosynthesis in Streptomyces nodosus: deductions from analysis of polyketide synthase and late genes. *Chem Biol*, **8**, 713-723.

24. Cases, I., de Lorenzo, V. and Ouzounis, C.A. (2003) Transcription regulation and environmental adaptation in bacteria. *Trends Microbiol*, **11**, 248-253.

25. Chatterji, M., Unniraman, S., Maxwell, A. and Nagaraja, V. (2000) The additional 165 amino acids in the B protein of Escherichia coli DNA gyrase have an

important role in DNA binding. *J Biol Chem*, **275**, 22888-22894.

26. Chen, V.B., Arendall, W.B., 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S. and Richardson, D.C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr*, **66**, 12-21.

27. Chopra, I. and Roberts, M. (2001) Tetracycline antibiotics: mode of action, applications, molecular biology, and epidemiology of bacterial resistance. *Microbiol Mol Biol Rev*, **65**, 232-260 ; second page, table of contents.

28. Cohen, S.P., Hachler, H. and Levy, S.B. (1993) Genetic and functional analysis of the multiple antibiotic resistance (mar) locus in Escherichia coli. *J Bacteriol*, **175**, 1484-1492.

29. Corbett, K.D. and Berger, J.M. (2003) Structure of the topoisomerase VI-B subunit: implications for type II topoisomerase mechanism and evolution. *Embo J*, **22**, 151-163.

30. Corbett, K.D. and Berger, J.M. (2004) Structure, molecular mechanisms, and evolutionary relationships in DNA topoisomerases. *Annu Rev Biophys Biomol Struct*, **33**, 95-118.

31. Cos, P., Tote, K., Horemans, T. and Maes, L. (2010) Biofilms: an extra hurdle for effective antimicrobial therapy. *Curr Pharm Des*, **16**, 2279-2295.

32. Cowtan, K. (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr D Biol Crystallogr*, **62**, 1002-1011.

33. Cowtan, K. (2010) Recent developments in classical density modification. *Acta Crystallogr D Biol Crystallogr*, **66**, 470-478.

34. Cundliffe, E. and Demain, A.L. (2010) Avoidance of suicide in antibiotic-producing microbes. *J Ind Microbiol Biotechnol*, **37**, 643-672.

35. Datsenko, K.A. and Wanner, B.L. (2000) One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc Natl Acad Sci U S A*, **97**, 6640-6645.

36. Davies, J. (1994) Inactivation of antibiotics and dissemination of resistance genes. *Science*, **264**, 375-381.

37. DeLano, W.L. (2002) The PyMol User's Manual. *DeLano Scientific, San Carlos, Ca, USA.*

38. Dickerson, R.E. (1998) DNA bending: the prevalence of kinkiness and the

virtues of normality. *Nucleic Acids Res*, **26**, 1906-1926.

39. Dong, K.C. and Berger, J.M. (2007) Structural basis for gate-DNA recognition and bending by type IIA topoisomerases. *Nature*, **450**, 1201-1205.

40. Doublie, S. (1997) Preparation of selenomethionyl proteins for phase determination. *Methods Enzymol*, **276**, 523-530.

41. Dover, L.G., Corsino, P.E., Daniels, I.R., Cocklin, S.L., Tatituri, V., Besra, G.S. and Futterer, K. (2004) Crystal structure of the TetR/CamR family repressor Mycobacterium tuberculosis EthR implicated in ethionamide resistance. *J Mol Biol*, **340**, 1095-1105.

42. Dyson, H.J. and Wright, P.E. (2005) Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol*, **6**, 197-208.

43. Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*, **32**, 1792-1797.

44. Edgar, R.C. (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, **26**, 2460-2461.

45. Edwards, M.J., Flatman, R.H., Mitchenall, L.A., Stevenson, C.E., Le, T.B., Clarke, T.A., McKay, A.R., Fiedler, H.P., Buttner, M.J., Lawson, D.M. *et al.* (2009) A crystal structure of the bifunctional antibiotic simocyclinone D8, bound to DNA gyrase. *Science*, **326**, 1415-1418.

46. Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr*, **60**, 2126-2132.

47. Engohang-Ndong, J., Baillat, D., Aumercier, M., Bellefontaine, F., Besra, G.S., Locht, C. and Baulard, A.R. (2004) EthR, a repressor of the TetR/CamR family implicated in ethionamide resistance in mycobacteria, octamerizes cooperatively on its operator. *Mol Microbiol*, **51**, 175-188.

48. Evans, P. (2006) Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr*, **62**, 72-82.

49. Fernandez-Moreno, M.A., Caballero, J.L., Hopwood, D.A. and Malpartida, F. (1991) The act cluster contains regulatory and antibiotic export genes, direct targets for translational control by the bldA tRNA gene of Streptomyces. *Cell*, **66**, 769-780.

50. Flatman, R.H., Howells, A.J., Heide, L., Fiedler, H.P. and Maxwell, A. (2005) Simocyclinone D8, an inhibitor of DNA gyrase with a novel mode of action. *Antimicrob Agents Chemother*, **49**, 1093-1100.

51. Frenois, F., Engohang-Ndong, J., Locht, C., Baulard, A.R. and Villeret, V. (2004) Structure of EthR in a ligand bound conformation reveals therapeutic perspectives against tuberculosis. *Mol Cell*, **16**, 301-307.

52. Fu, G., Wu, J., Liu, W., Zhu, D., Hu, Y., Deng, J., Zhang, X.E., Bi, L. and Wang, D.C. (2009) Crystal structure of DNA gyrase B' domain sheds lights on the mechanism for T-segment navigation. *Nucleic Acids Res*, **37**, 5908-5916.

53. Galm, U., Schimana, J., Fiedler, H.P., Schmidt, J., Li, S.M. and Heide, L. (2002) Cloning and analysis of the simocyclinone biosynthetic gene cluster of Streptomyces antibioticus Tu 6040. *Arch Microbiol*, **178**, 102-114.

54. Gellert, M., Mizuuchi, K., O'Dea, M.H. and Nash, H.A. (1976) DNA gyrase: an enzyme that introduces superhelical turns into DNA. *Proc Natl Acad Sci U S A*, **73**, 3872-3876.

55. George, A.M. and Levy, S.B. (1983) Amplifiable resistance to tetracycline, chloramphenicol, and other antibiotics in Escherichia coli: involvement of a non-plasmid-determined efflux of tetracycline. *J Bacteriol*, **155**, 531-540.

56. Gormley, N.A., Orphanides, G., Meyer, A., Cullis, P.M. and Maxwell, A. (1996) The interaction of coumarin antibiotics with fragments of DNA gyrase B protein. *Biochemistry*, **35**, 5083-5092.

57. Gregory, M.A., Till, R. and Smith, M.C. (2003) Integration site for Streptomyces phage phiBT1 and development of site-specific integrating vectors. *J Bacteriol*, **185**, 5320-5323.

58. Gregory, M.A., Till, R. and Smith, M.C. (2003) Integration site for Streptomyces phage phiBT1 and development of site-specific integrating vectors. *J Bacteriol*, **185**, 5320-5323.

59. Grkovic, S., Brown, M.H., Roberts, N.J., Paulsen, I.T. and Skurray, R.A. (1998) QacR is a repressor protein that regulates expression of the Staphylococcus aureus multidrug efflux pump QacA. *J Biol Chem*, **273**, 18665-18673.

60. Grkovic, S., Brown, M.H., Schumacher, M.A., Brennan, R.G. and Skurray, R.A. (2001) The staphylococcal QacR multidrug regulator binds a correctly spaced operator as a pair of dimers. *J Bacteriol*, **183**, 7102-7109.

61. Gust, B., Challis, G.L., Fowler, K., Kieser, T. and Chater, K.F. (2003) PCR-targeted Streptomyces gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. *Proc Natl Acad Sci U S A*, **100**, 1541-1546.

62. Guthrie, E.P., Flaxman, C.S., White, J., Hodgson, D.A., Bibb, M.J. and Chater,

K.F. (1998) A response-regulator-like activator of antibiotic synthesis from Streptomyces coelicolor A3(2) with an amino-terminal domain that lacks a phosphorylation pocket. *Microbiology*, **144 (Pt 3)**, 727-738.

63. Hanahan, D. (1983) Studies on transformation of Escherichia coli with plasmids. *J Mol Biol*, **166**, 557-580.

64. Hecht, B., Muller, G. and Hillen, W. (1993) Noninducible Tet repressor mutations map from the operator binding motif to the C terminus. *J Bacteriol*, **175**, 1206-1210.

65. Heide, L. (2009) Genetic engineering of antibiotic biosynthesis for the generation of new aminocoumarins. *Biotechnol Adv*, **27**, 1006-1014.

66. Heide, L. (2009) Aminocoumarins mutasynthesis, chemoenzymatic synthesis, and metabolic engineering. *Methods Enzymol*, **459**, 437-455.

67. Heide, L., Gust, B., Anderle, C. and Li, S.M. (2008) Combinatorial biosynthesis, metabolic engineering and mutasynthesis for the generation of new aminocoumarin antibiotics. *Curr Top Med Chem*, **8**, 667-679.

68. Helbl, V., Berens, C. and Hillen, W. (1995) Proximity probing of Tet repressor to tet operator by dimethylsulfate reveals protected and accessible functions for each recognized base-pair in the major groove. *J Mol Biol*, **245**, 538-548.

69. Henikoff, S., Haughn, G.W., Calvo, J.M. and Wallace, J.C. (1988) A large family of bacterial activator proteins. *Proc Natl Acad Sci U S A*, **85**, 6602-6606.

70. Hesketh, A., Chen, W.J., Ryding, J., Chang, S. and Bibb, M. (2007) The global role of ppGpp synthesis in morphological differentiation and antibiotic production in Streptomyces coelicolor A3(2). *Genome Biol*, **8**, R161.

71. Heuer, H., Szczepanowski, R., Schneiker, S., Puhler, A., Top, E.M. and Schluter, A. (2004) The complete sequences of plasmids pB2 and pB3 provide evidence for a recent ancestor of the IncP-1beta group without any accessory genes. *Microbiology*, **150**, 3591-3599.

72. Hillen, W. and Berens, C. (1994) Mechanisms underlying expression of Tn10 encoded tetracycline resistance. *Annu Rev Microbiol*, **48**, 345-369.

73. Hinrichs, W., Kisker, C., Duvel, M., Muller, A., Tovar, K., Hillen, W. and Saenger, W. (1994) Structure of the Tet repressor-tetracycline complex and regulation of antibiotic resistance. *Science*, **264**, 418-420.

74. Hoffmann, K.M., Williams, D., Shafer, W.M. and Brennan, R.G. (2005) Characterization of the multiple transferable resistance repressor, MtrR, from Neisseria

gonorrhoeae. *J Bacteriol*, **187**, 5008-5012.

75. Holm, L. and Sander, C. (1995) Dali: a network tool for protein structure comparison. *Trends Biochem Sci*, **20**, 478-480.

76. Holt, R.J. and Stewart, G.T. (1964) Production of Amidase and Beta-Lactamase by Bacteria. *J Gen Microbiol*, **36**, 203-213.

77. Holzenkampfer, M., Walker, M., Zeeck, A., Schimana, J. and Fiedler, H.P. (2002) Simocyclinones, novel cytostatic angucyclinone antibiotics produced by Streptomyces antibioticus Tu 6040 II. Structure elucidation and biosynthesis. *J Antibiot (Tokyo)*, **55**, 301-307.

78. Hong, H.J., Hutchings, M.I., Hill, L.M. and Buttner, M.J. (2005) The role of the novel Fem protein VanK in vancomycin resistance in Streptomyces coelicolor. *J Biol Chem*, **280**, 13055-13061.

79. Hopwood, D.A. (2007) How do antibiotic-producing bacteria ensure their self-resistance before antibiotic biosynthesis incapacitates them? *Mol Microbiol*, **63**, 937-940.

80. Itou, H., Okada, U., Suzuki, H., Yao, M., Wachi, M., Watanabe, N. and Tanaka, I. (2005) The CGL2612 protein from Corynebacterium glutamicum is a drug resistance-related transcriptional repressor: structural and functional analysis of a newly identified transcription factor from genomic DNA analysis. *J Biol Chem*, **280**, 38711-38719.

81. Itou, H., Watanabe, N., Yao, M., Shirakihara, Y. and Tanaka, I. (2010) Crystal structures of the multidrug binding repressor Corynebacteriumglutamicum CgmR in complex with inducers and with an operator. *J Mol Biol*, **403**, 174-184.

82. Jenks, P.J. and Edwards, D.I. (2002) Metronidazole resistance in Helicobacter pylori. *Int J Antimicrob Agents*, **19**, 1-7.

83. Jiang, H. and Hutchinson, C.R. (2006) Feedback regulation of doxorubicin biosynthesis in Streptomyces peucetius. *Res Microbiol*, **157**, 666-674.

84. Joshi, R., Passner, J.M., Rohs, R., Jain, R., Sosinsky, A., Crickmore, M.A., Jacob, V., Aggarwal, A.K., Honig, B. and Mann, R.S. (2007) Functional specificity of a Hox protein mediated by the recognition of minor groove structure. *Cell*, **131**, 530-543.

85. Kamionka, A., Bogdanska-Urbaniak, J., Scholz, O. and Hillen, W. (2004) Two mutations in the tetracycline repressor change the inducer anhydrotetracycline to a corepressor. *Nucleic Acids Res*, **32**, 842-847.

86. Kaneko, M., Yamaguchi, A. and Sawai, T. (1985) Energetics of tetracycline

efflux system encoded by Tn10 in Escherichia coli. *FEBS Lett*, **193**, 194-198.

87. Keatinge-Clay, A.T., Maltby, D.A., Medzihradszky, K.F., Khosla, C. and Stroud, R.M. (2004) An antibiotic factory caught in action. *Nat Struct Mol Biol*, **11**, 888-893.

88. Kieser, T., Bibb, M.J., Buttner, M.J., Chater, K.F. and Hopwood, D.A. (2000) Practical Streptomyces Genetics.

89. Kisker, C., Hinrichs, W., Tovar, K., Hillen, W. and Saenger, W. (1995) The complex formed between Tet repressor and tetracycline-Mg2+ reveals mechanism of antibiotic resistance. *J Mol Biol*, **247**, 260-280.

90. Krissinel, E. and Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr*, **60**, 2256-2268.

91. Krissinel, E. and Henrick, K. (2005) Detection of protein assemblies in crystals. In Computational Life Sciences (Berthold, M. R., Glen, R. C., Diederichs, K., Kohlbacher, O. & Fischer, I., eds.). *Springer-Verlag, Berlin Heidelberg*, **3695**, 163-174.

92. Lamour, V., Hoermann, L., Jeltsch, J.M., Oudet, P. and Moras, D. (2002) Crystallization of the 43 kDa ATPase domain of Thermus thermophilus gyrase B in complex with novobiocin. *Acta Crystallogr D Biol Crystallogr*, **58**, 1376-1378.

93. Laskowski, R.A. (2001) PDBsum: summaries and analyses of PDB structures. *Nucleic Acids Res*, **29**, 221-222.

94. Lavery, R., Moakher, M., Maddocks, J.H., Petkeviciute, D. and Zakrzewska, K. (2009) Conformational analysis of nucleic acids revisited: Curves+. *Nucleic Acids Res*, **37**, 5917-5929.

95. Le, T.B., Fiedler, H.P., den Hengst, C.D., Ahn, S.K., Maxwell, A. and Buttner, M.J. (2009) Coupling of the biosynthesis and export of the DNA gyrase inhibitor simocyclinone in Streptomyces antibioticus. *Mol Microbiol*, **72**, 1462-1474.

96. Le, T.B., Stevenson, C.E., Buttner, M.J. and Lawson, D.M. (2011a) Crystallization and preliminary X-ray analysis of the TetR-like efflux pump regulator SimR. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **67**: 307-309.

97. Le, T.B., Stevenson, C.E., Fiedler, H.P., Maxwell, A., Lawson, D.M., *et al.* (2011b) Structures of the TetR-like simocyclinone efflux pump repressor, SimR, and the mechanism of ligand-mediated derepression. *J Mol Biol* **408**: 40-56.

98. Le, T.B., Schumacher, M.A., Lawson, D.M., Brennan, R.G., Buttner, M.J. (2011c) The crystal structure of the TetR family transcriptional repressor SimR bound to

DNA and the role of a flexible N-terminal extension in minor groove binding. *Nucleic Acids Research*. Accepted.

99. Leclercq, R. and Courvalin, P. (1991) Intrinsic and unusual resistance to macrolide, lincosamide, and streptogramin antibiotics in bacteria. *Antimicrob Agents Chemother*, **35**, 1273-1276.

100. Lederer, T., Takahashi, M. and Hillen, W. (1995) Thermodynamic analysis of tetracycline-mediated induction of Tet repressor by a quantitative methylation protection assay. *Anal Biochem*, **232**, 190-196.

101. Leslie, A.G. (2006) The integration of macromolecular diffraction data. *Acta Crystallogr D Biol Crystallogr*, **62**, 48-57.

102. Levy, S.B. (1984) Resistance to tetracyclines p.191-204. in L. E. Bryan (ed.), Antimicrobial drug resistance. *Academic Press, New York, N.Y.*

103. Levy, S.B. (1988) Tetracycline resistance determinants are widespread. *ASM News*, **54**, 418-421.

104. Lewis, K. (2010) Persister cells. *Annu Rev Microbiol*, **64**, 357-372.

105. Li, S.M. and Heide, L. (2006) The biosynthetic gene clusters of aminocoumarin antibiotics. *Planta Med*, **72**, 1093-1099.

106. Long, F., Vagin, A.A., Young, P. and Murshudov, G.N. (2008) BALBES: a molecular-replacement pipeline. *Acta Crystallogr D Biol Crystallogr*, **64**, 125-132.

107. MacNeil, D.J., Gewain, K.M., Ruby, C.L., Dezeny, G., Gibbons, P.H. and MacNeil, T. (1992) Analysis of Streptomyces avermitilis genes required for avermectin biosynthesis utilizing a novel integration vector. *Gene*, **111**, 61-68.

108. Mahren, S. and Braun, V. (2003) The FecI extracytoplasmic-function sigma factor of Escherichia coli interacts with the beta' subunit of RNA polymerase. *J Bacteriol*, **185**, 1796-1802.

109. Martinez-Bueno, M., Molina-Henares, A.J., Pareja, E., Ramos, J.L. and Tobes, R. (2004) BacTregulators: a database of transcriptional regulators in bacteria and archaea. *Bioinformatics*, **20**, 2787-2791.

110. Matsushima, P., Broughton, M.C., Turner, J.R. and Baltz, R.H. (1994) Conjugal transfer of cosmid DNA from Escherichia coli to Saccharopolyspora spinosa: effects of chromosomal insertions on macrolide A83543 production. *Gene*, **146**, 39-45.

111. Maxam, A.M. and Gilbert, W. (1980) Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol*, **65**, 499-560.

112. Maxwell, A. and Lawson, D.M. (2003) The ATP-binding site of type II topoisomerases as a target for antibacterial drugs. *Curr Top Med Chem*, **3**, 283-303.

113. McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C. and Read, R.J. (2007) Phaser crystallographic software. *J Appl Crystallogr*, **40**, 658-674.

114. McMurry, L., Petrucci, R.E., Jr. and Levy, S.B. (1980) Active efflux of tetracycline encoded by four genetically different tetracycline resistance determinants in Escherichia coli. *Proc Natl Acad Sci U S A*, **77**, 3974-3977.

115. Meier, I., Wray, L.V. and Hillen, W. (1988) Differential regulation of the Tn10-encoded tetracycline resistance genes tetA and tetR by the tandem tet operators O1 and O2. *EMBO J*, **7**, 567-572.

116. Meyer, C.E. (1964) Rubradirin, a New Antibiotic. Ii. Isolation and Characterization. *Antimicrob Agents Chemother (Bethesda)*, **10**, 97-99.

117. Miller, D.J., Zhang, Y.M., Subramanian, C., Rock, C.O. and White, S.W. (2010

1.) Structural basis for the transcriptional regulation of membrane lipid homeostasis. *Nat Struct Mol Biol*, **17**, 971-975.

118. Morais Cabral, J.H., Jackson, A.P., Smith, C.V., Shikotra, N., Maxwell, A. and Liddington, R.C. (1997) Crystal structure of the breakage-reunion domain of DNA gyrase. *Nature*, **388**, 903-906.

119. Muller, G., Hecht, B., Helbl, V., Hinrichs, W., Saenger, W. and Hillen, W. (1995) Characterization of non-inducible Tet repressor mutants suggests conformational changes necessary for induction. *Nat Struct Biol*, **2**, 693-703.

120. Murshudov, G.N., Vagin, A.A. and Dodson, E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr*, **53**, 240-255.

121. Noguchi, N., Emura, A., Matsuyama, H., O'Hara, K., Sasatsu, M. and Kono, M. (1995) Nucleotide sequence and characterization of erythromycin resistance determinant that encodes macrolide 2'-phosphotransferase I in Escherichia coli. *Antimicrob Agents Chemother*, **39**, 2359-2363.

122. Nollmann, M., Crisona, N.J. and Arimondo, P.B. (2007) Thirty years of Escherichia coli DNA gyrase: from in vivo function to single-molecule mechanism. *Biochimie*, **89**, 490-499.

123. Olesky, M., Zhao, S., Rosenberg, R.L. and Nicholas, R.A. (2006) Porin-mediated antibiotic resistance in Neisseria gonorrhoeae: ion, solute, and antibiotic

permeation through PIB proteins with penB mutations. *J Bacteriol*, **188**, 2300-2308.

124. Oppegard, L.M., Hamann, B.L., Streck, K.R., Ellis, K.C., Fiedler, H.P., Khodursky, A.B. and Hiasa, H. (2009) In vivo and in vitro patterns of the activity of simocyclinone D8, an angucyclinone antibiotic from Streptomyces antibioticus. *Antimicrob Agents Chemother*, **53**, 2110-2119.

125. Orth, P., Cordes, F., Schnappinger, D., Hillen, W., Saenger, W. and Hinrichs, W. (1998) Conformational changes of the Tet repressor induced by tetracycline trapping. *J Mol Biol*, **279**, 439-447.

126. Orth, P., Saenger, W. and Hinrichs, W. (1999) Tetracycline-chelated Mg2+ ion initiates helix unwinding in Tet repressor induction. *Biochemistry*, **38**, 191-198.

127. Orth, P., Schnappinger, D., Hillen, W., Saenger, W. and Hinrichs, W. (2000) Structural basis of gene regulation by the tetracycline inducible Tet repressor-operator system. *Nat Struct Biol*, **7**, 215-219.

128. Ostash, I., Ostash, B., Luzhetskyy, A., Bechthold, A., Walker, S. and Fedorenko, V. (2008) Coordination of export and glycosylation of landomycins in Streptomyces cyanogenus S136. *FEMS Microbiol Lett*, **285**, 195-202.

129. Otten, S.L., Ferguson, J. and Hutchinson, C.R. (1995) Regulation of daunorubicin production in Streptomyces peucetius by the dnrR2 locus. *J Bacteriol*, **177**, 1216-1224.

130. Paget, M.S., Chamberlin, L., Atrih, A., Foster, S.J. and Buttner, M.J. (1999) Evidence that the extracytoplasmic function sigma factor sigmaE is required for normal cell wall structure in Streptomyces coelicolor A3(2). *J Bacteriol*, **181**, 204-211.

131. Painter, J. and Merritt, E.A. (2006) Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr D Biol Crystallogr*, **62**, 439-450.

132. Pal, D. and Chakrabarti, P. (2001) Non-hydrogen bond interactions involving the methionine sulfur atom. *J Biomol Struct Dyn*, **19**, 115-128.

133. Pansegrau, W., Lanka, E., Barth, P.T., Figurski, D.H., Guiney, D.G., Haas, D., Helinski, D.R., Schwab, H., Stanisich, V.A. and Thomas, C.M. (1994) Complete nucleotide sequence of Birmingham IncP alpha plasmids. Compilation and comparative analysis. *J Mol Biol*, **239**, 623-663.

134. Paulsen, I.T., Brown, M.H., Littlejohn, T.G., Mitchell, B.A. and Skurray, R.A. (1996) Multidrug resistance proteins QacA and QacB from Staphylococcus aureus: membrane topology and identification of residues involved in substrate specificity. *Proc Natl Acad Sci U S A*, **93**, 3630-3635.

135. Perera, I.C., Lee, Y.H., Wilkinson, S.P. and Grove, A. (2009) Mechanism for attenuation of DNA binding by MarR family transcriptional regulators by small molecule ligands. *J Mol Biol*, **390**, 1019-1029.

136. Pojer, F., Li, S.M. and Heide, L. (2002) Molecular cloning and sequence analysis of the clorobiocin biosynthetic gene cluster: new insights into the biosynthesis of aminocoumarin antibiotics. *Microbiology*, **148**, 3901-3911.

137. Poole, K. (2005) Efflux-mediated antimicrobial resistance. *J Antimicrob Chemother*, **56**, 20-51.

138. Raghavan, V. and Groisman, E.A. (2010) Orphan and hybrid two-component system proteins in health and disease. *Curr Opin Microbiol*, **13**, 226-231.

139. Ramos, J.L., Martinez-Bueno, M., Molina-Henares, A.J., Teran, W., Watanabe, K., Zhang, X., Gallegos, M.T., Brennan, R. and Tobes, R. (2005) The TetR family of transcriptional repressors. *Microbiol Mol Biol Rev*, **69**, 326-356.

140. Receveur-Brechot, V., Bourhis, J.M., Uversky, V.N., Canard, B. and Longhi, S. (2006) Assessing protein disorder and induced folding. *Proteins*, **62**, 24-45.

141. Reichheld, S.E., Yu, Z. and Davidson, A.R. (2009) The induction of folding cooperativity by ligand binding drives the allosteric response of tetracycline repressor. *Proc Natl Acad Sci U S A*, **106**, 22263-22268.

142. Roca, J. (1995) The mechanisms of DNA topoisomerases. *Trends Biochem Sci*, **20**, 156-160.

143. Roca, J. and Wang, J.C. (1992) The capture of a DNA double helix by an ATP-dependent protein clamp: a key step in DNA transport by type II DNA topoisomerases. *Cell*, **71**, 833-840.

144. Rohr, J. and Thiericke, R. (1992) Angucycline group antibiotics. *Nat Prod Rep*, **9**, 103-137.

145. Rohs, R., Jin, X., West, S.M., Joshi, R., Honig, B. and Mann, R.S. (2010) Origins of specificity in protein-DNA recognition. *Annu Rev Biochem*, **79**, 233-269.

146. Rohs, R., West, S.M., Sosinsky, A., Liu, P., Mann, R.S. and Honig, B. (2009) The role of DNA shape in protein-DNA recognition. *Nature*, **461**, 1248-1253.

147. Rojo, F. (1999) Repression of transcription initiation in bacteria. *J Bacteriol*, **181**, 2987-2991.

148. Rosenfeld, N., Elowitz, M.B. and Alon, U. (2002) Negative autoregulation speeds the response times of transcription networks. *J Mol Biol*, **323**, 785-793.

149. Ross, J.I., Eady, E.A., Cove, J.H., Cunliffe, W.J., Baumberg, S. and Wootton, J.C. (1990) Inducible erythromycin resistance in staphylococci is encoded by a member of the ATP-binding transport super-gene family. *Mol Microbiol*, **4**, 1207-1214.

150. Ruiz, N., Falcone, B., Kahne, D. and Silhavy, T.J. (2005) Chemical conditionality: a genetic strategy to probe organelle assembly. *Cell*, **121**, 307-317.

151. Ruthenburg, A.J., Graybosch, D.M., Huetsch, J.C. and Verdine, G.L. (2005) A superhelical spiral in the Escherichia coli DNA gyrase A C-terminal domain imparts unidirectional supercoiling bias. *J Biol Chem*, **280**, 26177-26184.

152. Saenger, W., Orth, P., Kisker, C., Hillen, W. and Hinrichs, W. (2000) The Tetracycline Repressor-A Paradigm for a Biological Switch. *Angew Chem Int Ed Engl*, **39**, 2042-2052.

153. Sambrook, J., Fritsh, E.F. and Maniatis, T. (1989) Molecular Cloning: a laboratory manual. *Cold Spring Habor Press, Cold Spring Harbor, N.Y. 2nd ed*.

154. Schimana, J., Fiedler, H.P., Groth, I., Sussmuth, R., Beil, W., Walker, M. and Zeeck, A. (2000) Simocyclinones, novel cytostatic angucyclinone antibiotics produced by Streptomyces antibioticus Tu 6040. I. Taxonomy, fermentation, isolation and biological activities. *J Antibiot (Tokyo)*, **53**, 779-787.

155. Schimana, J., Walker, M., Zeeck, A. and Fiedler, P. (2001) Simocyclinones: diversity of metabolites is dependent on fermentation conditions. *J Ind Microbiol Biotechnol*, **27**, 144-148.

156. Schmutz, E., Muhlenweg, A., Li, S.M. and Heide, L. (2003) Resistance genes of aminocoumarin producers: two type II topoisomerase genes confer resistance against coumermycin A1 and clorobiocin. *Antimicrob Agents Chemother*, **47**, 869-877.

157. Schoeffler, A.J. and Berger, J.M. (2008) DNA topoisomerases: harnessing and constraining energy to govern chromosome topology. *Q Rev Biophys*, **41**, 41-101.

158. Scholz, O., Henssler, E.M., Bail, J., Schubert, P., Bogdanska-Urbaniak, J., Sopp, S., Reich, M., Wisshak, S., Kostner, M., Bertram, R. *et al.* (2004) Activity reversal of Tet repressor caused by single amino acid exchanges. *Mol Microbiol*, **53**, 777-789.

159. Schumacher, M.A., Miller, M.C., Grkovic, S., Brown, M.H., Skurray, R.A. and Brennan, R.G. (2001) Structural mechanisms of QacR induction and multidrug recognition. *Science*, **294**, 2158-2163.

160. Schumacher, M.A., Miller, M.C., Grkovic, S., Brown, M.H., Skurray, R.A. and Brennan, R.G. (2002) Structural basis for cooperative DNA binding by two dimers of the multidrug-binding protein QacR. *Embo J*, **21**, 1210-1218.

161. Sheldrick, G.M. (2008) A short history of SHELX. *Acta Crystallogr A*, **64**, 112-122.

162. Singh, M.P., Petersen, P.J., Weiss, W.J., Janso, J.E., Luckman, S.W., Lenoy, E.B., Bradford, P.A., Testa, R.T. and Greenstein, M. (2003) Mannopeptimycins, new cyclic glycopeptide antibiotics produced by Streptomyces hygroscopicus LL-AC98: antibacterial and mechanistic activities. *Antimicrob Agents Chemother*, **47**, 62-69.

163. Skeggs, P.A., Holmes, D.J. and Cundliffe, E. (1987) Cloning of aminoglycoside-resistance determinants from Streptomyces tenebrarius and comparison with related genes from other actinomycetes. *J Gen Microbiol*, **133**, 915-923.

164. Sosio, M., Giusino, F., Cappellano, C., Bossi, E., Puglia, A.M. and Donadio, S. (2000) Artificial chromosomes for antibiotic-producing actinomycetes. *Nat Biotechnol*, **18**, 343-345.

165. Steffensky, M., Muhlenweg, A., Wang, Z.X., Li, S.M. and Heide, L. (2000) Identification of the novobiocin biosynthetic gene cluster of Streptomyces spheroides NCIB 11891. *Antimicrob Agents Chemother*, **44**, 1214-1222.

166. Studier, F.W. and Moffatt, B.A. (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol*, **189**, 113-130.

167. Surette, M.G. and Bassler, B.L. (1998) Quorum sensing in Escherichia coli and Salmonella typhimurium. *Proc Natl Acad Sci U S A*, **95**, 7046-7050.

168. Szczepanowski, R., Krahn, I., Bohn, N., Puhler, A. and Schluter, A. (2007) Novel macrolide resistance module carried by the IncP-1beta resistance plasmid pRSB111, isolated from a wastewater treatment plant. *Antimicrob Agents Chemother*, **51**, 673-678.

169. Szczepanowski, R., Krahn, I., Linke, B., Goesmann, A., Puhler, A. and Schluter, A. (2004) Antibiotic multiresistance plasmid pRSB101 isolated from a wastewater treatment plant is related to plasmids residing in phytopathogenic bacteria and carries eight different resistance determinants including a multidrug transport system. *Microbiology*, **150**, 3613-3630.

170. Taguchi, T., Itou, K., Ebizuka, Y., Malpartida, F., Hopwood, D.A., Surti, C.M., Booker-Milburn, K.I., Stephenson, G.R. and Ichinose, K. (2000) Chemical characterisation of disruptants of the Streptomyces coelicolor A3(2) actVI genes involved in actinorhodin biosynthesis. *J Antibiot (Tokyo)*, **53**, 144-152.

171. Tahlan, K., Yu, Z., Xu, Y., Davidson, A.R. and Nodwell, J.R. (2008) Ligand recognition by ActR, a TetR-like regulator of actinorhodin export. *J Mol Biol*, **383**,

753-761.

172. Tauch, A., Puhler, A., Kalinowski, J. and Thierbach, G. (2000) TetZ, a new tetracycline resistance determinant discovered in gram-positive bacteria, shows high homology to gram-negative regulated efflux systems. *Plasmid*, **44**, 285-291.

173. Theobald, U., Schimana, J. and Fiedler, H.P. (2000) Microbial growth and production kinetics of Streptomyces antibioticus Tu 6040. *Antonie Van Leeuwenhoek*, **78**, 307-313.

174. Thiara, A.S. and Cundliffe, E. (1988) Cloning and characterization of a DNA gyrase B gene from Streptomyces sphaeroides that confers resistance to novobiocin. *EMBO J*, **7**, 2255-2259.

175. Thiara, A.S. and Cundliffe, E. (1989) Interplay of novobiocin-resistant and -sensitive DNA gyrase activities in self-protection of the novobiocin producer, Streptomyces sphaeroides. *Gene*, **81**, 65-72.

176. Thiara, A.S. and Cundliffe, E. (1993) Expression and analysis of two gyrB genes from the novobiocin producer, Streptomyces sphaeroides. *Mol Microbiol*, **8**, 495-506.

177. Trefzer, A., Pelzer, S., Schimana, J., Stockert, S., Bihlmaier, C., Fiedler, H.P., Welzel, K., Vente, A. and Bechthold, A. (2002) Biosynthetic gene cluster of simocyclinone, a natural multihybrid antibiotic. *Antimicrob Agents Chemother*, **46**, 1174-1182.

178. Vagin, A. and Teplyakov, A. (2000) An approach to multi-copy search in molecular replacement. *Acta Crystallogr D Biol Crystallogr*, **56**, 1622-1624.

179. Walsh, C. (2003) Antibiotics: Actions, origins, resistance. *ASM Press, Washington, DC*.

180. Wang, L., Tian, X., Wang, J., Yang, H., Fan, K., Xu, G., Yang, K. and Tan, H. (2009) Autoregulation of antibiotic biosynthesis by binding of the end product to an atypical response regulator. *Proc Natl Acad Sci U S A*, **106**, 8617-8622.

181. Wang, Z.X., Li, S.M. and Heide, L. (2000) Identification of the coumermycin A(1) biosynthetic gene cluster of Streptomyces rishiriensis DSM 40489. *Antimicrob Agents Chemother*, **44**, 3040-3048.

182. Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M. and Barton, G.J. (2009) Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, **25**, 1189-1191.

183. Weisblum, B. (1995) Erythromycin resistance by ribosome modification.

Antimicrob Agents Chemother, **39**, 577-585.

184. Wigley, D.B., Davies, G.J., Dodson, E.J., Maxwell, A. and Dodson, G. (1991) Crystal structure of an N-terminal fragment of the DNA gyrase B protein. *Nature*, **351**, 624-629.

185. Wilkinson, S.P. and Grove, A. (2006) Ligand-responsive transcriptional regulation by members of the MarR family of winged helix proteins. *Curr Issues Mol Biol*, **8**, 51-62.

186. Willems, A.R., Tahlan, K., Taguchi, T., Zhang, K., Lee, Z.Z., Ichinose, K., Junop, M.S. and Nodwell, J.R. (2008) Crystal structures of the Streptomyces coelicolor TetR-like protein ActR alone and in complex with actinorhodin or the actinorhodin biosynthetic precursor (S)-DNPA. *J Mol Biol*, **376**, 1377-1387.

187. Wright, G.D. (2007) The antibiotic resistome: the nexus of chemical and genetic diversity. *Nat Rev Microbiol*, **5**, 175-186.

188. Yang, Z.R., Thomson, R., McNeil, P. and Esnouf, R.M. (2005) RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics*, **21**, 3369-3376.

189. Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene*, **33**, 103-119.

190. Yu, Z., Reichheld, S.E., Savchenko, A., Parkinson, J. and Davidson, A.R. (2010) A comprehensive analysis of structural and sequence conservation in the TetR family transcriptional regulators. *J Mol Biol*, **400**, 847-864.

191. Zhang, Y.M., Zhu, K., Frank, M.W. and Rock, C.O. (2007) A Pseudomonas aeruginosa transcription factor that senses fatty acid structure. *Mol Microbiol*, **66**, 622-632.

192. Zhao, J. and Aoki, T. (1992) Nucleotide sequence analysis of the class G tetracycline resistance determinant from Vibrio anguillarum. *Microbiol Immunol*, **36**, 1051-1060.

193. Zheng, J., Sagar, V., Smolinsky, A., Bourke, C., LaRonde-LeBlanc, N. and Cropp, T.A. (2009) Structure and function of the macrolide biosensor protein, MphR(A), with and without erythromycin. *J Mol Biol*, **387**, 1250-1260.

194. Zhu, K., Choi, K.H., Schweizer, H.P., Rock, C.O. and Zhang, Y.M. (2006) Two aerobic pathways for the formation of unsaturated fatty acids in Pseudomonas aeruginosa. *Mol Microbiol*, **60**, 260-273.

Publications

The coupling of the biosynthesis and export of the DNA gyrase inhibitor simocyclinone in *Streptomyces antibioticus*

By

Tung Ba Khanh Le

2011

A thesis submitted to the University of East Anglia, Norwich, for the Degree of Doctor of Philosophy

A Crystal Structure of the Bifunctional Antibiotic Simocyclinone D8, Bound to DNA Gyrase

Marcus J. Edwards, *et al.*

Science **326**, 1415 (2009);
DOI: 10.1126/science.1179123

the canonical TnaC sequence, indicating that even residues unrelated to the stalling process can adopt a distinct conformation within the exit tunnel. This notion is supported by a cryo-EM structure of a yeast 80*S* ribosome-nascent chain complex stalled during the translation of a truncated dipeptidyl-aminopeptidase B (DP120) mRNA at 6.1 Å resolution (*27*). Although the DP120 sequence has no stalling capacity, density for this nascent chain is visible, indicating a preferred conformation within the exit tunnel (Fig. 4D). Notably, the DP120 nascent chain follows a different path from that reported here for TnaC (Fig. 4, E and F). Clearly the chemical and electrostatic properties of the tunnel environment play a pivotal role in facilitating this kind of distinct nascent chain behavior (*3*, *4*). The finding that nascent chains with little or no sequence conservation interact with the exit tunnel in a distinct manner and adopt individual conformations may be important not only for initial folding events (*1*–*3*) but also for the variety of nascent chain–mediated regulatory mechanisms (*5*).

References and Notes

1. C. A. Woolhead, P. J. McCormick, A. E. Johnson, *Cell* **116**, 725 (2004).
2. J. Lu, C. Deutsch, *Biochemistry* **44**, 8230 (2005).
3. J. Lu, C. Deutsch, *Nat. Struct. Mol. Biol.* **12**, 1123 (2005).
4. J. Lu, C. Deutsch, *J. Mol. Biol.* **384**, 73 (2008).
5. T. Tenson, M. Ehrenberg, *Cell* **108**, 591 (2002).
6. H. Nakatogawa, K. Ito, *Cell* **108**, 629 (2002).
7. C. A. Woolhead, A. E. Johnson, H. D. Bernstein, *Mol. Cell* **22**, 587 (2006).
8. M. N. Yap, H. D. Bernstein, *Mol. Cell* **34**, 201 (2009).
9. D. Oliver, J. Norman, S. Sarker, *J. Bacteriol.* **180**, 5240 (1998).
10. F. Gong, C. Yanofsky, *Science* **297**, 1864 (2002).
11. L. Cruz-Vera, S. Rajagopal, C. Squires, C. Yanofsky, *Mol. Cell* **19**, 333 (2005).
12. L. R. Cruz-Vera, C. Yanofsky, *J. Bacteriol.* **190**, 4791 (2008).
13. F. Gong, K. Ito, Y. Nakamura, C. Yanofsky, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 8997 (2001).
14. Materials and methods are available as supporting material on *Science* Online.
15. L. G. Trabuco, E. Villa, K. Mitra, J. Frank, K. Schulten, *Structure* **16**, 673 (2008).
16. P. I. de Bakker, N. Furnham, T. L. Blundell, M. A. DePristo, *Curr. Opin. Struct. Biol.* **16**, 160 (2006).
17. M. Simonovic, T. A. Steitz, *Biochim. Biophys. Acta* **1789**, 612 (2009).
18. R. Yang, L. R. Cruz-Vera, C. Yanofsky, *J. Bacteriol.* **191**, 3445 (2009).
19. N. Vazquez-Laslop, C. Thum, A. S. Mankin, *Mol. Cell* **30**, 190 (2008).
20. L. R. Cruz-Vera, A. New, C. Squires, C. Yanofsky, *J. Bacteriol.* **189**, 3140 (2007).
21. T. M. Schmeing, K. S. Huang, S. A. Strobel, T. A. Steitz, *Nature* **438**, 520 (2005).
22. T. M. Schmeing, K. S. Huang, D. E. Kitchen, S. A. Strobel, T. A. Steitz, *Mol. Cell* **20**, 437 (2005).
23. L. R. Cruz-Vera, M. Gong, C. Yanofsky, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 3598 (2006).
24. A. Weixlbaumer *et al.*, *Science* **322**, 953 (2008).
25. K. Mitra *et al.*, *Mol. Cell* **22**, 533 (2006).
26. P. Nissen, J. Hansen, N. Ban, P. B. Moore, T. A. Steitz, *Science* **289**, 920 (2000).
27. T. Becker *et al.*, *Science* **326**, 1369 (2009). Published online 29 October 2009; 10.1126/science.1178535.

28. This research was supported by grants from the Deutsche Forschungsgemeinschaft SFB594 and SFB646 (to R.B.), SFB740 (to T.M.), and WI3285/1-1 (to D.N.W.); by NIH grants GM022778 (to T.A.S.) and P41-RR05969 (to K.S.); by NSF grant PHY0822613 (to K.S.); and by the European Union and Senatsverwaltung für Wissenschaft, Forschung und Kultur Berlin (UltraStructureNetwork, Anwenderzentrum). Computer time for MDFF was provided through an NSF Large Resources Allocation Committee grant MCA93S028. Coordinates of the atomic models of TnaC-70*S* complex have been deposited in the PDB under accession numbers 2WWL (30*S*) and 2WWQ (50*S*). The cryo-EM map has been deposited in the 3D-EM database under accession number EMD-1657.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1177662/DC1
Materials and Methods
Figs. S1 to S9
Table S1
Movie S1
References

12 June 2009; accepted 21 October 2009
Published online 29 October 2009;
10.1126/science.1177662
Include this information when citing this paper.

A Crystal Structure of the Bifunctional Antibiotic Simocyclinone D8, Bound to DNA Gyrase

Marcus J. Edwards,[1] Ruth H. Flatman,[1] Lesley A. Mitchenall,[1] Clare E.M. Stevenson,[1] Tung B.K. Le,[2] Thomas A. Clarke,[3] Adam R. McKay,[4] Hans-Peter Fiedler,[5] Mark J. Buttner,[2] David M. Lawson,[1] Anthony Maxwell[1]*

Simocyclinones are bifunctional antibiotics that inhibit bacterial DNA gyrase by preventing DNA binding to the enzyme. We report the crystal structure of the complex formed between the N-terminal domain of the *Escherichia coli* gyrase A subunit and simocyclinone D8, revealing two binding pockets that separately accommodate the aminocoumarin and polyketide moieties of the antibiotic. These are close to, but distinct from, the quinolone-binding site, consistent with our observations that several mutations in this region confer resistance to both agents. Biochemical studies show that the individual moieties of simocyclinone D8 are comparatively weak inhibitors of gyrase relative to the parent compound, but their combination generates a more potent inhibitor. Our results should facilitate the design of drug molecules that target these unexploited binding pockets.

Bacterial diseases remain a major problem because of the emergence of drug-resistant bacteria combined with the dearth of new antibacterial agents. Despite extensive efforts, there remain relatively few effective drug targets for antibacterials. One of these is the enzyme DNA gyrase, a DNA topoisomerase that controls the topology of DNA (*1*, *2*). Topoisomerases are classified into two types, I and II, depending on whether they catalyze reactions involving the transient breakage of one or both strands of DNA. Gyrase is the only type II DNA topoisomerase that can catalyze DNA supercoiling; this reaction is driven by the free energy of adenosine triphosphate (ATP) hydrolysis (*3*). Gyrase consists of two subunits, GyrA and GyrB (97 kD and 90 kD, respectively, in *Escherichia coli*), which form an A_2B_2 complex in the active enzyme. Because gyrase is essential in bacteria and lacking in humans, it is a valuable drug target (*4*). The complexity of the gyrase supercoiling reaction presents multiple opportunities for intervention. Two well-known groups of gyrase-specific antibacterial agents are quinolones and aminocoumarins. Fluoroquinolones, such as ciprofloxacin, are highly successful drugs (*5*), but their usefulness is diminishing as a consequence of bacterial resistance (*6*). Aminocoumarins, e.g., novobiocin and clorobiocin, are less successful clinically because of toxicity and solubility issues but are very well characterized in terms of their mode of action on gyrase (*7*), including several crystal structures (*8*–*11*). Aminocoumarins act by competitively inhibiting the binding of ATP to the GyrB subunit (*11*). The cloning and sequencing of the biosynthetic pathways for the aminocoumarins novobiocin, clorobiocin, and coumermycin A_1 and the application of bioengineering methodologies (*12*) have enabled the production of a series of modified aminocoumarins with varying potencies against their targets, gyrase and topoisomerase IV (*13*, *14*). This work has raised the possibility of engineering antibacterial agents targeted to gyrase that are based on natural antibiotics.

Simocyclinone D8 (SD8) was isolated from *Streptomyces antibioticus* Tü 6040 (*15*–*18*). The antibiotic consists of a chlorinated aminocoumarin (AC) linked to an angucyclic polyketide (PK) via

[1]Department of Biological Chemistry, John Innes Centre, Colney, Norwich NR4 7UH, UK. [2]Department of Molecular Microbiology, John Innes Centre, Colney, Norwich NR4 7UH, UK. [3]School of Biological Sciences, University of East Anglia, Norwich NR4 7TJ, UK. [4]Department of Chemistry, University College London, 20 Gordon St, London WC1H 0AJ, UK. [5]Mikrobiologisches Institut, Eberhard-Karls-Universität Tübingen, Auf der Morgenstelle 28, D-72076 Tübingen, Germany.

*To whom correspondence should be addressed. E-mail: tony.maxwell@bbsrc.ac.uk

a tetraene linker and a D-olivose sugar (Fig. 1). Because of the presence of the AC moiety, the expectation was that SD8 would target the adenosine triphosphatase (ATPase) domain of GyrB. Although SD8 is a potent inhibitor of *E. coli* gyrase, it does not inhibit the intrinsic GyrB ATPase activity. Instead, SD8 binds to the N-terminal domain of GyrA and prevents DNA binding (*19*). In hindsight, this is not surprising because SD8 lacks the decorated noviose sugar that is attached to the 7-OH of the AC ring and is involved in the majority of the interactions with GyrB in other aminocoumarins (*8*–*11*). In contrast to quinolones and aminocoumarins, which can act on both gyrase and its close relative topoisomerase IV, SD8 is potent against gyrases from *E. coli* and *Staphylococcus aureus* but much less effective against topoisomerase IV from these species (*20*).

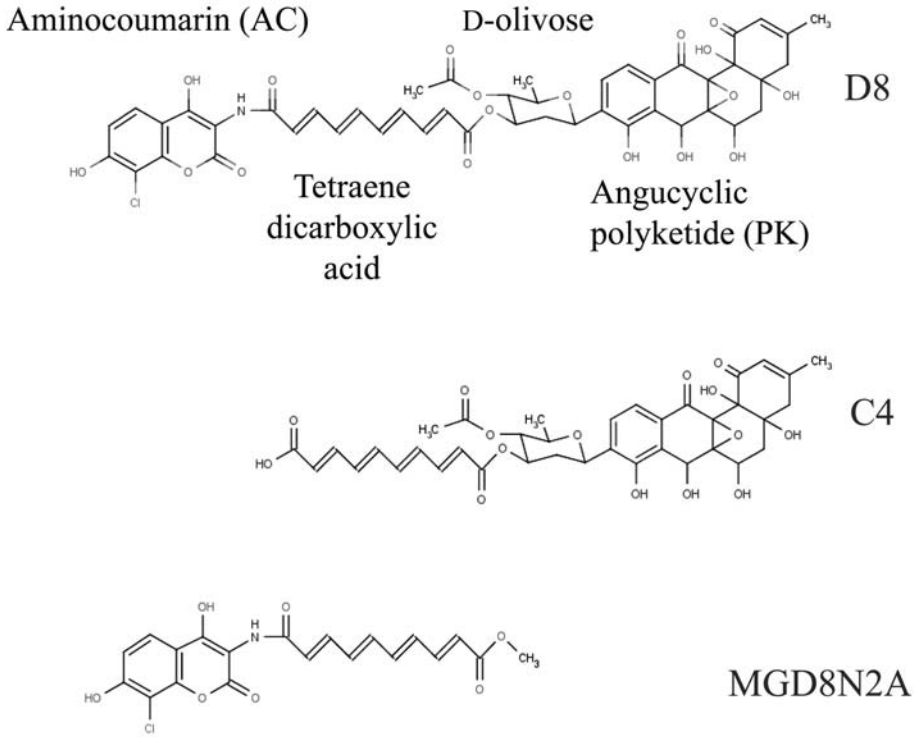


Fig. 1. Structure of simocyclinone D8 and analogs. IC_{50} values for inhibition of supercoiling by gyrase: for D8, 0.6 μM; C4, 70 μM; MGD8N2A, 50 μM.

We have crystallized the N-terminal domain of GyrA (GyrA59) complexed with SD8 and determined the crystal structure at 2.6-Å resolution by molecular replacement from the structure of unliganded GyrA59 that was previously determined (*21*). Our structure reveals a ligand-stabilized homotetramer of GyrA59 subunits consisting of two A59 dimers cross-linked by four molecules of SD8 (Fig. 2). The tetraene linker of SD8 acts as an extended rod, about 10 Å long, that holds the AC and PK moieties apart. Each GyrA subunit has distinct pockets that accept the AC and PK groups, respectively, of two separate SD8 molecules; both pockets lie in the predicted DNA binding saddle. Additional lobes of electron density adjacent to the PK moiety have been modeled as Mg^{2+} ions (Fig. 3A and fig. S1). Although each subunit interacts with two SD8 molecules, because each of these molecules is shared by two subunits from opposing dimers, the stoichiometry remains 1:1, consistent with previous experiments (*19*). In addition to the SD8-mediated dimer-dimer interactions, there is about 1500 $Å^2$ of protein-protein interface. This includes 12 hydrogen bonds, 10 of which involve residues spanning Leu^{17} to Asp^{23}, a region just before α-helix 1 that was not visible in the original GyrA59 structure (*21*). Superposition of the SD8 complex and ligand-free GyrA59 structures gives root mean square deviation values below 1 Å, both for

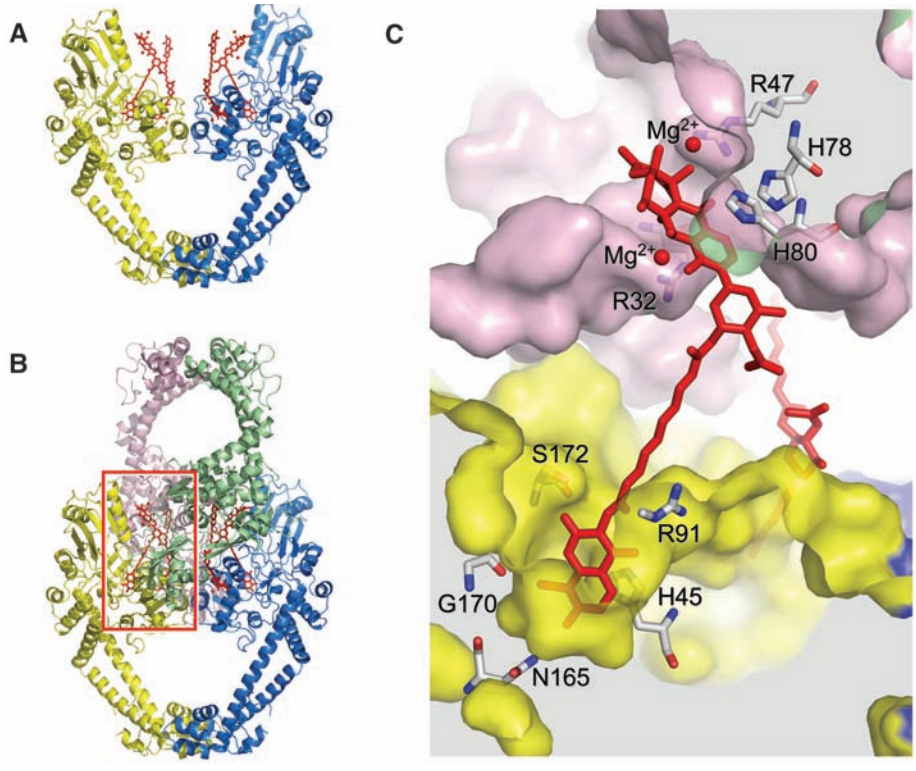


Fig. 2. Crystal structure of the GyrA59-simocyclinone complex. The protein is depicted in cartoon representation, the SD8 molecules are shown as red sticks, and their associated Mg^{2+} ions as small spheres. (**A**) Structure of the GyrA59 dimer and the four SD8 molecules it interacts with. (**B**) Structure of the SD8-mediated tetramer (dimer of dimers). (**C**) Close-up of the red boxed region in (B) showing a section through the complex containing two SD8 molecules. The same color scheme is adopted as for (A) and (B), but the GyrA59 subunits are represented as semi-transparent molecular surfaces. Regions with a white background are either outside the complex or between the subunits; regions with a gray background are within the molecular envelopes. Key residues (*23*) that are close to the SD8 molecule in the foreground are displayed in stick representation.

subunit-subunit and dimer-dimer comparisons, indicating that there are no major conformational changes upon ligand binding (fig. S2). Intriguingly, a single SD8 molecule can be modeled in a "bent" conformation such that it bridges between the AC and PK pockets of the same subunit, while maintaining essentially the same contacts with the protein seen in the crystal structure (fig. S3). This places the tetraene linker close to, and possibly interacting with, α-helix 4.

To investigate the oligomeric state in solution, we performed a series of molecular weight (MW) studies. Analytical ultracentrifugation (table S2) showed that the MW of GyrA59 in the absence of SD8 or at low ligand:protein ratios (<~3:1) was ~120 kD, suggesting a dimer, whereas at high ligand:protein ratios (>~4:1) GyrA59 had a MW of ~250 kD, consistent with a tetramer. By using nanoelectrospray ionization mass spectrometry (nanoESI MS) under conditions where noncovalent interactions are preserved, we assessed the binding of SD8 to GyrA59 and full-length GyrA. In the absence of SD8, both proteins had MWs consistent with dimers (fig. S4). Titration of SD8 into solutions containing the proteins showed dimeric species with either one or two SD8 molecules bound at ligand:protein ratios of <~2:1. With a ligand:protein ratio of 3:1, we began to see formation of a tetrameric species, which increased with increasing SD8 concentrations to become the predominant species at ~7.5:1 (fig. S4). These experiments showed small amounts of three SD8 molecules bound per dimer but no evidence of four, suggesting that a tetramer readily forms once four molecules are bound (fig. S4). We suggest that the dimeric species, observed at limiting ligand concentrations, might represent a single SD8 molecule bound to the AC and PK pockets within the same subunit (fig. S3).

To probe the importance of the two binding pockets, we analyzed the interactions of GyrA59 with simocyclinone analogs lacking either the PK or the AC moiety (Fig. 1). Simocyclinone C4 is a naturally occurring intermediate in the SD8 pathway that lacks the AC moiety; MGD8N2A, which lacks the PK moiety, was generated by chemical hydrolysis of SD8. The parent compound has a minimum inhibitory concentration (IC_{50}) value of 0.6 μM for inhibition of gyrase supercoiling, whereas IC_{50} values of 70 and 50 μM were obtained for analogs lacking either the AC or the PK moiety, respectively (Fig. 1). Although inhibition is greatly reduced, the fact that these SD8 analogs have some activity suggests that cross-linking of the two GyrA dimers is not a prerequisite for inhibition.

One key issue was to establish the in vivo target of simocyclinones; recent transcriptional profiling studies (*20*) suggest but do not prove that gyrase is the target. To address this, we selected spontaneous resistant mutants in *E. coli*. Wild-type *E. coli* and other Gram-negative bacteria are resistant to simocyclinones because the compounds cannot penetrate the outer membrane (*15*); we therefore used an *E. coli* strain (NR698) that is sensitive because it carries an in-frame de-

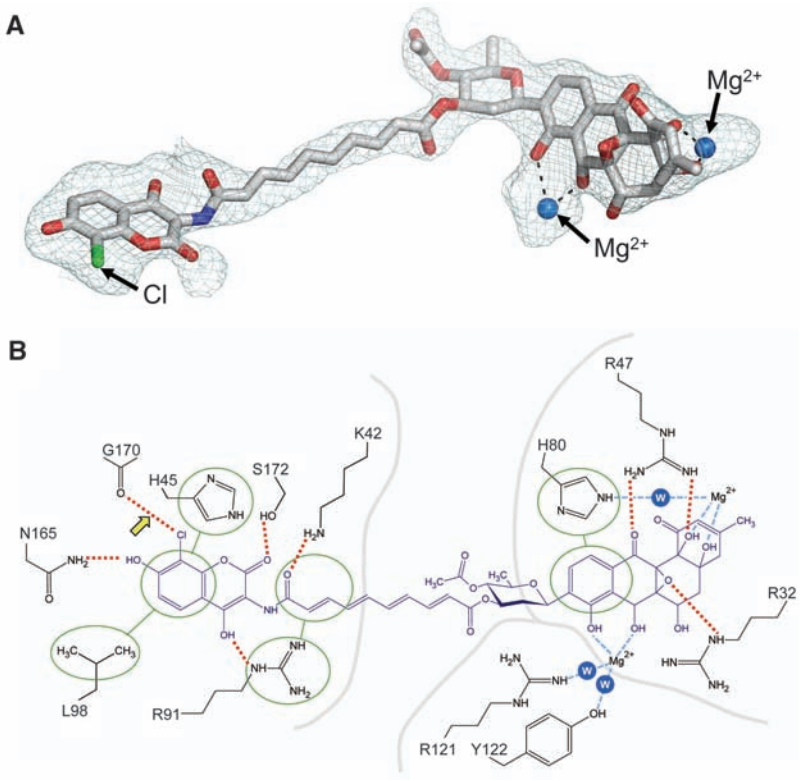


Fig. 3. Binding of SD8 to GyrA. (**A**) Simulated annealing omit electron density map for SD8 contoured at 6 σ and superposed on the final coordinates of the ligand (*27*). (**B**) Schematic figure detailing protein-ligand interactions in the GyrA59-SD8 complex. Red dotted lines represent hydrogen bonds and the single halogen bond (indicated by yellow arrow; see fig. S1). Interactions with the Mg^{2+} ions are shown as pale blue dashed lines. The side chains of Arg^{121} and Tyr^{122} (in the active site) point toward the SD8 molecule across the dimer interface and could interact with it via water molecules (blue shaded circles labeled "W") coordinated to a Mg^{2+} ion. Similarly, His^{80} likely makes a water-mediated interaction with the second Mg^{2+} ion. Nonbonded interactions are represented by the linked green ovals, which encircle the groups involved; subunit boundaries are delineated by gray lines.

Table 1. Properties of simocyclinone- and quinolone-resistant GyrA mutants. NA indicates not applicable. Mutant H78A is not active; K_D for simocyclinone is ~10 times that of wild type. S83W data is from (*19*).

GyrA mutation	Relative IC_{50} (supercoiling)	
	Simocyclinone	Ciprofloxacin
Wild type	1 (0.6 μM)*	1 (0.7 μM)*
Mutations in the simocyclinone-binding site		
Aminocoumarin-binding pocket mutations		
H45A	9.1	2.2
R91A	20	1.1
Polyketide-binding pocket mutations		
H78A	NA	NA
H80A	230	2.8
Mutations in both pockets		
H45A and H80A	>500	2.3
Mutations in the quinolone-resistance–determining region of GyrA		
G81D	40	24
S83W	10	30
A84P	38	28
D87A	7.2	5.2
D87Y	57	30
S83A and D87A	8.3	12

*Actual IC_{50} values are given in parentheses.

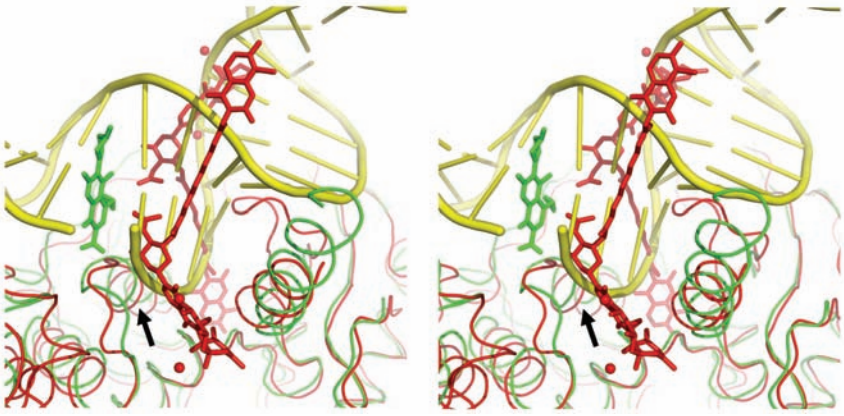


Fig. 4. Stereo view showing a superposition of GyrA59-SD8 and the clinafloxacin-DNA cleavage complex of topo IV [Protein Data Bank (PDB) code 3FOE], focusing on the drug-binding sites in one half of the GyrA/ParC dimers (*28*). The protein backbones are shown as ribbons and the drug molecules in stick representation, with GyrA59-SD8 in red, ParC-clinafloxacin in green, and the DNA from the latter structure in yellow. For clarity, the ParE subunits have been omitted from 3FOE. The SD8 and clinafloxacin binding sites are adjacent but do not overlap. This figure illustrates how SD8 would interfere with DNA binding; this would also be the case if the drug were bound in the "bent-over" conformation proposed in fig. S3. The arrow indicates the position of α-helix 4.

letion in the *imp* (increased membrane permeability) gene (*22*). We isolated 31 spontaneous simocyclinone-resistant mutants and in each case sequenced a ~500–base pair (bp) region of *gyrA*, corresponding to residues Met^{26} to Ser^{172} of the protein. We found *gyrA* mutations in 22 of them, conferring one of the following amino acid changes: V^{44}→G^{44} (V44G) (*23*), H45Y, H45Q, G81S, and D87Y (fig. S5). These amino acids are close to the bound SD8 molecule in the crystal structure, consistent with gyrase being the in vivo target. Unlike the 22 *gyrA* mutants, the remaining nine isolates had also acquired resistance to bile salts, suggesting that they could be accounted for by spontaneous second-site mutations that are known to restore outer membrane impermeability to the *imp* mutant (*22*).

On the basis of the spontaneous mutations and the crystal structure information, we made selected site-directed mutations in GyrA to probe its interaction with SD8 in vitro. The mutant proteins, together with wild-type GyrB, were assayed for DNA supercoiling in the presence of SD8 (Table 1). Mutations in either the AC (His^{45} and Arg^{91}) or the PK (His^{80} and Gly^{81}) pocket showed simocyclinone-resistant supercoiling. SD8 binding to an inactive mutant in the PK pocket, H78A (*23*), was investigated by surface-plasmon resonance using the GyrA59 protein. The mutant showed decreased binding affinity for SD8 (K_D values were 1.3 μM for wild type and 10.4 μM for H78A) and had a near-identical far-ultraviolet circular dichroism spectrum to wild-type GyrA, suggesting that it was properly folded.

Quinolone-resistant mutations map to both *gyrA* and *gyrB* in regions known as the quinolone-resistance determining regions [QRDRs (*24*, *25*)]. In the case of *E. coli* GyrA, the QRDR occurs between amino acids 67 and 106 with mutations identified at Ala^{67}, Gly^{81}, Asp^{82}, Ser^{83}, Ala^{84}, Asp^{87}, and Gln^{106} (*26*), mostly occurring either in or just before α-helix 4 (*21*) (fig. S5). From the published structures of quinolone-DNA cleavage complexes of *Streptococcus pneumoniae* topo IV (*27*), we can infer the location of the quinolone-binding site in GyrA, which is adjacent to but not overlapping the SD8 binding sites (Fig. 4). The quinolones do not make substantive contacts with the topo IV ParC protein, being closest to the equivalents of residues Gly^{81} to Ala^{84} and Asp^{87} in GyrA. Therefore, at least some of the mutations in the QRDR of GyrA most likely have indirect effects on quinolone binding. Given the proximity of the quinolone and SD8 binding sites, we investigated whether there was any cross-resistance between the two types of inhibitor. SD8-resistant mutants were tested for their susceptibility to the fluoroquinolone ciprofloxacin, and a range of ciprofloxacin-resistant mutants were tested for their susceptibility to SD8 (Table 1).

Mutations in the simocyclinone-binding pockets (AC and PK) result in near–wild-type amounts of susceptibility to ciprofloxacin (Table 1); QRDR mutations in α-helix 4 of GyrA confer increased resistance to both ciprofloxacin and SD8. None of these amino acids makes direct contacts with bound SD8 (Fig. 3B); given the low resolution of the quinolone-DNA-topo IV complex structures, it is not possible to precisely define any ligand-protein interactions, but it is likely that substitutions at positions 81 to 84 and 87 in GyrA would have an effect on drug binding. The prevalence of mutations at Ser^{83} and Asp^{87} in quinolone-resistant clinical isolates supports this assertion (*26*). In the case of SD8, it is possible that mutations in α-helix 4 of GyrA, which lies between the AC and PK binding pockets, can affect the proposed bridging of the two binding sites by the tetraene linker (fig. S3).

Given the global concerns over drug-resistant bacterial diseases, work on SD8 raises the prospect of developing agents that exploit its bifunctional mode of antibiotic action on a well-validated target. Alternatively, designing monofunctional compounds with enhanced affinity for one or the other of the binding sites may prove fruitful.

References and Notes

1. A. D. Bates, A. Maxwell, *DNA Topology* (Oxford Univ. Press, Oxford, 2005).
2. A. J. Schoeffler, J. M. Berger, *Q. Rev. Biophys.* **41**, 41 (2008).
3. A. D. Bates, A. Maxwell, *Biochemistry* **46**, 7929 (2007).
4. A. Maxwell, *Trends Microbiol.* **5**, 102 (1997).
5. C. M. Oliphant, G. M. Green, *Am. Fam. Physician* **65**, 455 (2002).
6. D. Livermore, *J. Antimicrob. Chemother.* **60** (suppl. 1), i59 (2007).
7. A. Maxwell, D. M. Lawson, *Curr. Top. Med. Chem.* **3**, 283 (2003).
8. V. Lamour, L. Hoermann, J. M. Jeltsch, P. Oudet, D. Moras, *J. Biol. Chem.* **277**, 18947 (2002).
9. F. T. F. Tsai *et al.*, *Proteins* **28**, 41 (1997).
10. G. A. Holdgate *et al.*, *Biochemistry* **36**, 9663 (1997).
11. R. J. Lewis *et al.*, *EMBO J.* **15**, 1412 (1996).
12. S.-M. Li, L. Heide, *Curr. Med. Chem.* **12**, 419 (2005).
13. C. Anderle *et al.*, *Antimicrob. Agents Chemother.* **52**, 1982 (2008).
14. R. H. Flatman, A. Eustaquio, S. M. Li, L. Heide, A. Maxwell, *Antimicrob. Agents Chemother.* **50**, 1136 (2006).
15. J. Schimana *et al.*, *J. Antibiot. (Tokyo)* **53**, 779 (2000).
16. U. Theobald, J. Schimana, H. P. Fiedler, *Antonie Leeuwenhoek* **78**, 307 (2000).
17. U. Galm *et al.*, *Arch. Microbiol.* **178**, 102 (2002).
18. A. Trefzer *et al.*, *Antimicrob. Agents Chemother.* **46**, 1174 (2002).
19. R. H. Flatman, A. J. Howells, L. Heide, H.-P. Fiedler, A. Maxwell, *Antimicrob. Agents Chemother.* **49**, 1093 (2005).
20. L. M. Oppegard *et al.*, *Antimicrob. Agents Chemother.* **53**, 2110 (2009).
21. J. H. Morais Cabral *et al.*, *Nature* **388**, 903 (1997).
22. N. Ruiz, B. Falcone, D. Kahne, T. J. Silhavy, *Cell* **121**, 307 (2005).
23. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
24. H. Yoshida, M. Bogaki, M. Nakamura, L. M. Yamanaka, S. Nakamura, *Antimicrob. Agents Chemother.* **35**, 1647 (1991).
25. H. Yoshida, M. Bogaki, M. Nakamura, S. Nakamura, *Antimicrob. Agents Chemother.* **34**, 1271 (1990).
26. D. C. Hooper, E. Rubinstein, *Quinolone Antimicrobial Agents* (American Society for Microbiology, Washington, DC, ed. 3, 1993), p. 485.
27. Materials and methods are available as supporting material on *Science* Online.
28. I. Laponogov *et al.*, *Nat. Struct. Mol. Biol.* **16**, 667 (2009).

29. This work was funded by UK Biotechnology and Biological Sciences Research Council (BBSRC) and the European Commission (CombiGyrase LSHB-CT-2004-503466); M.J.E. was supported by a CASE studentship funded by BBSRC and Plant Bioscience Limited, and T.B.K.L. was supported by a John Innes Centre Rotation studentship. We thank A. Zeeck (University of Göttingen) for providing us with a sample of MGD8N2A; P. Johnson for assistance with CD experiments; R. Field, L. Heide, and D. Hopwood for helpful discussions; and N. Ruiz and T. Silhavy for their advice and for the *E. coli imp4213* mutant. Atomic coordinates and structure factors have been deposited in the Protein Data Bank with accession code 2WL2. We dedicate this paper to the memory of Chris Lamb, director of the John Innes Centre from 1999 to 2009.

Supporting Online Material

www.sciencemag.org/cgi/content/full/326/5958/1415/DC1
Materials and Methods
SOM Text
Figs. S1 to S5
Tables S1 and S2

15 July 2009; accepted 8 October 2009
10.1126/science.1179123

Molecular Microbiology (2009) 72(6), 1462–1474 ■
doi:10.1111/j.1365-2958.2009.06735.x
First published online 22 May 2009

Coupling of the biosynthesis and export of the DNA gyrase inhibitor simocyclinone in *Streptomyces antibioticus*

Tung B. K. Le,[1]* Hans-Peter Fiedler,[2] Chris D. den Hengst,[1] Sang Kyun Ahn,[3] Anthony Maxwell[4] and Mark J. Buttner[1]
Departments of [1]Molecular Microbiology and [4]Biological Chemistry, John Innes Centre, Colney Lane, Norwich NR4 7UH, UK.
[2]Mikrobiologisches Institut, Eberhard-Karls-Universität Tübingen, Auf der Morgenstelle 28, D-72076 Tübingen, Germany.
[3]Department of Biochemistry and Biomedical Sciences, McMaster University, 1200 Main St. W., Hamilton, Ontario, Canada L8N 3Z5.

Summary

Because most antibiotics are potentially lethal to the producing organism, there must be mechanisms to ensure that the machinery responsible for export of the mature antibiotic is in place at the time of biosynthesis. Simocyclinone D8 is a potent DNA gyrase inhibitor produced by *Streptomyces antibioticus* Tü 6040. Within the simocyclinone biosynthetic cluster are two divergently transcribed genes, *simR* and *simX*, encoding proteins that resemble the TetR/TetA repressor–efflux pump pair that cause widespread resistance to clinically important tetracyclines. Engineered expression of *simX* from a strong, heterologous promoter conferred high level simocyclinone D8 resistance on *Streptomyces lividans*, showing that *simX* encodes a simocyclinone efflux pump. Transcription of *simX* is controlled by SimR, which directly represses the *simX* and *simR* promoters by binding to two operator sites in the *simX–simR* intergenic region. Simocyclinone D8 abolishes DNA binding by SimR, providing a mechanism that couples the biosynthesis of simocyclinone to its export. In addition, an intermediate in the biosynthetic pathway, simocyclinone C4, which is essentially inactive as a DNA gyrase inhibitor, also induces *simX* expression *in vivo* and relieves *simX* repression by SimR *in vitro*.

Accepted 5 May, 2009. *For correspondence. E-mail tung.le@bbsrc.ac.uk; Tel. (+44) 1603 450757; Fax (+44) 1603 450778.

Introduction

Aminocoumarin antibiotics are active against Gram-positive bacteria and function principally by inhibiting DNA gyrase (Maxwell and Lawson, 2003), an essential DNA topoisomerase found in all bacteria, which catalyses DNA supercoiling (Nollmann *et al*., 2007). In addition, a likely secondary target of these compounds is topoisomerase IV, which is involved in chromosome decatenation (Hardy and Cozzarelli, 2003; Oppegard *et al*., 2009). All four known aminocoumarins are produced by *Streptomyces* species. The first three to be discovered, novobiocin, clorobiocin and coumermycin A_1, each competitively inhibit the ATPase activity of the GyrB subunit of DNA gyrase and exhibit K_i values in the nanomolar range (Gormley *et al*., 1996). The most recently identified aminocoumarin antibiotic, simocyclinone D8 (Schimana *et al*., 2000) (Fig. 1), also inhibits DNA gyrase but was unexpectedly found to have a completely novel mode of action, binding instead to the GyrA subunit of the enzyme and preventing its binding to DNA (Flatman *et al*., 2005; M.J. Edwards *et al*., unpubl. data). Simocyclinone D8 is a potent inhibitor of DNA gyrase supercoiling with an IC_{50} lower than that of novobiocin (Flatman *et al*., 2005; Oppegard *et al*., 2009).

The genus *Streptomyces* accounts for the production of approximately two-thirds of the known antibiotics. They expel these compounds into their environment, typically the soil, most probably to give them a competitive advantage over other organisms that share the same ecological niche. Because the antibiotic is often potentially lethal to the producing organism, there must be mechanisms to ensure that the machinery responsible for export of the mature antibiotic is in place at the time of biosynthesis. This export machinery may be sufficient to confer resistance to the antibiotic, or there may be additional resistance mechanisms. For example, in the case of the novobiocin, clorobiocin and coumermycin A_1 producers, the principal mechanism of resistance is production of an aminocoumarin-resistant GyrB subunit ($GyrB^R$), encoded within the biosynthetic cluster and activated during antibiotic production (Thiara and Cundliffe, 1988; 1989; 1993; Schmutz *et al*., 2003). The $GyrB^R$ subunit replaces the sensitive subunit ($GyrB^s$) in the $(GyrA)_2(GyrB)_2$ heterotetramer during the production phase. In addition, an aminocoumarin-resistant topoisomerase IV subunit,

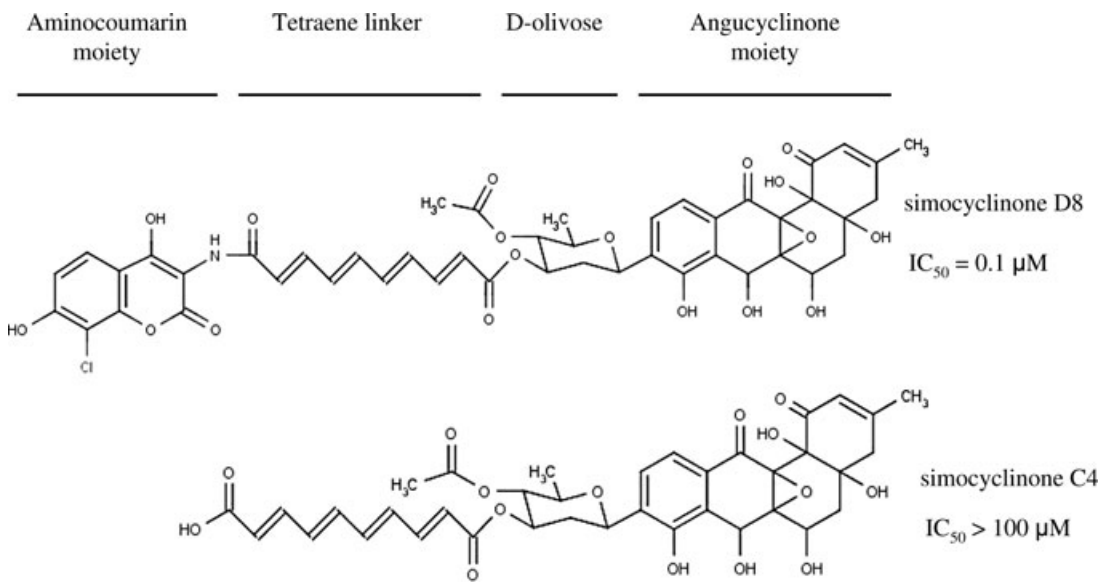


Fig. 1. Structures of simocyclinone D8 and its biosynthetic intermediate, simocyclinone C4, and their IC_{50}s for *Escherichia coli* DNA gyrase.

ParY^R, encoded within the clorobiocin and coumermycin A_1 biosynthetic clusters (but not the novobiocin biosynthetic cluster), also confers resistance to these antibiotics when introduced into a naïve host, presumably by an analogous mechanism (Schmutz *et al*., 2003). However, no DNA gyrase or topoisomerase IV subunits are encoded within the simocyclinone D8 (*sim*) biosynthetic cluster (Galm *et al*., 2002; Trefzer *et al*., 2002), leaving unknown the mechanism of resistance in the producing organism, *Streptomyces antibioticus* Tü 6040.

Simocyclinone D8 consists of four different parts, with a halogenated aminocoumarin at one end connected to an angucyclic polyketide at the other end via a tetraene linker and a D-olivose sugar (Galm *et al*., 2002; Trefzer *et al*., 2002) (Fig. 1). Within the *sim* cluster, among the genes responsible for the biosynthesis and linking of the four constituents of the antibiotic, are two divergently transcribed genes, *simR2* (hereafter, *simR*) and *simEX1* (hereafter, *simX*) (Galm *et al*., 2002; Trefzer *et al*., 2002). The SimR/SimX pair resembles the TetR/TetA repressor–efflux pump pair that causes widespread resistance to clinically important tetracyclines in several human pathogens (Chopra and Roberts, 2001). TetR represses transcription of the divergently transcribed *tetA* gene, encoding a proton-dependent tetracycline efflux pump. Binding of tetracycline to the C-terminal domain of TetR causes it to lose affinity for its operators, derepressing expression of *tetA*, which confers high-level resistance to the drug (Hillen and Berens, 1994; Kisker *et al*., 1995; Orth *et al*., 2000; Ramos *et al*., 2005). The similarity of SimR/SimX to TetR/TetA suggested they might be involved in simocyclinone efflux and, potentially, in simocyclinone resistance (Galm *et al*., 2002; Trefzer *et al*., 2002).

Here we show that *simX* encodes a simocyclinone efflux pump, and that transcription of *simX* is controlled by SimR, which directly represses the *simX* and *simR* promoters by binding to two operator sites in the *simR–simX* intergenic region. We show that simocyclinone D8 abolishes DNA binding by SimR, providing an intimate mechanism that couples the biosynthesis of simocyclinone to its export. In addition, we show that an intermediate in the biosynthetic pathway, simocyclinone C4 (Fig. 1), which is essentially inactive as a DNA gyrase inhibitor, also induces *simX* expression *in vivo* and relieves DNA binding by SimR *in vitro*, suggesting a potential 'feed-forward' mechanism (Tahlan *et al*., 2007) that might ensure expression of the SimX efflux pump prior to the build-up of a toxic concentration of the mature, active antibiotic.

Results

SimX encodes a simocyclinone efflux pump

There are two pump-like transmembrane proteins encoded in the *sim* cluster, SimEX1 (hereafter SimX) and SimEX2 (Trefzer *et al*., 2002). To determine if either of these two proteins is involved in simocyclinone D8 efflux, we expressed *simX* and *simEX2* from the strong constitutive promoter *ermEp** using the integrative, single-copy vector pIJ10257 (Hong *et al*., 2005). We introduced these constructs into the heterologous host *S. lividans* and compared the susceptibility of the resulting strains to simocyclinone D8. The strain carrying *ermEp*–simEX2* had an MIC of 2 μg ml^{-1}, as did *S. lividans* alone, or *S. lividans* containing the parent vector, pIJ10257. In contrast, the strain carrying the *ermEp*–simX* construct had an MIC of

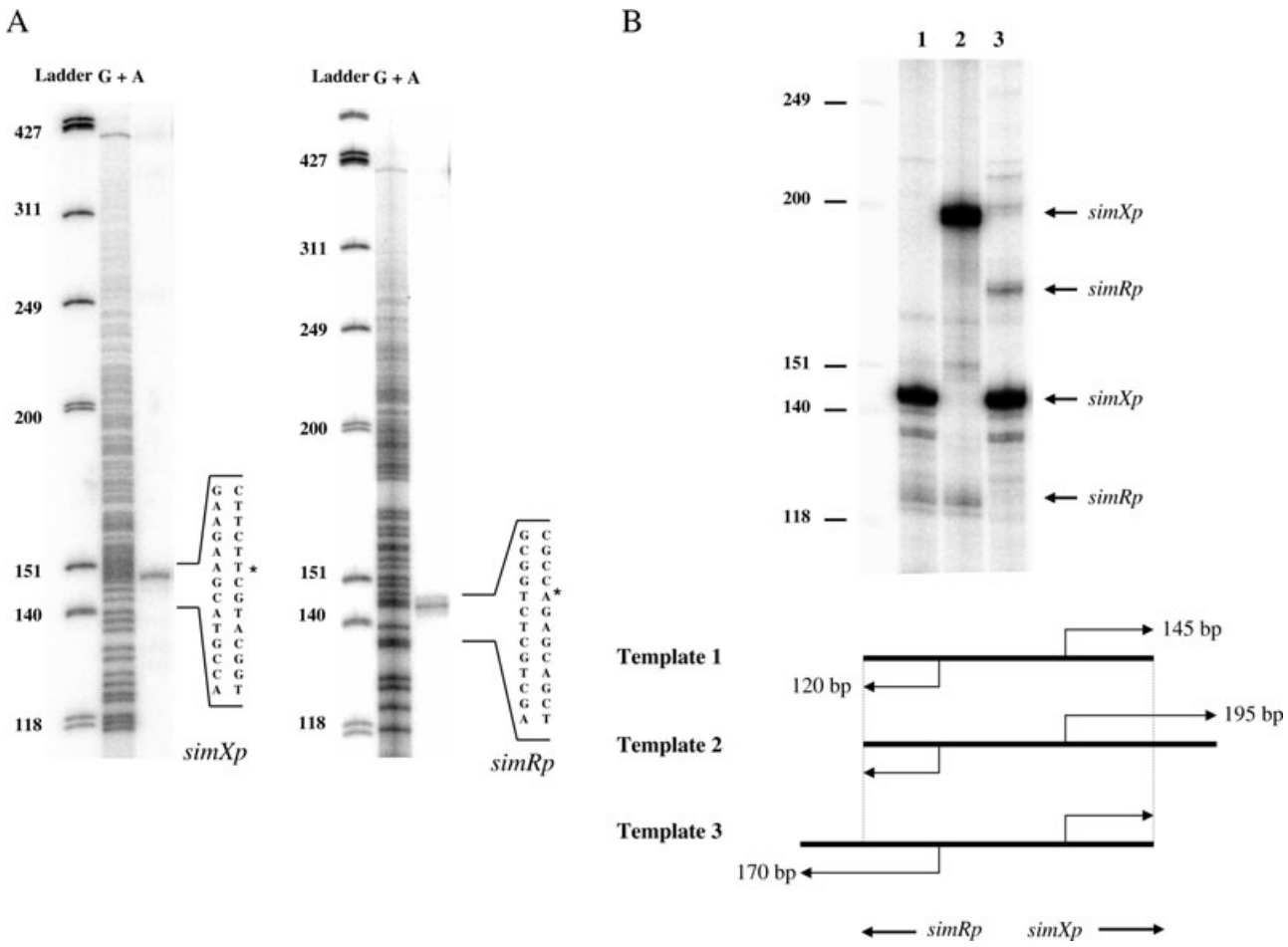


Fig. 2. A. High-resolution S1 nuclease mapping of the 5′ ends of the *simR* and *simX* transcripts using PCR-generated probes and RNA from the simocyclinone D8 producing organism, *S. antibioticus* Tü 6040. The most likely transcription start points are indicated by the asterisks. The G+A Maxam–Gilbert chemical sequencing ladder was generated using the same probe as the one used for S1 nuclease mapping assays. The size markers are a radiolabelled *Hin*fI digest of ΦX174 DNA.
B. Run-off transcription from the *simR* and *simX* promoters *in vitro* using purified *S. coelicolor* RNA polymerase and the templates illustrated containing the *simR–simX* intergenic region. Lane numbers correspond to the different templates shown in the illustration below.

65 μg ml^{-1}. The *ermEp*–simX* construct did not confer resistance to the structurally related aminocoumarin DNA gyrase inhibitor novobiocin, nor to unrelated antibiotics such as erythromycin, lincomycin, rifampicin, bacitracin and nisin. These results suggested that *simX* encodes a simocyclinone-specific efflux pump. Adjacent to *simX* is the divergent gene *simR*, encoding a potential transcriptional repressor of *simX*. We also cloned *simX* into an integrative vector under the control of its native promoter and in the absence of *simR*. This construct (pIJ10461) only mildly enhanced resistance to simocyclinone D8 in *S. lividans*, giving an MIC of 4 μg ml^{-1}.

Mapping the transcription start points of simR *and* simX

High-resolution S1 nuclease mapping of the *simR* and *simX* promoters was performed using RNA isolated from the simocyclinone D8 producing organism, *S. antibioticus* Tü 6040 (Fig. 2A). A single *simR* promoter (*simRp*) was identified, initiating transcription 20 bp upstream of the *simR* ATG start codon, and a single *simX* promoter (*simXp*) was identified, initiating transcription 47 bp upstream of the *simX* ATG start codon (Fig. 2A and 3). *In vitro* run-off transcription experiments with purified *S. coelicolor* RNA polymerase confirmed the presence and locations of the *simR* and *simX* promoters (Fig. 2B).

SimR regulates expression of simR *and* simX

To investigate the regulation of *simR* and *simX*, we measured *simR* and *simX* promoter activities in the presence and absence of SimR, using the integrative luciferase (*luxAB*) reporter plasmid pIJ5972 (Aigle *et al.*, 2000; M. Paget, pers. comm.). Fragments carrying *simRp* and *simXp* were individually cloned into pIJ5972, and the resulting reporter constructs were introduced into

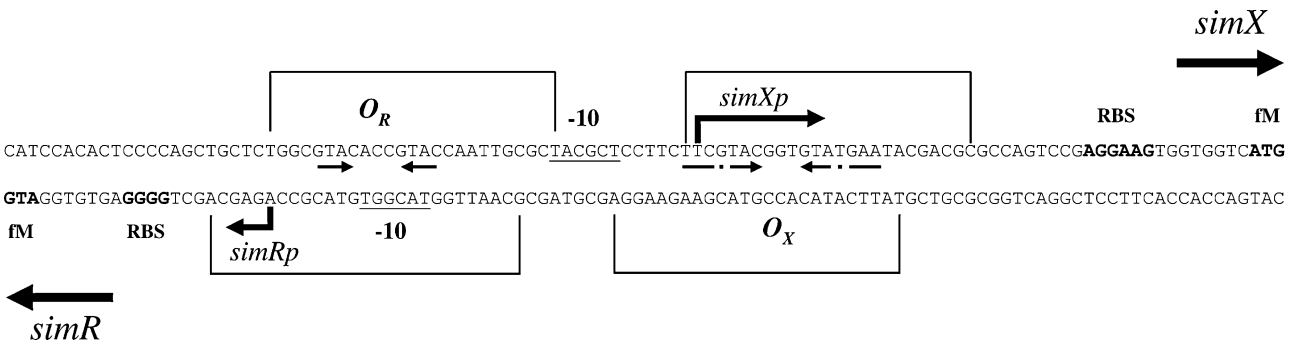


Fig. 3. Sequence of the *simR–simX* intergenic region showing the *simRp* and *simXp* transcription start points and putative −10 sequences, the *simR* and *simX* ribosome binding sites (RBS), the extent of the SimR DNaseI footprints on the O_X and O_R operators, and the imperfect inverted repeats within the footprints that may represent SimR binding motifs.

S. lividans in order to probe *simR* and *simX* promoter activities in the absence of SimR. To measure *simR* and *simX* promoter activities in the presence SimR, an integrative plasmid (pIJ10469) carrying *simR* under the control of its own promoter was introduced into the strains already harbouring the promoter-probe plasmids. Transformants were grown on Difco Nutrient Agar to promote vegetative growth and delay the formation of aerial hyphae, which may interfere with diffusion of the luciferase substrate (n-decanal) and with light emission. Figure 4 shows that *simX* and *simR* promoter activities were repressed 100-fold and 12-fold, respectively, in the presence of *simR*.

Purified SimR binds to the simR–simX *intergenic region at two distinct operator sites*

The *lux* reporter data suggested that SimR is a repressor that regulates its own expression as well as that of *simX*. To test this idea, we monitored SimR binding to the *simR–simX* intergenic region by electrophoretic mobility shift assay (EMSA). An N-terminally His_6-tagged derivative of SimR was overexpressed in *Escherichia coli* and purified to homogeneity. Increasing concentrations of SimR were incubated with a radiolabelled probe spanning the *simR–simX* intergenic region and the complexes were resolved on a native gel. Purified SimR bound to the intergenic region at concentrations as low as 0.23 nM and, as the concentration of SimR increased, two sets of shifted protein–DNA complexes became evident, suggesting that there are two SimR binding sites in the *simR–simX* intergenic region (Fig. 5A).

DNaseI footprinting on both DNA strands was used to map precisely the SimR operator sites within the *simR–simX* intergenic region. Two separate SimR binding sites were observed, consistent with the two shifted species seen in the EMSA experiments: the operator closer to *simX* was designated O_X and the one closer to *simR* was

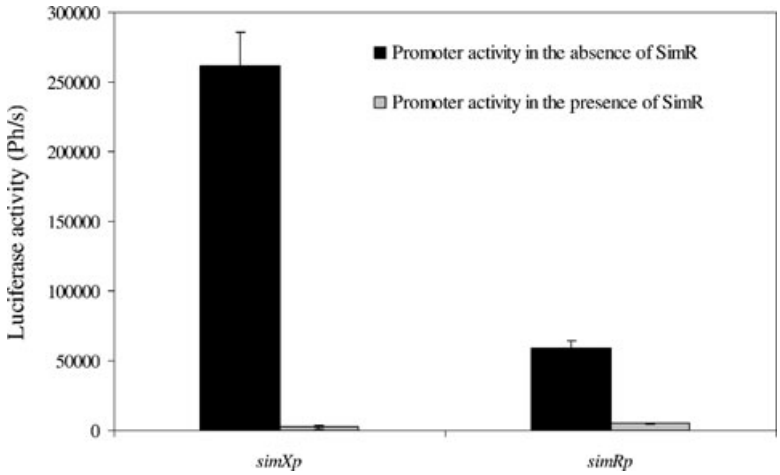


Fig. 4. Promoter activities of *simR* and *simX* in the presence and absence of *simR*. *simRp–luxAB* and *simXp–luxAB* transcriptional fusions were created in an integrative luciferase promoter-probe vector (pIJ5972) and assayed in *S. lividans*, in either the presence or absence of *simR*. Plasmid-containing *S. lividans* strains were grown on Difco Nutrient Agar in single wells of a 96-well microtitre plate (Sterilin) for 3 days. Each well was inoculated with approximately 5×10^4 spores. Plates were exposed to filter paper impregnated with n-decanal for 5 min and luciferase activities were quantified using a NightOwl camera (Berthold) equipped with WinLight software (Berthold) using a 1 min exposure time. Values given correspond to the average of three biological replicates from three different spore stocks and standard errors are shown.

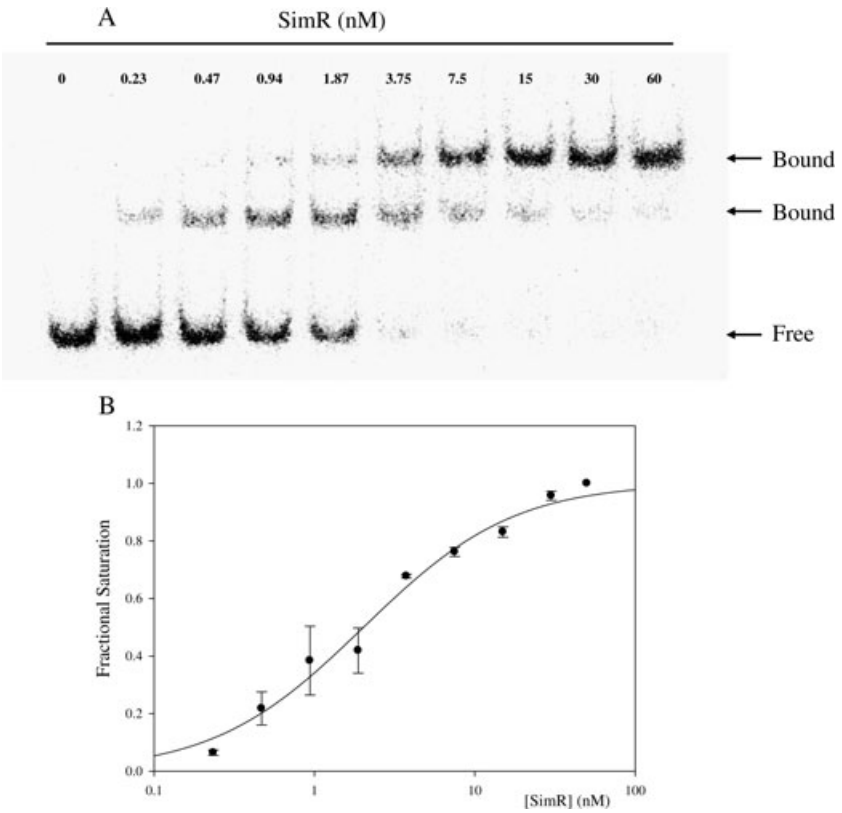


Fig. 5. A. Electrophoretic mobility shift assay showing binding of purified SimR to the *simR–simX* intergenic region. Bands corresponding to protein–DNA complexes (Bound) and free DNA (Free) are indicated. The final concentration of SimR is indicated above each lane.
B. Saturation curve of the data from EMSA experiments. EMSA data were collected and analysed on a PhosphorImager (FujiFilm) using Multi Gauge image analysis software (FujiFilm). Two independent EMSAs were carried out [one of which is shown in (A)] and the mean values calculated. Standard errors are shown. Saturation curves (saturation fraction against concentration of protein) were fitted with SigmaPlot (see *Experimental procedures*) to determine K_ds. In all EMSA experiments, SimR was present in molar excess over the probe.

designated O_R (Figs 3 and 6). Within the O_X and O_R footprints we identified imperfect inverted repeats that may represent the binding sequences for the SimR homodimer. Figure 3 shows these imperfect inverted repeats, the extent of the SimR DNaseI footprints on the sequence of the *simR–simX* intergenic region and the positions of the *simRp* and *simXp* transcription start points and putative −10 promoter sequences in relation to the two SimR operators.

SimR binding to the two operators is non-cooperative

In the DNaseI footprinting analysis, O_X was occupied at a lower concentration of SimR than was O_R, suggesting that O_X has a higher affinity for SimR than O_R. Competitive EMSA was used to explore this issue further. In these experiments, unlabelled fragments containing either O_X or O_R were used to compete with a radioactively labelled *simR–simX* intergenic fragment containing both O_X and O_R. The final concentration of SimR used in the competitive EMSA experiments was set at 20 nM such that all the labelled probes were in complex with SimR in the absence of unlabelled competitor DNA. A 1000-fold excess of unlabelled O_X-containing fragment outcompeted the labelled intergenic probe (no complex formation between SimR and the labelled intergenic probe) (Fig. 7). However, the same excess of O_R-containing fragment could not completely abolish SimR complex formation with the labelled intergenic probe (Fig. 7), confirming that SimR binds O_X more tightly than O_R.

Using the EMSA data shown in Fig. 5, we determined the approximate equilibrium dissociation constants (K_ds) for the two complexes to be 1.2 ± 0.4 nM for SimR-O_X and 3.5 ± 1.4 nM for SimR-O_R. In order to determine whether there is cooperativity between SimR binding at O_X and O_R, we also determined the K_ds of each SimR–operator complex by EMSA using probes containing only O_X or only O_R (data not shown), instead of the full *simR–simX* intergenic region. The K_ds were found to be 0.9 ± 0.2 nM for SimR-O_X and 3.6 ± 0.3 nM for SimR-O_R. The K_ds for the two SimR–operator complexes did not change substantially when the two operators were separated, suggesting that SimR binding to its two operators is non-cooperative.

Exogenous simocyclinone D8 induces expression of the SimX efflux pump in vivo

In order to determine whether the *simR* and *simX* promoters respond to simocyclinone D8 *in vivo*, we used the *S. lividans luxAB* reporter system to measure *simRp* and *simXp* activities in the presence of SimR and in response to exogenously added antibiotic. Figure 8A shows the response curve of *simR* and *simX* promoter activity to

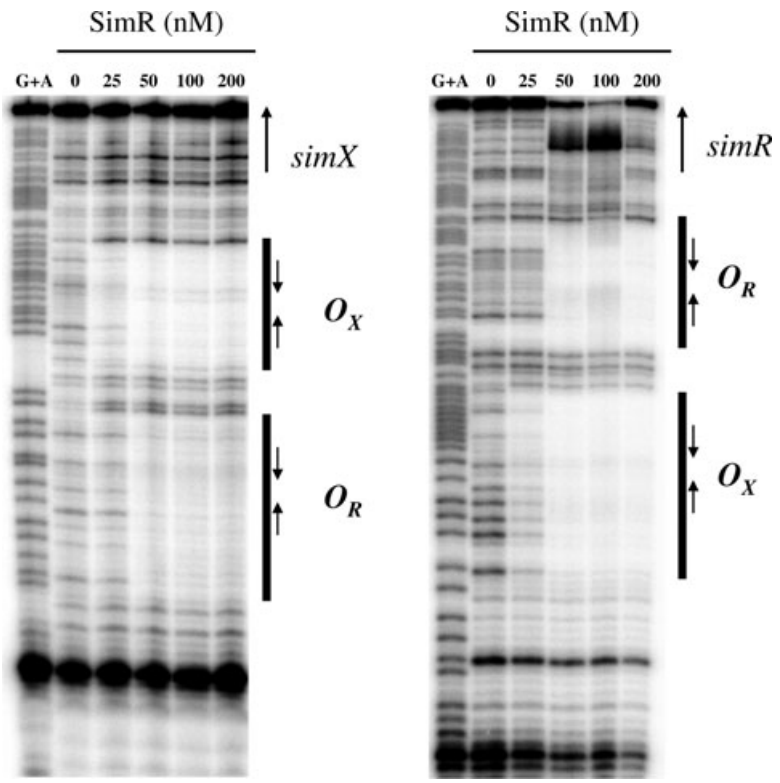


Fig. 6. DNaseI footprinting analysis of SimR binding to the *simR–simX* intergenic region. A DNA fragment containing the *simR–simX* intergenic region, 5′ end-labelled on either the upper strand (left panel) or the lower strand (right panel), was exposed to DNaseI in the presence of increasing concentrations of SimR. The sequencing ladders were generated by subjecting the probes to Maxam–Gilbert G+A chemical sequencing. Regions protected from DNaseI cleavage (operators O_X and O_R) are indicated with vertical bars and inverted repeats within the DNaseI protected regions are indicated by convergent arrows; these features are also highlighted on the DNA sequence in Fig. 3.

increasing concentrations of simocyclinone D8. Both promoters were induced by simocylinone D8, suggesting that the antibiotic can relieve SimR-mediated repression of both *simRp* and *simXp*. The highest concentration tested was 0.6 μg ml^{-1} because *S. lividans* is sensitive to simocylinone D8 (MIC = 2 μg ml^{-1}).

Simocyclinone D8 dissociates SimR from the simR–simX *intergenic region*

To determine if SimR responds directly to simocyclinone D8, we examined the effect of the antibiotic on SimR–operator complex formation by EMSA (Fig. 9A). The SimR concentration was held constant and an increasing concentration of simocyclinone D8 was introduced into the binding reaction. As the simocyclinone D8 concentration increased, there was a progressive decrease in SimR–DNA complex formation and a concomitant liberation of free probe; 62.5 μM simocyclinone D8 was sufficient to dissociate the SimR–DNA complexes almost completely. This effect was not due to DMSO, the simocyclinone D8 solvent, as equivalent amounts of pure DMSO had no effect on SimR–DNA complex formation (data not shown). Furthermore, to test specificity, we examined the effect of simocyclinone D8 on the DNA binding activity of ActR, a TetR homologue that regulates actinorhodin export in

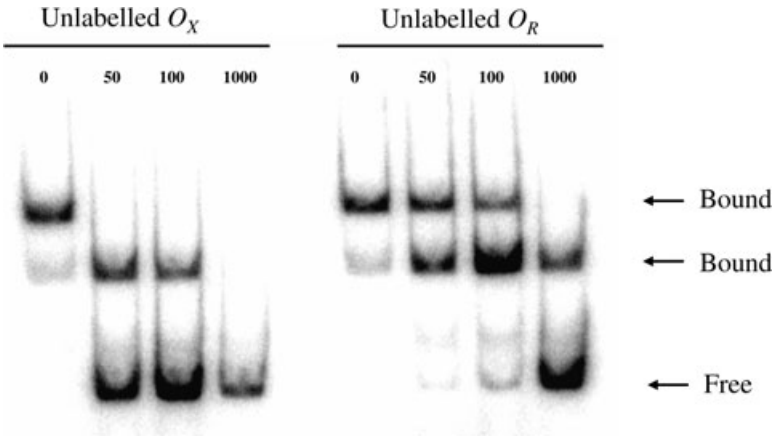


Fig. 7. Competitive electrophoretic mobility shift assay comparing the binding affinity of SimR to O_X and O_R. All lanes contain SimR at a final concentration of 20 nM and a constant amount of a radiolabelled *simR–simX* intergenic probe that carries both O_X and O_R. The fold excess (over the radiolabelled probe) of a competing unlabelled fragment containing either O_X or O_R is indicated above each lane.

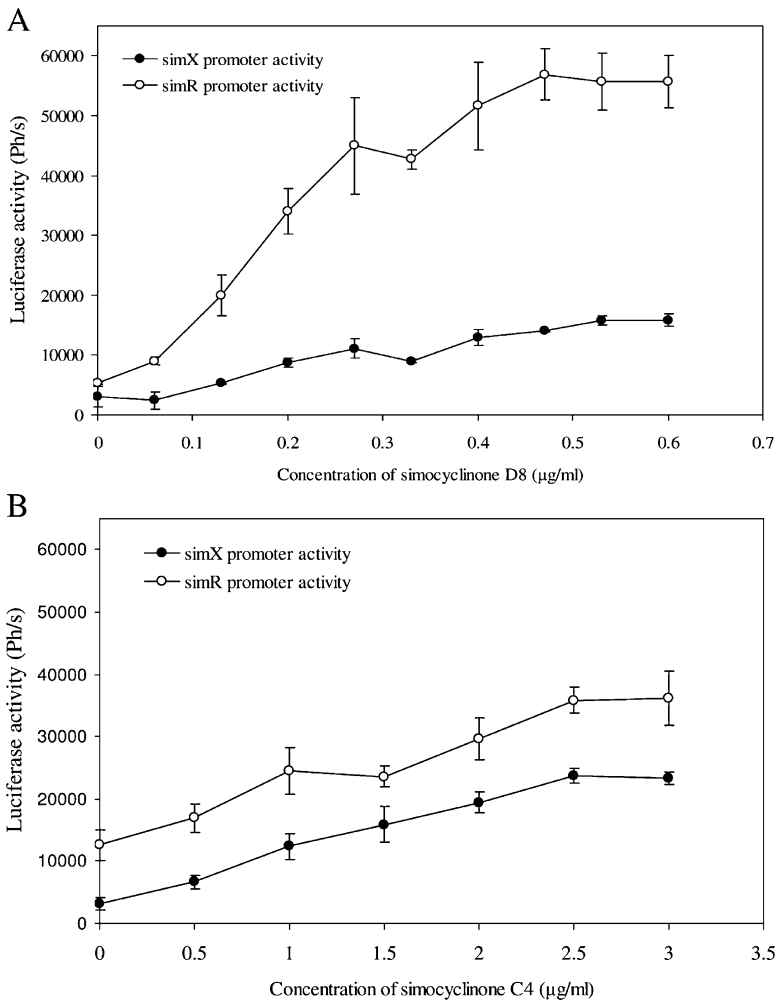


Fig. 8. Induction of the *simR* and *simX* promoters *in vivo* by (A) simocyclinone D8 and (B) the biosynthetic intermediate simocyclinone C4. *S. lividans* containing pIJ10469 (carrying *simR* under its native promoter) together with luciferase promoter-probe plasmids pIJ10465 (*simXp–luxAB*) or pIJ10466 (*simRp–luxAB*) were assayed. Values given correspond to the average of three biological replicates from three different spore stocks and standard errors are shown. For further details, see the legend of Fig. 4.

Streptomyces coelicolor (see *Discussion*) and found that 100 μM simocyclinone D8 had no effect on ActR binding to its cognate operator (data not shown). These results show that simocyclinone D8 is able to specifically disrupt SimR–DNA complex formation.

An intermediate in the simocyclinone biosynthetic pathway induces simX in vivo *and dissociates SimR from its operators* in vitro

In addition to simocyclinone D8, we also tested the ability of simocyclinone C4, a natural intermediate in the D8 biosynthetic pathway, to induce the *simR* and *simX* promoters *in vivo* and to dissociate SimR from its operators *in vitro*. Simocylinone C4 lacks the aminocoumarin ring present in the mature antibiotic (Fig. 1) and is essentially inactive as a DNA gyrase inhibitor; the simocyclinone D8 $IC_{50} = 0.1$ μM, whereas the simocyclinone C4 $IC_{50} > 100$ μM (M.J. Edwards *et al.*, unpubl. data). Simocyclinone C4 induced the *simR* and *simX* promoters *in vivo*, although somewhat more weakly than the mature antibiotic (Fig. 8B). Because simocyclinone C4 is not an active antibiotic, we were able to test higher concentrations than for simocyclinone D8. Consistent with the *in vivo* inductions, simocyclinone C4 also caused SimR to dissociate from its binding sites *in vitro*, although again somewhat more weakly than simocyclinone D8; 250 μM simocyclinone C4 was sufficient to abolish all complex formation (Fig. 9B).

Discussion

This report shows that *simX* encodes a simocyclinone D8 efflux pump and that the *simX* promoter is directly repressed by SimR, which binds to two operator sites in the *simR–simX* intergenic region. Simocyclinone D8 abolishes DNA binding by SimR, coupling the biosynthesis of simocyclinone to its export.

Simocyclinone D8 is a potent inhibitor of DNA gyrase supercoiling ($IC_{50} = 0.1$ μM) (Flatman *et al.*, 2005; Oppegard *et al.*, 2009; M.J. Edwards *et al.*, unpubl. data). The recent structure of the antibiotic bound to the DNA gyrase

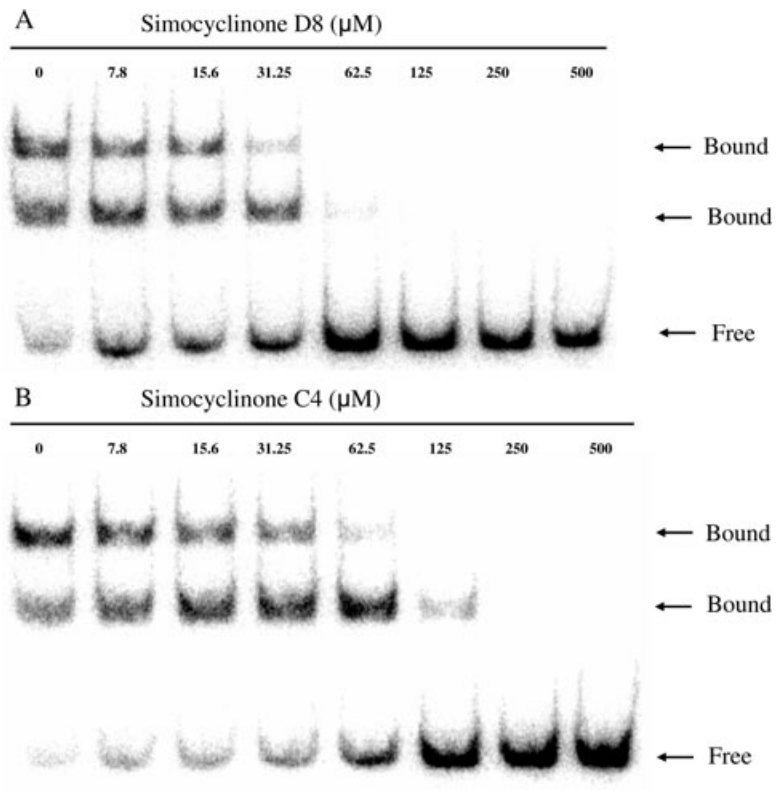


Fig. 9. Electrophoretic mobility shift assays showing that (A) simocyclinone D8 and (B) its biosynthetic intermediate simocyclinone C4 abolish SimR DNA binding activity. All lanes contain a constant amount of SimR and radiolabelled *simR–simX* intergenic fragment containing both O_X and O_R. Bands corresponding to protein–DNA complexes (Bound) and free DNA (Free) are indicated. The final concentrations of simocyclinone D8 or simocyclinone C4 are indicated above each lane.

A subunit shows that the two moieties at the ends of simocyclinone D8, the aminocoumarin ring and the angucyclic polyketide, bind to two separate, well-defined pockets within the GyrA DNA binding saddle, linked by the intervening tetraene linker and D-olivose moieties (M.J. Edwards *et al.*, unpubl. data). Given the prominence of the aminocoumarin ring in the overall binding of the antibiotic to the GyrA subunit, it is not surprising that the biosynthetic intermediate simocyclinone C4, which lacks the aminocoumarin ring (Fig. 1), is essentially inactive as a DNA gyrase inhibitor ($IC_{50} > 100$ μM). Importantly, however, although simocyclinone C4 is not a DNA gyrase inhibitor, it can efficiently derepress SimR *in vivo* and *in vitro*.

The most striking analogy to the research presented here is the work of Nodwell and colleagues on the regulation of efflux of actinorhodin (Ahn *et al.*, 2007; Hopwood, 2007; Tahlan *et al.*, 2007; 2008; Willems *et al.*, 2008). Actinorhodin is a blue-pigmented, six-ring polyketide antibiotic made by *S. coelicolor* (Bystrykh *et al.*, 1996). Within the actinorhodin biosynthetic cluster are two co-transcribed genes, *actA* and *act*II-ORF3, encoding integral membrane proteins implicated in actinorhodin export (Caballero *et al.*, 1991; Fernandez-Moreno *et al.*, 1991). Expression of these two genes is regulated by a TetR-like protein, ActR, the product of the adjacent, divergently transcribed gene. Nodwell and colleagues have been characterizing ligands that relieve repression by ActR. Importantly, they showed that, in addition to the mature six-ring antibiotic, three-ring intermediates from the biosynthetic pathway also relieve repression by ActR. From this, they suggested that the ability of actinorhodin intermediates to relieve repression by ActR might provide a 'feed-forward' mechanism that would ensure expression of the ActA efflux pump prior to the build-up of a toxic concentration of the mature antibiotic (Ahn *et al.*, 2007; Hopwood, 2007; Tahlan *et al.*, 2007; 2008; Willems *et al.*, 2008). Similarly, the ability of an inactive simocyclinone intermediate to relieve repression by SimR might also act as a biosynthetic checkpoint to ensure feed-forward regulation of simocyclinone export.

As applied to simocyclinone (or actinorhodin) export, the feed-forward hypothesis is speculative. We have shown that the pathway intermediate simocyclinone C4 induces *simX* expression *in vivo* when applied exogenously. However, for a feed-forward mechanism to operate in the producing organism, simocyclinone C4 or other SimR-binding pathway intermediates would have to accumulate in the cytoplasm to a concentration high enough to trigger *simX* expression, and the cytoplasmic concentrations of pathway intermediates are hard to determine experimentally. However, there are other examples in the literature where the TetR-like protein blocks expression not only of the exporter gene, but also of late biosynthetic genes, and in these cases it seems that an intermediate in the pathway must be responsible for inducing expression of these genes (i.e. feed-forward

activation), as induction of the late biosynthetic enzymes is required to generate the mature antibiotic.

In the landomycin A producer, *Streptomyces cyanogenus*, a TetR-family regulator, LanK, represses expression of *lanJ*, encoding a putative landomycin A efflux pump (Ostash *et al*., 2008). LanK is derepressed by mature landomycin A, which carries a hexasaccharide chain, but also by intermediates in the pathway that carry only a pentasaccharide or a trisaccharide chain (Ostash *et al*., 2008). However, *lanJ* is co-transcribed with several downstream biosynthetic genes involved in late glycosylation steps. As a consequence, LanK couples production of intermediates not only to assembly of the export machinery, but also to expression of late biosynthetic enzymes that attach the final sugars to produce mature landomycin A (Ostash *et al*., 2008). A more complex example concerns the biosynthesis of the clinically important anticancer agents daunorubicin and doxorubicin made by *Streptomyces peucetius*. In this system, binding of daunorubicin–doxorubicin pathway intermediates like rhodomycin D appears to derepress the TetR-like regulator DnrO, activating a cascade involving two further transcription factors, DnrN and DnrI, which leads to expression of the resistance genes and of late biosynthetic genes (Otten *et al*., 1995; Jiang and Hutchinson, 2006).

In addition to repressing the *simX* promoter, SimR also directly negatively regulates its own expression. In other systems, negative autoregulation has been shown to confer specific functions that are absent in systems that have simple regulation. Many of these studies on the design principles of genetic circuits that incorporate negative autoregulation have exploited TetR, the founding member of the family to which SimR belongs (Becskei and Serrano, 2000; Rosenfeld *et al*., 2002; D. Madar and U. Alon, pers. comm.). These studies show that negative autoregulatory feedback loops in gene circuits can provide stability, thereby limiting stochastic fluctuations in the system (Becskei and Serrano, 2000). In addition, negative autoregulation can speed the response time of the transcription network to a stimulus (Rosenfeld *et al*., 2002), and it can broaden the dynamic range of the input signal to which the downstream genes respond (D. Madar and U. Alon, pers. comm.).

An unresolved issue is the basis of resistance to simocyclinone D8 in the producing organism. In the streptomycetes that produce the aminocoumarins novobiocin (*Streptomyces sphaeroides*), clorobiocin (*Streptomyces roseochromogenes*) and coumermycin A_1 (*Streptomyces rishiriensis*), expression of an aminocoumarin-resistant $GyrB^R$ gyrase subunit encoded within the biosynthetic gene cluster is turned on during antibiotic production, conferring resistance (Thiara and Cundliffe, 1988; 1989; 1993; Schmutz *et al*., 2003). The absence of an equivalent *gyrB* resistance gene in the simocyclinone biosynthetic gene cluster was an initial surprise when the *sim* locus was sequenced, until it was subsequently and unexpectedly shown that simocyclinone has a completely different mode of action, binding instead to the GyrA subunit of the enzyme (Flatman *et al*., 2005; M.J. Edwards *et al*., unpubl. data). However, no $GyrA^R$ subunit is encoded within the *sim* cluster either (Galm *et al*., 2002; Trefzer *et al*., 2002). Expressing *simX* from the strong, constitutive promoter, *ermEp**, conferred simocyclinone resistance on the heterologous host *S. lividans* (MIC 65 μg ml^{-1} instead of 2 μg ml^{-1} for *S. lividans* alone), showing that SimX can act as an effective resistance determinant when expressed at appropriate levels. However, the MIC was only 4 μg ml^{-1} when *simX* was expressed from its own promoter in the absence of SimR. This may reflect additional levels of *simX* regulation that have yet to be uncovered. For example, in addition to its repression by SimR, it is possible that *simX* expression in *S. antibioticus* might also be activated by a simocyclinone pathway-specific activator that is absent from *S. lividans*. Alternatively, it is possible that SimX-mediated efflux is not the principal simocyclinone resistance mechanism in *S. antibioticus*.

Experimental procedures

Bacterial strains, plasmids, oligonucleotides, culture conditions and conjugal plasmid transfer from E. coli to Streptomyces *spp.*

Strains, plasmids and oligonucleotides are described in Table 1. Unless stated otherwise, *S. lividans* 1326 was grown on SFM (Kieser *et al*., 2000) and *S. antibioticus* Tü 6040 was grown on MYMTap [0.4% (w/v) maltose, 0.4% (w/v) Yeast Extract, 1% (w/v) Malt Extract, supplemented trace elements; made up in tap water]. For conjugal transfer into *Streptomyces* (Paget *et al*., 1999), plasmids were introduced into the *dam dcm hsdS E. coli* strain ET12567, which carries the non-transmissible, *oriT*-mobilizing plasmid pUZ8002 (Paget *et al*., 1999). All cloned PCR fragments were verified by sequencing.

Fermentation, isolation and purification of simocyclinone D8 and simocyclinone C4

Simocyclinone D8 was isolated as described by Schimana *et al*. (2000). Briefly, *S. antibioticus* Tü 6040 was fermented in a complex medium consisting of 2% (w/v) mannitol and 2% (w/v) soybean meal in a 20 l fermentor, and simocyclinones were extracted from the mycelium with methanol. Pure simocyclinone D8 was obtained after reversed-phase HPLC using Nucleosil-100 C-18 material and 0.01% trifluoroacetic acid-acetonitrile gradient elution, resulting in a dark yellow powder after drying. Pure simocyclinone C4 was isolated using essentially the same procedure, but fermentation was carried out in a defined medium containing 20% (v/v) glycerol and 0.15% (w/v) L-arginine to maximize simocyclinone C4

Table 1. Bacterial strains, plasmids and primers used in this study.

	Relevant genotype/comments	Source/reference
Strains		
Streptomyces		
S. antibioticus Tü 6040	Environmental isolate; original producer of simocyclinone D8	Schimana *et al.* (2000)
S. lividans 1326	Wild-type SLP2$^+$ SLP3$^+$	Kieser *et al.* (2000)
E. coli		
ET12567 (pUZ8002)	ET12567 containing helper plasmid pUZ8002	Paget *et al.* (1999)
Plasmids		
pSET152	ΦC311 *attP-int* derived integration vector for the conjugal transfer of DNA from *E. coli* to *Streptomyces* (Apr^R)	Bierman *et al.* (1992)
pMS82	ΦBT1 *attP-int* derived integration vector for the conjugal transfer of DNA from *E. coli* to *Streptomyces* (Hyg^R)	Gregory *et al.* (2003)
pIJ5972	Integrative *Streptomyces* promoter-probe plasmid based on TTA codon-free derivatives of the *luxAB* reporter genes	Aigle *et al.* (2000); M. Paget, pers. comm.
pIJ10257	Integrative expression vector based on the strong, constitutive *ermE** promoter (*ermEp**)	Hong *et al.* (2005)
pIJ10461	pSET152 carrying *simX* under the control of its own promoter, without *simR*	This study
pIJ10465	pIJ5972 *simXp–luxAB*	This study
pIJ10466	pIJ5972 *simRp–luxAB*	This study
pIJ10469	pMS82 carrying *simR* under the control of its native promoter	This study
pIJ10480	pIJ10257 *ermEp*–simX*	This study
pIJ10481	pIJ10257 *ermEp*–simEX2*	This study
pIJ10490	pET28a derivative expressing His_6-tagged SimR	This study
Primers		
pX-F-EcoRI	GAATTCGAGCACGAACTCCTGCTGGC	
pX-R-BamHI	GGATCCGACCACCACTTCCTCGGACTGG	
pR-F-BamHI	GGATCCTCCCCAGCTGCTCTGGCGTACACC	
pR-R-EcoRI	GAATTCGTGAACGTACCGACCATCAGGCCG	
MS82-simR-F-HindIII	GGCAAGCTTTCAAGCCAGTGCTGGACGTTCC	
MS82-simR-R-KpnI	AACGGTACCAACGGCATCCTCATCTGGCATGACC	
intRX-138-F	AAAGATATCCTCGTTCATCCACACTCCCC	
intRX-138-R	AAAGGATCCATCTGGCATGACCACCACTTC	
intOX-123-F	CCAATTGCGCTACGCTCCTTC	
intOX-123-R	CCTGCGCGGAGCCTCCGGAC	
intOR-130-F	CACCCTCGGTGTCGGCCACC	
intOR-130-R	AACGAGAACGAACCCGTCAG	
simEX1-F-NdeI	GAGCATATGCCAGATGAGGATGCCGTTGC	
simEX1-R-HindIII	TAGAAGCTTCTATCCGGCATTCCGAGCCG	
simEX2-F-NdeI	GGGCATATGACCAGTTTCCAAGTCCAG	
simEX2-R-HindIII	GGGAAGCTTACCTCCCGGCCGCAGTAGACC	
simR-int-simX-R	CTATCCGGCATTCCGAGCCG	
S1-probeX-F	GTAGAGGGACATCGTGCCGGC	
S1-probeX-R	GGCCGAGCAGTACGGCCAGC	
S1-probeR-F	GTGTCGGCCACCTTGACGGC	
S1-probeR-R	CCACCGAGCTCTCCGACGATCG	
Invitro1-F	GAACGATCTGGTCACGGCTC	
Invitro1-R	AGCAGTACGGCCAGCACCCC	
Invitro2-R	GCAGCGCCGTACCGACGATCACC	
Invitro3-F	CATCGACGCCGCCTCGACACCC	

production (Theobald *et al.*, 2000). The nature and purity of the simocyclinone D8 and C4 samples were confirmed by nuclear magnetic resonance spectroscopy and liquid chromatography-mass spectrometry (Holzenkämpfer *et al.*, 2002; M.J. Edwards, pers. comm.)

Construction of simX *and* simEX2 *expression plasmids and minimal inhibitory concentration (MIC) determinations*

simX and *simEX2* were amplified by PCR using primers carrying NdeI (upstream) and HindIII (downstream) sites [for *simX*, simEX1-F-NdeI and simEX1-R-HindIII; for *simEX2*, simEX2-F-NdeI and simEX2-R-HindIII (Table 1)]. To express the *simX* and *simEX2* genes from the *ermEp** promoter, the fragments were cloned into NdeI-HindIII-cut pIJ10257 (Table 1) to generate pIJ10480 and pIJ10481 respectively. To express *simX* from its own promoter, a fragment containing the *simX* coding sequence and 122 bp of DNA upstream of the *simX* translation start codon was amplified by PCR using the primers intRX-138-F and simR-int-simX-R (Table 1), and cloned into the SmaI site of pUC19. The resulting construct was digested with BamHI and EcoRI and the insert was cloned into BamHI-EcoRI-cut pSET152 to give pIJ10461. Constructs were transferred into *S. lividans* by conjugation.

For MIC determinations, approximately 5×10^4 spores were added to each well of a 96-well microtitre plate (Sterilin), containing a twofold simocyclinone D8 dilution series in SMMS medium (Kieser *et al*., 2000). The plates were evaluated everyday for 5 days after inoculation, and the MIC was defined as the lowest concentration at which no growth was observed. Simocyclinone D8 was dissolved in DMSO with the final concentration of DMSO in the agar not to exceed 1% (v/v).

Construction of luxAB *reporter plasmids and luciferase activity measurements*

To probe *simXp* and *simRp* activities in the absence of SimR, the promoter regions (500 bp upstream of the translation start codons) were amplified by PCR using upstream primers carrying an EcoRI site and downstream primers carrying a BamHI site [primers pX-F-EcoRI and pX-R-BamHI for *simXp*; pR-F-BamHI and pR-R-EcoRI for *simRp* (Table 1)]. The *simXp* and *simRp* promoter fragments were cloned into EcoRI-BamHI-cut pIJ5972, an integrative, *Streptomyces* promoter-probe plasmid based on TTA codon-free derivatives of the *luxAB* reporter genes (Aigle *et al*., 2000; M. Paget, pers. comm.), to create plasmids pIJ10465 and pIJ10466, respectively, and these reporter constructs were transferred by conjugation into *S. lividans*. To probe *simXp* and *simRp* activities in the presence of SimR, a plasmid carrying *simR* under the control of its own promoter (pIJ10469) was transferred by conjugation into strains already harbouring the promoter-probe plasmids pIJ0465 and pIJ10466. To construct pIJ10469, *simR* with its promoter was amplified by PCR using an upstream primer carrying a HindIII site and a downstream primer carrying a KpnI site [primers MS82-simR-F-HindIII and MS82-simR-R-KpnI (Table 1)] and the fragment was cloned into HindIII-KpnI-cut pMS82.

Plasmid-containing strains were grown on Difco Nutrient Agar in single wells of a 96-well microtitre plate (Sterilin) for 3 days. Each well was inoculated with approximately 5×10^4 spores. Plates were exposed to filter paper impregnated with n-decanal for 5 min and luciferase activities were quantified using a NightOwl camera (Berthold) equipped with WinLight software (Berthold) using a 1 min exposure time. Values given correspond to the average of three biological replicates from three different spore stocks.

Overexpression and purification of SimR

The *simR* gene was chemically synthesized with codon optimization (Genescript) for expression in *E. coli* and ligated into pET28a (Novagen) to give pIJ10490, which was introduced into *E. coli* BL21 (DE3) pLys. Recombinant His_6-SimR was purified from a 250 ml LB culture induced with a 0.5 mM final concentration of IPTG at 30°C for 4 h. The cell pellet was re-suspended in 20 ml Buffer A [10 mM Tris-HCl, pH 7.8, 150 mM NaCl, 50 mM imidazole, 1× Complete Mini, EDTA-free protease inhibitor (Roche)], sonicated (20 s at 10 microns for three cycles), and the lysate was clarified by centrifugation. The filtered cell lysate was loaded onto a 1 ml Ni-loaded Hi-Trap Chelating HD column (GE Healthcare) and His-tagged SimR was eluted with Buffer B [10 mM Tris-HCl, pH 7.8, 150 mM NaCl, 500 mM imidazole] and dialysed into Storage Buffer [10 mM Tris-HCl, pH 7.8, 150 mM NaCl, 10% (v/v) glycerol]. The resulting preparation of SimR was 98% pure as judged by SDS-PAGE and was stored at −80°C.

Electrophoretic mobility shift assays and determination of dissociation constants (K_ds)

The EMSA DNA probe spanning the entire *simR–simX* intergenic region was amplified by PCR using primers intRX-138-F and intRX-138-R (Table 1) and then 5′ end-labelled using [γ^{32}-P]-ATP and T4 polynucleotide kinase (New England Biolabs). The competitor DNA carrying only O_R was amplified using primers intOR-130-F and intOR-130-R, and the competitor DNA carrying only O_X was amplified using primers intOX-123-F and intOX-123-R (Table 1). Binding of SimR to DNA was carried out in 20 µl EMSA buffer [20 mM Tris, pH 8.0, 1 µg poly(dI-dC), 1 mM EDTA, 100 mM NaCl, 0.5 mM DTT, 8% (v/v) glycerol] containing 0.1 nM radiolabelled DNA (approximately 8000 c.p.m.) and varying amounts of SimR. After incubation at 30°C for 5 min, the binding reaction mixtures were loaded on 5% (w/v) native polyacrylamide gels and run in TBE buffer at 120 V for 45 min. The effect of simocyclinone D8 and simocyclinone C4 on the ability of SimR to bind DNA was tested by adding the compounds to the EMSA buffer. The simocyclinone D8 and simocyclinone C4 stock solutions were made up in 100% DMSO.

EMSA data were collected and analysed on a PhosphorImager (FujiFilm) using Multi Gauge image analysis software (FujiFilm). Two independent EMSAs were carried out for each probe and mean values were calculated. In order to calculate K_ds for SimR binding to each operator when the two operator sites were separated, saturation curves (percentage of probe bound against concentration of protein) were fitted using SigmaPlot (one site saturation model). For K_ds of each operator when the two sites were coupled, a random-order binding model was used, where $Y = [S]/(2(K_{d1} + [S])) + [S]/(2(K_{d2} + [S]))$, in which Y is the fractional saturation, K_{d1} and K_{d2} are dissociation constant of SimR binding to O_X and O_R operators, respectively, and $[S]$ is the concentration of protein. The fractional saturation was calculated from intensities of EMSA bands in each lane (Fig. 5A) as followed, $Y = (0.5 \star \text{intensity of middle band} + \text{intensity of top band})/(\text{intensity of all bands in a lane})$. The equation was then fitted using SigmaPlot to determine K_ds.

DNaseI footprinting

Templates for DNaseI footprinting were amplified by PCR using one unlabelled primer and one primer 5′ end-labelled using [γ^{32}-P]-ATP and T4 polynucleotide kinase (New England Biolabs). The primers were intRX-138-F and intRX-138-R (Table 1), the same pair used to generate the *simR–simX* intergenic region probe for the EMSA experiments. DNaseI footprinting assays were performed in 40 µl EMSA buffer containing approximately 180 000 c.p.m. radiolabelled DNA and varying amounts of SimR. After incubation at room temperature for 5 min, 10 µl of DNaseI (10 units in 10 mM $CaCl_2$) was added and the incubation was continued for a

further 60 s. Reactions were stopped by adding 140 µl DNaseI stop solution (200 mM unbuffered sodium acetate, 30 mM EDTA, 0.15% SDS and 0.1 mg ml^{-1} yeast tRNA), the samples were precipitated with ethanol, and the pellets were dried and dissolved in 5 µl Sequencing Loading Dye [80% (v/v) formamide, 10 mM NaOH, 1 mM EDTA, 0.1% (w/v) xylene cyanol and 0.1% (w/v) bromophenol blue]. After heating at 80°C for 3 min and cooling on ice, the samples were run on a 6% (w/v) polyacrylamide/8 M urea sequencing gel, which was dried and analysed using a PhosphorImager (FujiFilm). A G+A sequencing ladder was generated from the template DNA by chemical sequencing (Maxam and Gilbert, 1980).

RNA preparation, S1 nuclease protection analysis and in vitro *run-off transcription*

For RNA preparation, approximately 10^9 *S. antibioticus* spores were germinated by heat-shock treatment in 5 ml TES buffer (0.05 M, pH 8) at 50°C for 10 min, then diluted with an equal volume of double-strength germination medium [1% (w/v) Difco yeast extract, 1% (w/v) Difco casaminoacids, 0.01 M $CaCl_2$] and incubated with shaking at 37°C for 6 h (modified from Kieser *et al*., 2000). Germinated spores were inoculated into NMMP (Kieser *et al*., 2000) and incubated with shaking for a further 15 h at 30°C. RNA was prepared essentially as described by Hesketh *et al*. (2007), but with minor modifications taken from the Qiagen RNA Extraction Kit procedure (Qiagen). Probes for S1 nuclease protection analysis were generated by PCR using a 5′ end-labelled oligonucleotide internal to the ORF and an unlabelled upstream primer [for *simXp*, primers S1-probeX-F and S1-probeX-R; for *simRp*, primers S1-probeR-F and S1-probeR-R; (Table 1)]. Hybridizations were carried out in sodium trichloroacetic acid buffer at 45°C overnight after denaturation at 65°C for 10 min (Kieser *et al*., 2000). A G+A sequencing ladder was generated by chemical sequencing (Maxam and Gilbert, 1980).

RNA polymerase containing a mixture of sigma factors was purified as described previously (Buttner *et al*., 1988) from *S. coelicolor* M600 grown to exponential phase in YEME (Kieser *et al*., 2000). *In vitro* transcription was carried out as described previously (Buttner *et al*., 1987). Three different templates for *in vitro* run-off transcription were generated by PCR that differed in their left or right ends, to allow the *simR* and *simX* transcripts to be identified unambiguously. Templates were generated using the following primers: Template 1, Invitro1-F and Invitro1-R; Template 2, Invitro1-F and Invitro2-R; Template 3, Invitro3-F and Invitro1-R (Table 1).

Acknowledgements

We thank Ray Dixon, Justin Nodwell, David Hopwood and Daniel Madar for their comments on the manuscript, Richard Little, Nick Tucker, Stephen Bornemann and Maureen Bibb for practical help and advice, Luis Servín-González for the gift of *S. coelicolor* RNA polymerase, and Andreas Bechthold for providing us with cosmids covering the *sim* cluster. This work was funded by a John Innes Centre Rotation Studentship (to T.B.K.L.) and by a grant-in-aid to the John Innes Centre from the BBSRC.

References

Ahn, S.K., and Tahlan, K., Yu, Z., and Nodwell, J. (2007) Investigation of transcription repression and small-molecule responsiveness by TetR-like transcription factors using a heterologous *Escherichia coli*-based assay. *J Bacteriol* **189:** 6655–6664.

Aigle, B., Wietzorrek, A., Takano, E., and Bibb, M.J. (2000) A single amino acid substitution in region 1.2 of the principal sigma factor of *Streptomyces coelicolor* A3(2) results in pleiotropic loss of antibiotic production. *Mol Microbiol* **37:** 995–1004.

Becskei, A., and Serrano, L. (2000) Engineering stability in gene networks by autoregulation. *Nature* **405:** 590–593.

Bierman, M., Logan, R., O'Brien, K., Seno, E.T., Rao, R.N., and Schoner, B.E. (1992) Plasmid cloning vectors for the conjugal transfer of DNA from *Escherichia coli* to *Streptomyces* spp. *Gene* **116:** 43–49.

Buttner, M.J., Fearnley, I.M., and Bibb, M.J. (1987) The agarase gene (*dagA*) of *Streptomyces coelicolor* A3(2): nucleotide sequence and transcriptional analysis. *Mol Gen Genet* **209:** 101–109.

Buttner, M.J., Smith, A.M., and Bibb, M.J. (1988) At least three different RNA polymerase holoenzymes direct transcription of the agarase gene (*dagA*) of *Streptomyces coelicolor* A3(2). *Cell* **52:** 599–607.

Bystrykh, L.V., Fernandez-Moreno, M.A., Herrema, J.K., Malpartida, F., Hopwood, D.A., and Dijkhuizen, L. (1996) Production of actinorhodin-related 'blue pigments' by *Streptomyces coelicolor* A3(2). *J Bacteriol* **178:** 2238–2244.

Caballero, J.L., Malpartida, F., and Hopwood, D.A. (1991) Transcriptional organization and regulation of an antibiotic export complex in the producing *Streptomyces* culture. *Mol Gen Genet* **228:** 372–380.

Chopra, I., and Roberts, M. (2001) Tetracycline antibiotics: mode of action, applications, molecular biology, and epidemiology of bacterial resistance. *Microbiol Mol Biol Rev* **65:** 232–260.

Fernandez-Moreno, M.A., Caballero, J.L., Hopwood, D.A., and Malpartida, F. (1991) The *act* cluster contains regulatory and antibiotic export genes, direct targets for translational control by the *bldA* tRNA gene of *Streptomyces*. *Cell* **66:** 769–780.

Flatman, R.H., Howells, A.J., Heide, L., Fiedler, H.P., and Maxwell, A. (2005) Simocyclinone D8, an inhibitor of DNA gyrase with a novel mode of action. *Antimicrob Agents Chemother* **49:** 1093–1100.

Galm, U., Schimana, J., Fiedler, H.P., Schmidt, J., Li, S.M., and Heide, L. (2002) Cloning and analysis of the simocyclinone biosynthetic gene cluster of *Streptomyces antibioticus* Tu 6040. *Arch Microbiol* **178:** 102–114.

Gormley, N.A., Orphanides, G., Meyer, A., Cullis, P.M., and Maxwell, A. (1996) The interaction of coumarin antibiotics with fragments of DNA gyrase B protein. *Biochemistry* **35:** 5083–5092.

Gregory, M.A., Till, R., and Smith, M.C. (2003) Integration site for *Streptomyces* phage ΦBT1 and development of site-specific integrating vectors. *J Bacteriol* **185:** 5320–5323.

Hardy, C.D., and Cozzarelli, N.R. (2003) Alteration of

Escherichia coli topoisomerase IV to novobiocin resistance. *Antimicrob Agents Chemother* **47:** 941–947.

Hesketh, A., Chen, W.J., Ryding, J., Chang, S., and Bibb, M. (2007) The global role of ppGpp synthesis in morphological differentiation and antibiotic production in *Streptomyces coelicolor* A3(2). *Genome Biol* **8:** R161.

Hillen, W., and Berens, C. (1994) Mechanisms underlying expression of Tn*10* encoded tetracycline resistance. *Annu Rev Microbiol* **48:** 345–369.

Holzenkämpfer, M., Walker, M., Zeeck, A., Schimana, J., and Fiedler, H.P. (2002) Simocyclinones, novel cytostatic angucyclinone antibiotics produced by Streptomyces antibioticus Tü 6040. II. Structure elucidation and biosynthesis. *J Antibiot (Tokyo)* **55:** 301–307.

Hong, H.J., Hutchings, M.I., Hill, L.M., and Buttner, M.J. (2005) The role of the novel Fem protein VanK in vancomycin resistance in *Streptomyces coelicolor*. *J Biol Chem* **280:** 13055–13061.

Hopwood, D.A. (2007) How do antibiotic-producing bacteria ensure their self-resistance before antibiotic biosynthesis incapacitates them? *Mol Microbiol* **63:** 937–940.

Jiang, H., and Hutchinson, C.R. (2006) Feedback regulation of doxorubicin biosynthesis in *Streptomyces peucetius*. *Res Microbiol* **157:** 666–674.

Kieser, T., Bibb, M.J., Buttner, M.J., Chater, K.F., and Hopwood, D.A. (2000) *Practical Streptomyces Genetics*. Norwich: The John Innes Foundation.

Kisker, C., Hinrichs, W., Tovar, K., Hillen, W., and Saenger, W. (1995) The complex formed between Tet repressor and tetracycline-Mg^{2+} reveals mechanism of antibiotic resistance. *J Mol Biol* **247:** 260–280.

Maxam, A.M., and Gilbert, W. (1980) Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol* **65:** 499–560.

Maxwell, A., and Lawson, D.M. (2003) The ATP-binding site of type II topoisomerases as a target for antibacterial drugs. *Curr Top Med Chem* **3:** 283–303.

Nollmann, M., Crisona, N.J., and Arimondo, P.B. (2007) Thirty years of *Escherichia coli* DNA gyrase: from *in vivo* function to single-molecule mechanism. *Biochimie* **89:** 490–499.

Oppegard, L.M., Hamann, B.L., Streck, K.R., Ellis, K.C., Fiedler, H.-P., Khodursky, A.B., and Hiasa, H. (2009) *In vivo* and *in vitro* patterns of the activity of simocyclinone D8, an angucyclinone antibiotic from *Streptomyces antibioticus*. *Antimicrob Agents Chemother* **53:** 2110–2119.

Orth, P., Schnappinger, D., Hillen, W., Saenger, W., and Hinrichs, W. (2000) Structural basis of gene regulation by the tetracycline inducible Tet repressor-operator system. *Nat Struct Biol* **7:** 215–219.

Ostash, I., Ostash, B., Luzhetskyy, A., Bechthold, A., Walker, S., and Fedorenko, V. (2008) Coordination of export and glycosylation of landomycins in *Streptomyces cyanogenus* S136. *FEMS Microbiol Lett* **285:** 195–202.

Otten, S.L., Ferguson, J., and Hutchinson, C.R. (1995) Regulation of daunorubicin production in *Streptomyces peucetius* by the *dnrR2* locus. *J Bacteriol* **177:** 1216–1224.

Paget, M.S., Chamberlin, L., Atrih, A., Foster, S.J., and Buttner, M.J. (1999) Evidence that the extracytoplasmic function sigma factor σ^E is required for normal cell wall structure in *Streptomyces coelicolor* A3(2). *J Bacteriol* **181:** 204–211.

Ramos, J.L., Martinez-Bueno, M., Molina-Henares, A.J., Teran, W., Watanabe, K., Zhang, X., *et al.* (2005) The TetR family of transcriptional repressors. *Microbiol Mol Biol Rev* **69:** 326–356.

Rosenfeld, N., Elowitz, M.B., and Alon, U. (2002) Negative autoregulation speeds the response times of transcription networks. *J Mol Biol* **323:** 785–793.

Schimana, J., Fiedler, H.P., Groth, I., Sussmuth, R., Beil, W., Walker, M., and Zeeck, A. (2000) Simocyclinones, novel cytostatic angucyclinone antibiotics produced by *Streptomyces antibioticus* Tu 6040. I. Taxonomy, fermentation, isolation and biological activities. *J Antibiot (Tokyo)* **53:** 779–787.

Schmutz, E., Muhlenweg, A., Li, S.M., and Heide, L. (2003) Resistance genes of aminocoumarin producers: two type II topoisomerase genes confer resistance against coumermycin A1 and clorobiocin. *Antimicrob Agents Chemother* **47:** 869–877.

Tahlan, K., Ahn, S.K., Sing, A., Bodnaruk, T.D., Willems, A.R., Davidson, A.R., and Nodwell, J.R. (2007) Initiation of actinorhodin export in *Streptomyces coelicolor*. *Mol Microbiol* **63:** 951–961.

Tahlan, K., Yu Z., Xu, Y., Davidson, A.R., and Nodwell, J.R. (2008) Ligand recognition by ActR, a TetR-like regulator of actinorhodin export. *J Mol Biol* **383:** 753–761.

Theobald, U., Schimana, J., and Fiedler, H.P. (2000) Microbial growth and production kinetics of *Streptomyces antibioticus* Tu 6040. *Antonie Van Leeuwenhoek* **78:** 307–313.

Thiara, A.S., and Cundliffe, E. (1988) Cloning and characterization of a DNA gyrase B gene from *Streptomyces sphaeroides* that confers resistance to novobiocin. *EMBO J* **7:** 2255–2259.

Thiara, A.S., and Cundliffe, E. (1989) Interplay of novobiocin-resistant and -sensitive DNA gyrase activities in self-protection of the novobiocin producer, *Streptomyces sphaeroides*. *Gene* **81:** 65–72.

Thiara, A.S., and Cundliffe, E. (1993) Expression and analysis of two *gyrB* genes from the novobiocin producer, *Streptomyces sphaeroides*. *Mol Microbiol* **8:** 495–506.

Trefzer, A., Pelzer, S., Schimana, J., Stockert, S., Bihlmaier, C., Fiedler, H.P., *et al.* (2002) Biosynthetic gene cluster of simocyclinone, a natural multihybrid antibiotic. *Antimicrob Agents Chemother* **46:** 1174–1182.

Willems, A.R., Tahlan, K., Taguchi, T., Zhang, K., Lee, Z.Z., Ichinose, K., *et al.* (2008) Crystal structures of the *Streptomyces coelicolor* TetR-like protein ActR alone and in complex with actinorhodin or the actinorhodin biosynthetic precursor (S)-DNPA. *J Mol Biol* **376:** 1377–1387.

Acta Crystallographica Section F
Structural Biology and Crystallization Communications
ISSN 1744-3091

Crystallization and preliminary X-ray analysis of the TetR-like efflux pump regulator SimR

Tung B. K. Le,[a] Clare E. M. Stevenson,[b] Mark J. Buttner[a] and David M. Lawson[b]*

[a]Department of Molecular Microbiology, John Innes Centre, Norwich NR4 7UH, England, and [b]Department of Biological Chemistry, John Innes Centre, Norwich NR4 7UH, England

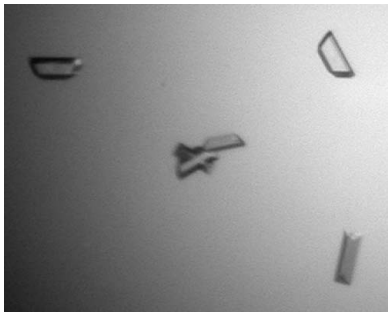
Correspondence e-mail: david.lawson@bbsrc.ac.uk

Received 15 November 2010
Accepted 17 December 2010

Crystals of SimR were grown by vapour diffusion. The protein crystallized with trigonal symmetry and X-ray data were recorded to a resolution of 2.3 Å from a single crystal at the synchrotron. SimR belongs to the TetR family of bacterial transcriptional regulators. In the absence of the antibiotic simocyclinone, SimR represses the transcription of a divergently transcribed gene encoding the simocyclinone efflux pump SimX in *Streptomyces antibioticus* by binding to operators in the *simR–simX* intergenic region. Simocyclinone binding causes SimR to dissociate from its operators, leading to expression of the SimX efflux pump. Thus, SimR represents an intimate link between the biosynthesis of simocyclinone and its export, which may also provide the mechanism of self-resistance to the antibiotic in the producer strain.

1. Introduction

Most antibiotics are potentially lethal to the producing organism and therefore there must be mechanisms to ensure that the machinery responsible for the export of the mature antibiotic is in place at the time of biosynthesis. Simocyclinone (Schimana *et al.*, 2000) is a potent DNA gyrase inhibitor that is produced by *Streptomyces antibioticus* Tü 6040 (Edwards *et al.*, 2009; Flatman *et al.*, 2005; Oppegard *et al.*, 2009). Within the simocyclinone-biosynthetic cluster are two divergently transcribed genes, *simR* and *simX*, that encode proteins that resemble the TetR/TetA repressor–efflux pump pair that causes widespread resistance to clinically important tetracyclines (Chopra & Roberts, 2001). Transcription of *simX* is controlled by SimR, which directly represses the *simX* and *simR* promoters by binding to two operator sites in the *simX–simR* intergenic region (Le *et al.*, 2009). Simocyclinone abolishes DNA binding by SimR, inducing expression of the SimX efflux pump and thus providing a mechanism that couples the biosynthesis of simocyclinone to its export. TetR-family transcriptional regulators are widespread in the bacterial kingdom (Ramos *et al.*, 2005) and there are representative structures of over 100 of these in the Protein Data Bank (Yu *et al.*, 2010). However, the mode of action has been fully elucidated at the molecular level for relatively few of these proteins (Orth *et al.*, 2000; Schumacher *et al.*, 2001; Itou *et al.*, 2010; Miller *et al.*, 2010; Reichheld *et al.*, 2009; Yu *et al.*, 2010). Here, we report the crystallization and preliminary X-ray analysis of SimR as the first step towards determining the molecular basis of its action.



2. Materials and methods

2.1. Protein expression and purification and crystallization

The *simR* gene of *S. antibioticus* Tü 6040, encoding a protein of 259 amino acids, was chemically synthesized and codon-optimized (GenScript) for expression in *Escherichia coli* and then ligated into the pETM11-*Nde*I plasmid to give the expression vector pIJ10495. The SimR protein produced from this vector has an N-terminal TEV protease-cleavable His tag. This adds a further 26 amino acids to the native protein (with sequence MKHHHHHHPMSDYDIPTTENLYFQGA), giving a total deduced molecular mass of 32 222 Da. The

expression vector pIJ10495 was transformed into *E. coli* strain BL21 (DE3) pLysS and a 7 ml overnight culture of the cells was used to inoculate a 500 ml culture of Luria–Bertani medium containing 50 µg ml^{-1} kanamycin and 30 µg ml^{-1} chloramphenicol. The cells were grown at 310 K to an $OD_{600\,nm}$ of around 0.4. The culture was then cooled to 293 K and dimethylsulfoxide (DMSO) was added to a final concentration of 0.15%(*v*/*v*) before induction of protein expression by the addition of isopropyl β-D-thiogalactopyranoside to a final concentration of 0.3 m*M*. The addition of DMSO was essential for a high yield of soluble SimR, probably because it induces the expression of chaperones that assist in the folding of SimR *in vivo*. The culture was left shaking for 3 h at 303 K. Harvested cells were resuspended in 50 m*M* Tris–HCl pH 8.4, 300 m*M* NaCl, 5%(*v*/*v*) glycerol containing Complete EDTA-free protease-inhibitor cocktail (Roche) and lysed by sonication (three cycles of 20 s with 40 s resting on ice in between each cycle). The cell debris was removed by centrifugation at 84 000*g* for 30 min and the supernatant was filtered through a 0.45 µm membrane. It was then applied onto a 1 ml Ni-loaded Hi-Trap Chelating HP column (GE Healthcare) connected to an ÄKTA FPLC system (Pharmacia) that had been equilibrated with buffer *A* (50 m*M* Tris–HCl pH 8.4, 300 m*M* NaCl, 50 m*M* imidazole). Protein was eluted from the column using an increasing (50–500 m*M*) imidazole gradient in the same buffer. SimR fractions were identified using SDS–PAGE, pooled together and concentrated to approximately 5–10 mg ml^{-1} using a Vivaspin 6 10 kDa cutoff concentrator (Vivascience). His-tagged SimR is prone to precipitation and further purification by gel-filtration chromatography resulted in very poor yields (data not shown). Although attempts to cleave the affinity tag with TEV protease were successful (as judged by SDS–PAGE), the resultant sample still adhered to the Ni-affinity resin when reapplied onto the Hi-Trap Chelating HP column and could not be eluted even with buffer containing 1.0 *M* imidazole (data not shown). Therefore, for all subsequent preparations a single-column purification procedure was used and the His tag was not cleaved. Directly after SDS–PAGE analysis the pooled fractions were exchanged into crystallization buffer (25 m*M* Tris–HCl pH 8.4, 300 m*M* NaCl) using a Zeba desalting micro column (Thermo Scientific).

Dynamic light scattering (DLS) was used to monitor the solution properties of the purified sample. For this purpose, approximately 30 µl protein solution was centrifuged through a 0.1 µm Ultrafree filter (Millipore) to remove particulate material before introduction into a 12 µl microsampling cell. The cell was inserted into a DynaPro Titan molecular-sizing instrument at 293 K (Wyatt Technology). A minimum of ten scattering measurements were obtained and the resulting data were analysed using the *DYNAMICS* software package (Wyatt Technology).

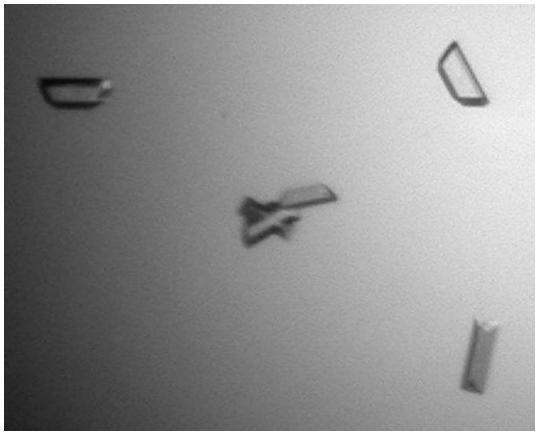


Figure 1
Single crystals of SimR, with approximate dimensions of 200 × 100 × 100 µm.

Crystallization trials of His-tagged SimR were set up using an OryxNano robot (Douglas Instruments Ltd) in sitting-drop vapour-diffusion format with 96-well MRC plates (Molecular Dimensions) using a variety of commercially available screens (Molecular Dimensions and Qiagen) at a constant temperature of 293 K. Drops consisted of 0.3 µl protein solution mixed with 0.3 µl precipitant solution and the reservoir volume was 50 µl; the protein concentration was approximately 5 mg ml^{-1}. Improved crystals were subsequently obtained by refining the successful conditions in a hanging-drop format using 24-well VDX plates (Molecular Dimensions) over a reservoir volume of 1 ml.

2.2. X-ray data collection

Crystals were mounted for X-ray data collection using LithoLoops (Molecular Dimensions) and were flash-cooled by plunging into liquid nitrogen and stored in Uni-puck cassettes prior to transport to the synchrotron. Crystals were subsequently transferred robotically to the goniostat on station I03 of the Diamond Light Source (Oxfordshire, England) and maintained at 100 K with a Cryojet cryocooler (Oxford Instruments). Diffraction data were recorded using an ADSC Quantum 315 CCD detector with the wavelength set to 0.9763 Å and were then processed using *MOSFLM* (Leslie, 2006) and *SCALA* (Evans, 2006).

3. Results and discussion

N-terminally His-tagged SimR was overexpressed and purified with an approximate yield of 40 mg protein from 500 ml culture and was judged to be greater than 95% pure by SDS–PAGE analysis. DLS analysis gave a peak accounting for 99.9% of the mass of the sample with a polydispersity of 16.5%. From these results the molecular size was estimated at 70 kDa, being close to the value expected for a SimR homodimer (64 kDa).

Preliminary crystals of SimR grew within two weeks of setup with 17%(*w*/*v*) PEG 10 000, 0.2 *M* ammonium acetate in 0.1 *M* bis-tris pH 5.5 as the precipitant. Improved crystals were subsequently obtained overnight from 2%(*w*/*v*) PEG 10 000, 0.2 *M* ammonium acetate in 0.1 *M* bis-Tris pH 5.5, with maximum dimensions of approximately 200 × 100 × 100 µm (Fig. 1). We found that subjecting SimR stock solutions to heat treatment prior to setting up crystallizations resulted in better quality crystals. SimR (10 mg ml^{-1}) was incubated at 310 K for 2 min and then cooled in running water at ∼283 K for 1 min and the solution was cleared of precipitated protein by centrifugation through a Millipore 0.1 µm Ultrafree filter before crystallization. SimR crystals were cryoprotected by a three-step transfer (25, 50 and 100% cryosolution) to the final cryosolution [crystallization solution with the addition of 30%(*w*/*v*) PEG 1500 in place of an equivalent volume of water].

Native X-ray data were collected from a single SimR crystal: a total of 200 × 0.45° oscillation images were recorded in a continuous sweep to a maximum resolution of 2.3 Å. Indexing was consistent with a rhombohedral lattice, with unit-cell parameters $a = b = 116.62$, $c = 110.58$ Å. Analysis using the program *POINTLESS* (Evans, 2006) indicated that the space group was *H*32. Data-collection and processing statistics are summarized in Table 1. Estimation of the content of the asymmetric unit indicated that a single His-tagged SimR subunit was most likely, giving a solvent content of 45% and a crystal-packing parameter (V_M) of 2.25 Å^3 Da^{-1} (Matthews, 1968).

Table 1
Summary of X-ray data for SimR.

Values in parentheses are for the outer resolution shell.

No. of crystals	1
Beamline	I03, Diamond Light Source, England
Wavelength (Å)	0.9763
Detector	ADSC Quantum 315 CCD
Crystal-to-detector distance (mm)	368.6
Rotation range per image (°)	0.45
Total rotation range (°)	90.0
Resolution range (Å)	48.50–2.30 (2.42–2.30)
Space group	*H*32
Unit-cell parameters (Å)	$a = b = 116.62$, $c = 110.58$
Mosaicity (°)	0.79
Total no. of measured intensities	63659 (6804)
Unique reflections	12693 (1786)
Multiplicity	5.0 (3.8)
Mean $I/\sigma(I)$	10.2 (2.3)
Completeness (%)	98.0 (95.2)
$R_{\rm merge}$†	0.093 (0.523)
$R_{\rm meas}$‡	0.104 (0.604)
Wilson *B* value (Å^2)	57.9

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$. ‡ $R_{\rm meas} = \sum_{hkl}[N/(N-1)]^{1/2} \times \sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the *i*th observation of reflection *hkl*, $\langle I(hkl)\rangle$ is the weighted average intensity for all observations *i* of reflection *hkl* and *N* is the number of observations of reflection *hkl*.

A partial molecular-replacement solution for SimR was obtained with the *BALBES* pipeline (Long *et al.*, 2008; with an MR score of 2.63) using the structure of an *S. coelicolor* putative transcriptional regulator (PDB code 2hxi; 25.4% sequence identity to SimR; K. Tan, X. Xu, H. Zheng, A. Savchenko, A. Edwards & A. Joachimiak, unpublished work) as the search model. This template was also the top hit that was found by the *FUGUE* fold-prediction server (http://tardis.nibio.go.jp/fugue/prfsearch.html; Shi *et al.*, 2001) based on the amino-acid sequence of SimR, with a *Z* score of 31.2. In addition, a 3.4 Å resolution single-wavelength anomalous dispersion data set was collected from a crystal of selenomethionine-substituted protein (not shown). The combination of these two sources of phase information enabled us to solve the SimR structure. Full details of the structure-determination process and refinement of the resultant model will be reported elsewhere.

This work is funded by a John Innes Centre Rotation Studentship (awarded to TBKL), by BBSRC grant BB/I002197/1 (to MJB and DML), and by the BBSRC Core Strategic Grant to the John Innes Centre. We are grateful to the beamline scientists for assistance with X-ray data collection at the Diamond Light Source and to N. Tucker for supplying the pETM11-*Nde*I plasmid.

References

Chopra, I. & Roberts, M. (2001). *Microbiol. Mol. Biol. Rev.* **65**, 232–260.
Edwards, M. J., Flatman, R. H., Mitchenall, L. A., Stevenson, C. E., Le, T. B., Clarke, T. A., McKay, A. R., Fiedler, H. P., Buttner, M. J., Lawson, D. M. & Maxwell, A. (2009). *Science*, **326**, 1415–1418.
Evans, P. (2006). *Acta Cryst.* D**62**, 72–82.
Flatman, R. H., Howells, A. J., Heide, L., Fiedler, H. P. & Maxwell, A. (2005). *Antimicrob. Agents Chemother.* **49**, 1093–1100.
Itou, H., Watanabe, N., Yao, M., Shirakihara, Y. & Tanaka, I. (2010). *J. Mol. Biol.* **403**, 174–184.
Le, T. B., Fiedler, H. P., den Hengst, C. D., Ahn, S. K., Maxwell, A. & Buttner, M. J. (2009). *Mol. Microbiol.* **72**, 1462–1474.
Leslie, A. G. W. (2006). *Acta Cryst.* D**62**, 48–57.
Long, F., Vagin, A. A., Young, P. & Murshudov, G. N. (2008). *Acta Cryst.* D**64**, 125–132.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Miller, D. J., Zhang, Y.-M., Subramanian, C., Rock, C. O. & White, S. W. (2010). *Nature Struct. Mol. Biol.* **17**, 971–975.
Oppegard, L. M., Hamann, B. L., Streck, K. R., Ellis, K. C., Fiedler, H. P., Khodursky, A. B. & Hiasa, H. (2009). *Antimicrob. Agents Chemother.* **53**, 2110–2119.
Orth, P., Schnappinger, D., Hillen, W., Saenger, W. & Hinrichs, W. (2000). *Nature Struct. Biol.* **7**, 215–219.
Ramos, J. L., Martinez-Bueno, M., Molina-Henares, A. J., Teran, W., Watanabe, K., Zhang, X., Gallegos, M. T., Brennan, R. & Tobes, R. (2005). *Microbiol. Mol. Biol. Rev.* **69**, 326–356.
Reichheld, S. E., Yu, Z. & Davidson, A. R. (2009). *Proc. Natl Acad. Sci. USA*, **106**, 22263–22268.
Schimana, J., Fiedler, H. P., Groth, I., Sussmuth, R., Beil, W., Walker, M. & Zeeck, A. (2000). *J. Antibiot. (Tokyo)*, **53**, 779–787.
Schumacher, M. A., Miller, M. C., Grkovic, S., Brown, M. H., Skurray, R. A. & Brennan, R. G. (2001). *Science*, **294**, 2158–2163.
Shi, J., Blundell, T. L. & Mizuguchi, K. (2001). *J. Mol. Biol.* **310**, 243–257.
Yu, Z., Reichheld, S. E., Savchenko, A., Parkinson, J. & Davidson, A. R. (2010). *J. Mol. Biol.* **400**, 847–864.

doi:10.1016/j.jmb.2011.02.035 J. Mol. Biol. (2011) 408, 40–56

Contents lists available at www.sciencedirect.com

Journal of Molecular Biology

journal homepage: http://ees.elsevier.com.jmb

Structures of the TetR-like Simocyclinone Efflux Pump Repressor, SimR, and the Mechanism of Ligand-Mediated Derepression

Tung B. K. Le[1], Clare E. M. Stevenson[2], Hans-Peter Fiedler[3], Anthony Maxwell[2], David M. Lawson[2]* and Mark J. Buttner[1]

[1]*Department of Molecular Microbiology, John Innes Centre, Norwich Research Park, Norwich NR4 7UH, UK*
[2]*Department of Biological Chemistry, John Innes Centre, Norwich Research Park, Norwich NR4 7UH, UK*
[3]*Mikrobiologisches Institut, Eberhard-Karls-Universität Tübingen, Auf der Morgenstelle 28, D-72076 Tübingen, Germany*

Received 20 December 2010; received in revised form 12 February 2011; accepted 15 February 2011
Available online 24 February 2011

Edited by I. Wilson

Keywords:
crystal structures; transcriptional regulation; antibiotic resistance; antibiotic biosynthesis; *Streptomyces*

Simocyclinone D8 (SD8), a potent DNA gyrase inhibitor made by *Streptomyces antibioticus*, is exported from the producing organism by the SimX efflux pump. The expression of *simX* is under the control of SimR, a member of the TetR family of transcriptional regulators. SimR represses *simX* transcription by binding to operators in the intergenic region between *simR* and *simX*. Previously, we have shown that the mature antibiotic SD8 or its biosynthetic intermediate, simocyclinone C4, can dissociate SimR from its operators, leading to derepression of *simX* and export of SD8 from the cell. This provides a mechanism that couples the biosynthesis of the antibiotic to its export. Here, we report the crystal structures of SimR alone and in complex with either SD8 or simocyclinone C4. The ligand-binding pocket is unusual compared to those of other characterized TetR-family transcriptional regulators: the structures show an extensive ligand-binding pocket spanning both monomers in the functional dimeric unit, with the aminocoumarin moiety of SD8 buried in the protein core, while the angucyclic polyketide moiety is partially exposed to bulk solvent. Through comparisons of the structures, we postulate a derepression mechanism for SimR that invokes rigid-body motions of the subunits relative to one another, coupled with a putative locking mechanism to restrict further conformational change.

Introduction

TetR-family transcriptional regulators (TFRs) are abundant and widespread in bacteria.[1] The majority function as repressors and they are known principally because well-characterized members of the family control genes whose products are involved in antibiotic resistance. Other members regulate genes involved in diverse processes, including the biosynthesis of secondary metabolites (e.g., *Streptomyces coelicolor* CprB) and the pathogenicity of both Gram-negative (e.g., *Vibrio cholerae* HapR) and Gram-positive bacteria (e.g., *Bacillus cereus* HlyIIR).[1] Most

*Corresponding author. E-mail address: david.lawson@bbsrc.ac.uk.

Abbreviations used: AC, aminocoumarin; ASU, asymmetric unit; DBD, DNA-binding domain; DMSO, dimethylsulfoxide; HTH, helix-turn-helix; LBD, ligand-binding domain; PDB, Protein Data Bank; PEG, polyethylene glycol; PK, angucyclic polyketide; SC4, simocyclinone C4; SD8, simocyclinone D8; SeMet, selenomethionine; SimR-apo, ligand-free SimR; SimR-SC4, complex of SimR with SC4; SimR-SD8, complex of SimR with SD8; TFR, TetR-family transcriptional regulator.

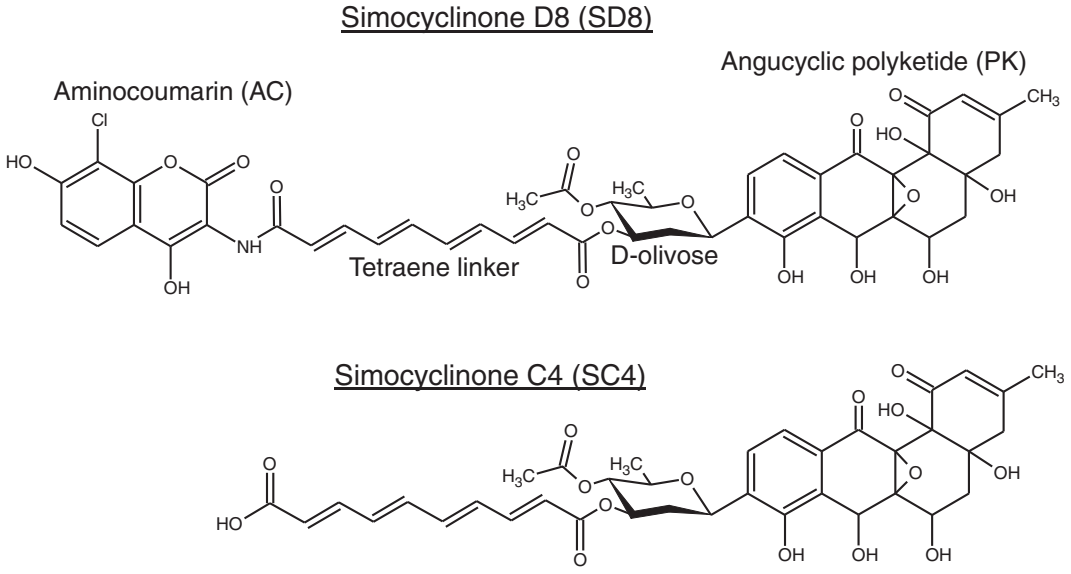


Fig. 1. Structures of simocyclinone D8 and its biosynthetic intermediate, simocyclinone C4.

TFRs are believed to respond to small molecules, but cognate ligands have been identified for only a handful of family members. The best-characterized member of the family is *Escherichia coli* TetR itself, which confers resistance to the antibiotic tetracycline by regulating the expression of the tetracycline efflux pump, TetA.[2] In the absence of tetracycline, TetR binds to the intergenic region between the divergently transcribed *tetR* and *tetA* at two operator sites.[3,4] This action represses the transcription of both *tetA* and *tetR* itself. When TetR binds tetracycline, it loses affinity for these operators, leading to derepression of *tetA* and the export of tetracycline from the cell.[5] Representative crystal structures for ligand-free, ligand-bound, and DNA-bound forms of TetR have been determined.[6,7]

TFRs function as homodimers. Each monomer consists of two domains, an N-terminal DNA-binding domain (DBD) containing a helix-turn-helix (HTH) motif and a C-terminal ligand-binding domain (LBD).[1] LBDs are extremely diverse in amino acid sequence, and it is therefore not generally possible to model these domains based on existing structures with any confidence, nor to predict potential ligands.[1,8] In addition to TetR, the structures of three further TFRs have been determined with relevant ligands, and also with bound DNA, enabling their mechanisms of repression and derepression to be elucidated. Two of these, *Staphylococcus aureus* QacR[9] and *Corynebacterium glutamicum* CgmR,[10] like TetR, repress the transcription of efflux pump genes and lose affinity for the operator sites of these genes upon binding small molecules. The fourth protein, DesT from *Pseudomonas aeruginosa*, differs in that it responds to the ratio of two ligands. It controls the composition of membrane lipids by acting as a repressor of a desaturase; when unsaturated fatty acid is bound to DesT, the protein has high affinity for its cognate DNA, but loses this affinity when saturated fatty acid is bound.[11] Four other TFRs have been crystallized in complex with cognate ligands. These include two further efflux pump regulators, *Pseudomonas putida* TtgR[12] and *S. coelicolor* ActR.[13] ActR controls the efflux of the antibiotic actinorhodin in the producer; the structure reveals a binding pocket that accommodates one molecule of the mature hexacyclic antibiotic or two molecules of a tricyclic intermediate that also efficiently derepresses ActR.[13,14] The remaining two examples are *Mycobacterium tuberculosis* EthR, which represses expression of a mono-oxygenase required for the activation of the pro-drug ethionamide,[15–17] and *E. coli* MphR, a repressor of the macrolide 2′-phosphotransferase I resistance

Fig. 2. Crystal structures of ligand-free (SimR-apo) and SD8-bound (SimR–SD8) forms of SimR. The top panel uses a cylindrical helix representation to highlight the secondary structure of SimR-apo with key features labelled. The two subunits of the homodimer are distinguished by cyan and yellow colouration, and the two recognition helices are shown in magenta. Where appropriate, this colour scheme is adopted throughout the subsequent figures. Note the α9–α10 arm that engages with the opposing subunit. The lower panels compare SimR-apo and SimR–SD8 in cartoon representation viewed from the front (left) and from the top (right), with the SD8 molecules shown in red space-filling representation. The separations of the recognition helices are indicated. The labels AC and PK refer to the aminocoumarin and polyketide moieties of SD8, respectively.

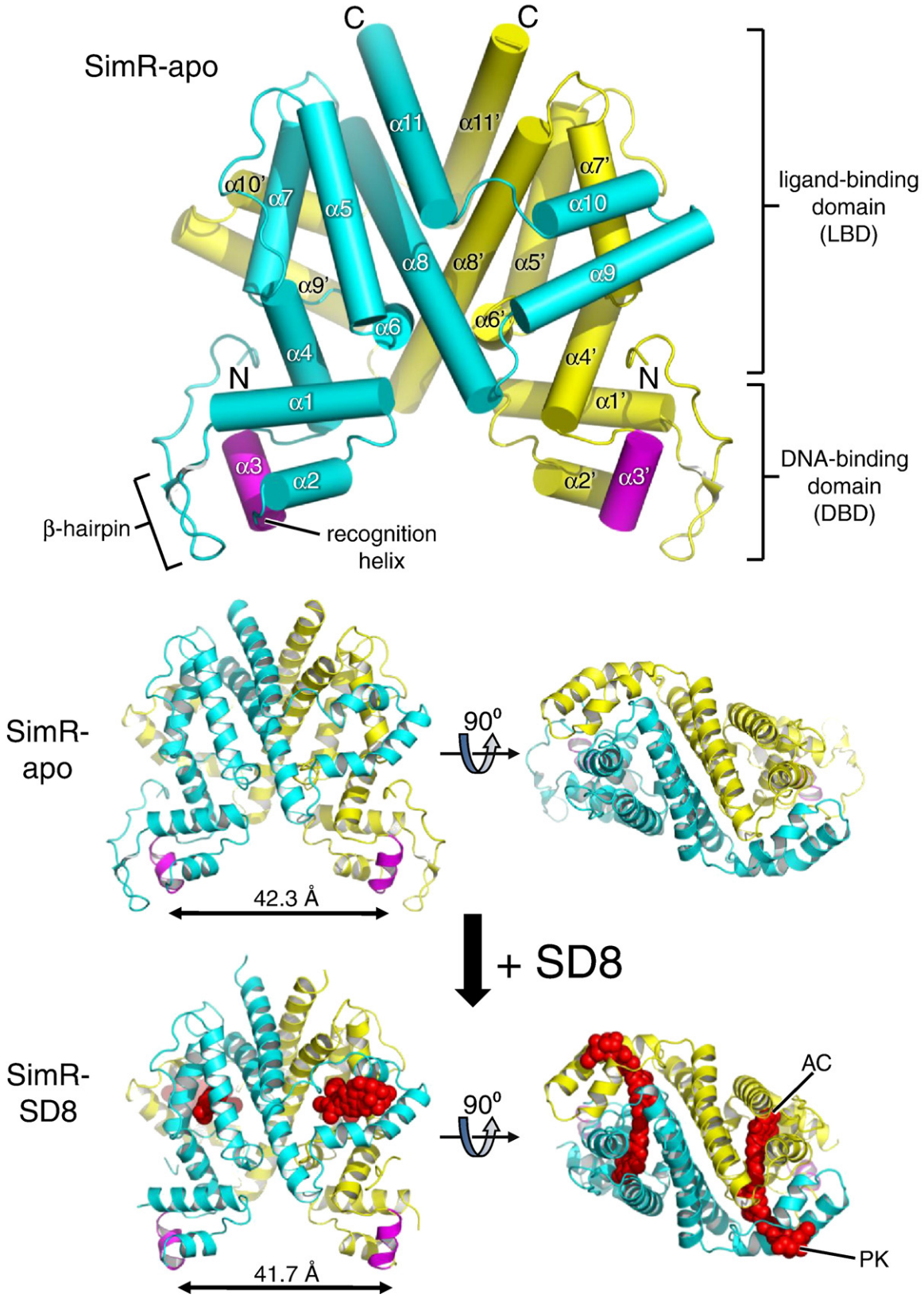


Fig. 2 (*legend on previous page*)

gene.[18] In those cases where derepression has been analysed by comparing DNA-bound and ligand-bound structures,[6,7,9–11] it is brought about through the stabilisation, by the binding of the ligand, of a conformation in which the recognition helices are too far apart to simultaneously bind to adjacent major grooves of DNA. These mechanisms necessarily invoke allosteric communication between the LBDs and DBDs.

Simocyclinone D8 (SD8) is a potent DNA gyrase inhibitor produced by *Streptomyces antibioticus*, with an IC_{50} ~0.1 μM for inhibition of DNA supercoiling.[19–21] It consists of four moieties: a chlorinated aminocoumarin (AC) at one end, connected to an angucyclic polyketide (PK) at the other end, via a tetraene linker and a D-olivose sugar (Fig. 1).[22–24] SD8 is exported from the producing organism by a specific efflux pump, SimX, and the transcription of *simX* is repressed by SimR, a TFR that binds to two distinct operators in the intergenic region between the divergently transcribed *simR* and *simX* genes.[25] SD8 abolishes DNA binding by SimR, inducing expression of the SimX efflux pump, and this provides an intimate mechanism that couples the biosynthesis of simocyclinone to its export.[25]

In order to gain insight into the molecular mechanism of ligand-mediated SimR derepression, we have solved the crystal structures of SimR, alone and in complexes with SD8 and its biosynthetic intermediate simocyclinone C4 (SC4), which lacks the AC moiety of the mature antibiotic (Fig. 1) but still derepresses SimR, albeit less efficiently.[25] The structures show an extensive ligand-binding pocket spanning both monomers in the functional dimeric unit and suggest a derepression mechanism that involves rigid-body motions of the subunits relative to one another, coupled with a putative locking mechanism to restrict further conformational change.

Results and Discussion

The structure of ligand-free SimR

The structure of ligand-free SimR (SimR-apo) was determined at 1.95-Å resolution. There is a single copy of the SimR monomer in the asymmetric unit (ASU), and the final model comprises residues 7–244 of the 259–residue native sequence. A biologically relevant homodimer is generated through the application of 2-fold crystallographic symmetry (Fig. 2). The secondary structure of SimR is almost entirely α-helical, being composed of 11 helices that span residues 29–43 (α1), 50–57 (α2), 61–67 (α3), 71–84 (α4), 96–113 (α5), 116–120 (α6), 130–144 (α7), 150–181 (α8), 185–202 (α9), 206–216 (α10), and 222–243 (α11) (Figs. 2 and 3). There are, in addition, two short β-strands near the N-terminus (residues 18–19 and 26–27), which form a β-hairpin. The SimR monomer can be divided into two domains, an N-terminal DBD and a C-terminal LBD. The DBD consists of the β-hairpin and helices α1 to α3. Helices α2 and α3 form an HTH motif that is stabilised by packing against α1. By analogy with other TFRs, the HTH is predicted to interact with the major groove of DNA, with the recognition helix, α3, providing the majority of base-specific interactions. In addition, the N-terminal β-hairpin motif of SimR appears to be ideally placed to make minor groove interactions. Consistent with this proposal, there are a total of four Arg residues in the range 18–25 (Fig. 3), which could interact favourably with the negatively charged phosphate backbone of DNA. These residues contribute to the prominent electropositive surface patch that overlaps the recognition helix in SimR-apo (Supplementary Fig. S1).

The last eight helices (α4–α11) make up the LBD responsible for small-molecule ligand binding and forming the interface between the two subunits of the dimer. Dimerization occurs largely through a pair of helices, α8 and α11, from each SimR monomer that form a four-helix bundle (Fig. 2). The dimerization interface is further enhanced by the α9–α10 segment, which forms an extended 'arm' that wraps around the LBD of the opposing monomer (Fig. 2). (Note that the distinctive kink in α9 is important for ligand binding; see below.) In fact, all of helices α7–α11 from each subunit are involved to varying degrees in dimerization. Overall, this gives rise to an extensive dimer interface of 2475 $Å^2$ accounting for 17.5% of the solvent-accessible surface of an isolated monomer, as calculated by the Protein Interactions, Surfaces and Assemblies server (PISA†).[28] Within the core of each subunit, there is a substantial cavity (Fig. 4a). The two cavities of the dimer are linked via a narrow pore that passes through the centre of the dimer, and they each connect to bulk solvent through channels that pass between the two helices of the α9–α10 arm (Fig. 4a).

Comparison of the SimR structure with other TetR family members

The closest structural homologues of SimR were identified using the DALI server‡ (Supplementary Fig. S2).[29] This gave a total of 12 hits from the nonredundant database with *Z* scores of greater than 15.0. The top hit was *S. coelicolor* ActR with bound actinorhodin [Protein Data Bank (PDB) code 3B6A],[13] with a *Z* score of 18.7, a sequence identity of 25%, and a root-mean-square deviation (rmsd) of 2.6 Å after superposition of single subunits. Hit number 6 was *E. coli* TetR with bound tetracycline

† http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html
‡ http://ekhidna.biocenter.helsinki.fi/dali_server/

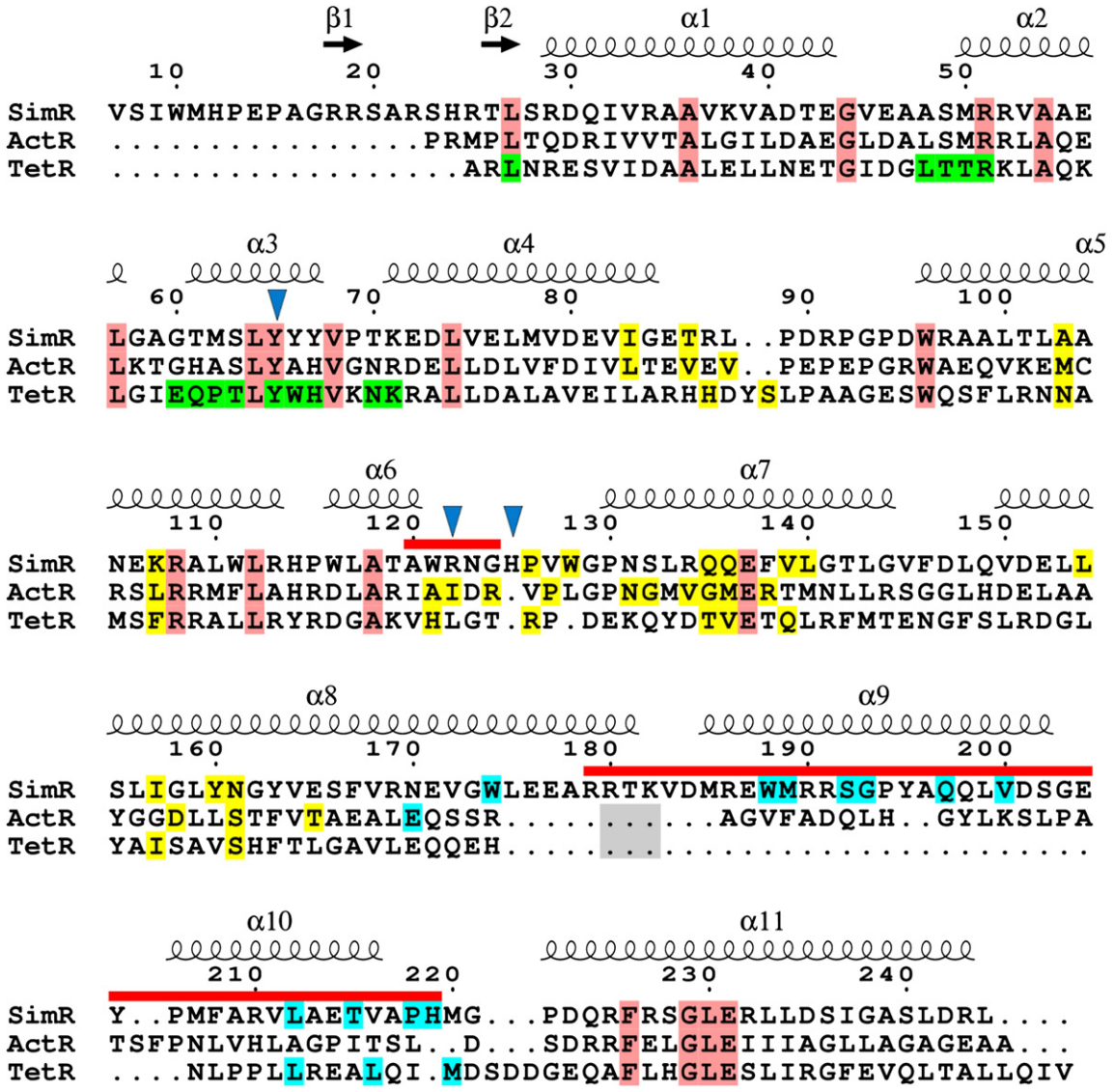


Fig. 3. Structure-based multiple sequence alignment of SimR with *S. coelicolor* ActR and *E. coli* TetR. The initial alignment was generated using the SSM server (http://www.ebi.ac.uk/msd-srv/ssm/),[26] with the ligand-free structures (PDB codes 2OPT and 1BJZ for ActR and TetR, respectively). This was subsequently adjusted manually with reference to the superposed structures and then displayed using ESPript.[27] Strictly conserved residues are highlighted with pink shading. Secondary-structure elements for SimR are shown above the alignment, where α stands for α-helix and β represents β-strand. Residues involved in binding SD8 (selected on the basis of being <4.0 Å from the ligand) are highlighted with yellow and cyan shading to denote the subunit. Similarly, residues involved in actinorhodin binding by ActR (PDB code 3B6A) and tetracycline binding by TetR (PDB code 1BJY) are highlighted with the same colour scheme. Green shading indicates residues involved in DNA binding by TetR (PDB code 1QPI). Grey shading indicates regions in ActR (residues 178–187) and TetR (residues 152–164) that were not modelled in the ligand-free structures. The red bars denote regions of SimR that show the most significant conformational changes between the ligand-bound and ligand-free structures. The inverted blue triangles mark key residues referred to in the text.

(PDB code 1BJY[6]; *Z* score, 17.0; sequence identity, 18%; rmsd, 2.6 Å). In a recent systematic survey of TFR protein structures, TetR itself, together with ActR, were placed in a distinct subclass of structures.[8] One of the characteristics of this subclass is a long insertion between α8 and the C-terminal α-helix. This corresponds to the α9–α10 arm of SimR. By contrast, CgmR from *C. glutamicum* (PDB code 2ZOY; ligand-free form)[30] is more typical of the majority of known TFR structures, having no such insertion (Supplementary Fig. S2b). As a consequence, the dimer interface of CgmR is less extensive than that of SimR-apo at 1819 Å^2 (as calculated by PISA). In addition, several TFRs have a significantly

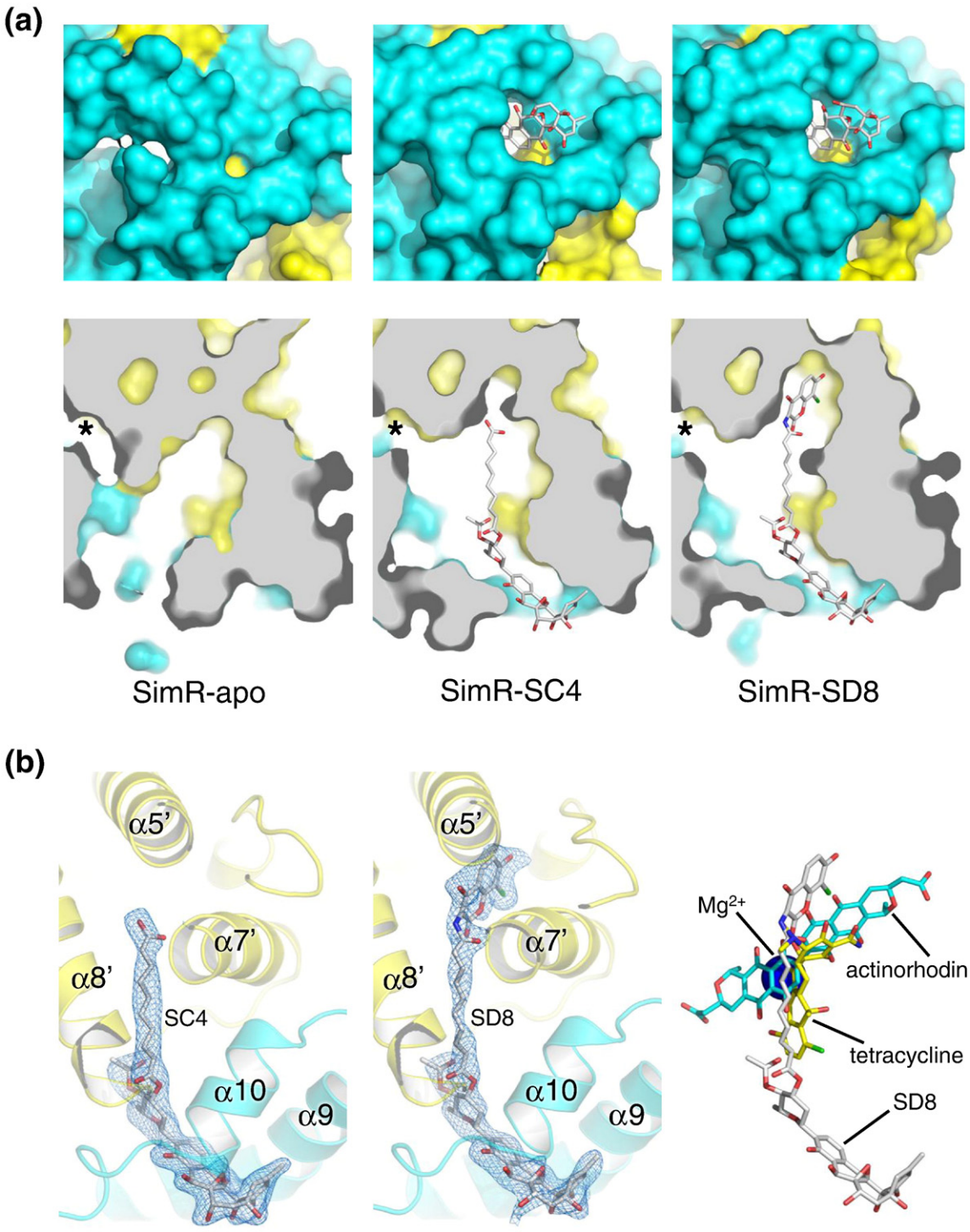


Fig. 4. The ligand-binding pocket of SimR. (a) Molecular surface representations of the ligand-free and ligand-bound structures of SimR, focussing on the right-hand pocket shown in Fig. 2; ligands are shown with grey carbon atoms. The top panels show a front view and the lower panels show a top view as a thin slice through the molecule. Grey shading denotes regions within the molecular boundary. The asterisks indicate narrow channels leading to equivalent cavities in the other halves of the dimers. (b) Simulated annealing omit maps at 2.3-Å resolution for SimR–SC4 and SimR–SD8 (contoured at 3σ) are shown in blue superposed on their respective ligands; the protein is shown in semi-transparent cartoon representation (see also Supplementary Fig. S3). The right-hand panel illustrates the relative positions of the ligands for SimR, ActR, and TetR after superposition of their respective structures.

longer α1 helix, the N-terminus of which may make additional contacts to DNA as seen in QacR, CgmR, and DesT.[10,11,31,32]

Tyr42 in TetR lies in the middle of the recognition helix (α3) and forms important interactions with DNA in the TetR–*tetO* complex.[7] This residue is reasonably well conserved in TFRs, and the C^{α} atom serves as a reference point for measuring the separation of the recognition helices in a homodimer.[8] In DNA-bound structures, this separation lies in the range 34.7 to 38.8 Å,[7,10,11,31,32] roughly corresponding to the distance of 34 Å between consecutive major grooves in standard B-form DNA. By contrast, in apo-TFR structures, the helix separation is highly variable but is typically not compatible with DNA binding.[8] The equivalent residue in SimR is Tyr65 and the C^{α}–C^{α} separation is 42.3 Å in SimR-apo (Fig. 2). Thus, we conclude that the conformation seen in SimR-apo is also not compatible with DNA binding (Supplementary Fig. S2d).

Ligand-bound structures of SimR

SC4 is a biosynthetic intermediate of SD8 (Fig. 1) that is essentially inactive as a DNA gyrase inhibitor.[19] However, like the mature antibiotic, SC4 derepresses SimR; when applied exogenously, it induces expression of the SimX efflux pump *in vivo*, and it relieves DNA binding by SimR *in vitro*.[25] Structures of the complexes formed between SimR and the mature antibiotic SD8 (SimR–SD8), and with SC4 (SimR–SC4), were determined at 2.3-Å resolution. In both cases, crystals of the complexes were formed by co-crystallization with the respective ligands and gave rise to new crystal forms. Each structure was solved by molecular replacement using the SimR-apo structure as the template, and in both cases, there were two copies of the SimR subunit in the ASU, corresponding to the biological dimer. In contrast to the SimR-apo structure, most of the N-terminal polypeptide chain prior to α1 in the ligand-bound structures was poorly resolved; only in one monomer of SimR–SC4 was the β-hairpin visible. For both complexes, electron density for two copies of the corresponding ligand per SimR dimer was clearly present, occupying equivalent sites related by 2-fold noncrystallographic symmetry (Fig. 4b; Supplementary Fig. S3). At the protein backbone level, the two ligand-bound structures of SimR were closely similar. In pairwise superpositions of individual subunits, the rmsd values were in the range 0.66 to 0.75 Å, and a dimer-upon-dimer superposition gave a value of 0.74 Å. In comparisons between the SimR-apo and the two ligand-bound structures at the subunit level, the most striking difference is the concerted upward shift of the α9–α10 arm upon ligand binding (Fig. 5a). Superpositions of individual ligand-bound subunits upon the SimR-apo subunit, based only on residues in the subunit core (i.e. excluding α9–α10, specifically residues 182–221), gave remarkably low rmsd values in the range 1.01 to 1.09 Å. Thus, the cores are extremely similar, and most importantly, the relative orientations of LBDs and DBDs within each subunit remain essentially unchanged upon ligand binding. At the dimer level, differences between the SimR-apo and ligand-bound structures of SimR become more pronounced due to a twisting motion at the dimer interface. As a result, dimer-upon-dimer superpositions (using core residues) gave larger values of 2.13 and 2.04 Å for SimR-SD8 *versus* SimR-apo and SimR–SC4 *versus* SimR-apo, respectively. This twisting motion is more noticeable in comparisons where only one subunit in each dimer is used for the superposition (Fig. 6). When viewed from the side of the dimer, one subunit rotates relative to the other by around 8°. The axis of rotation passes through equivalent points in each subunit roughly at the interface between the LBD and DBD just above the midpoint of α3. Because the pivot point is close to the DBD, this small rotation has very little effect on the separation of the recognition helices (Fig. 6); values of 41.7 and 42.6 Å were obtained for SimR-SD8 and SimR–SC4, respectively, being comparable to the value of 42.3 Å obtained for the SimR-apo structure (Fig. 2). The upward motion of the α9–α10 arm upon ligand binding is, in part, correlated with the movement of the adjacent subunit (Figs. 5 and 6). This is because each arm is closely associated with the opposing subunit and many of the interactions are retained, including several hydrogen bonds at the extremities of the arm (C-terminal end of α9 and N-terminal end of α10).

As a result of the subunit rotation, the C-termini move closer together. For example, the separation of the terminal residues of α11 (i.e., Arg243) decreases from 17.6 to 10.1 Å upon SD8 binding (Fig. 6); being remote from the pivot point, conformational differences are more exaggerated here. In addition, there is more subunit overlap, such that there is a modest increase in the dimer interface area from 2475 to 2759 $Å^2$ upon SD8 binding. Nonbonded protein–protein interactions across the interfaces are extensive in both conformational states, although there is a slight increase in the number of pairwise interatomic distances under 4 Å, from 297 in SimR-apo to 336 in SimR–SD8. More significant is the increase in the number of protein–protein hydrogen bonds, from 18 in SimR-apo to 30 in SimR–SD8 (values from PDBsum§).[33] Further analysis indicates that 12 of these are retained between the two structures. These comprise contacts between the α9–α10 arm of one subunit and the α6–α7 loop and α7 of the other, and between the α6–α7 loop of one subunit and α8 of the other.

§ http://www.ebi.ac.uk/pdbsum/

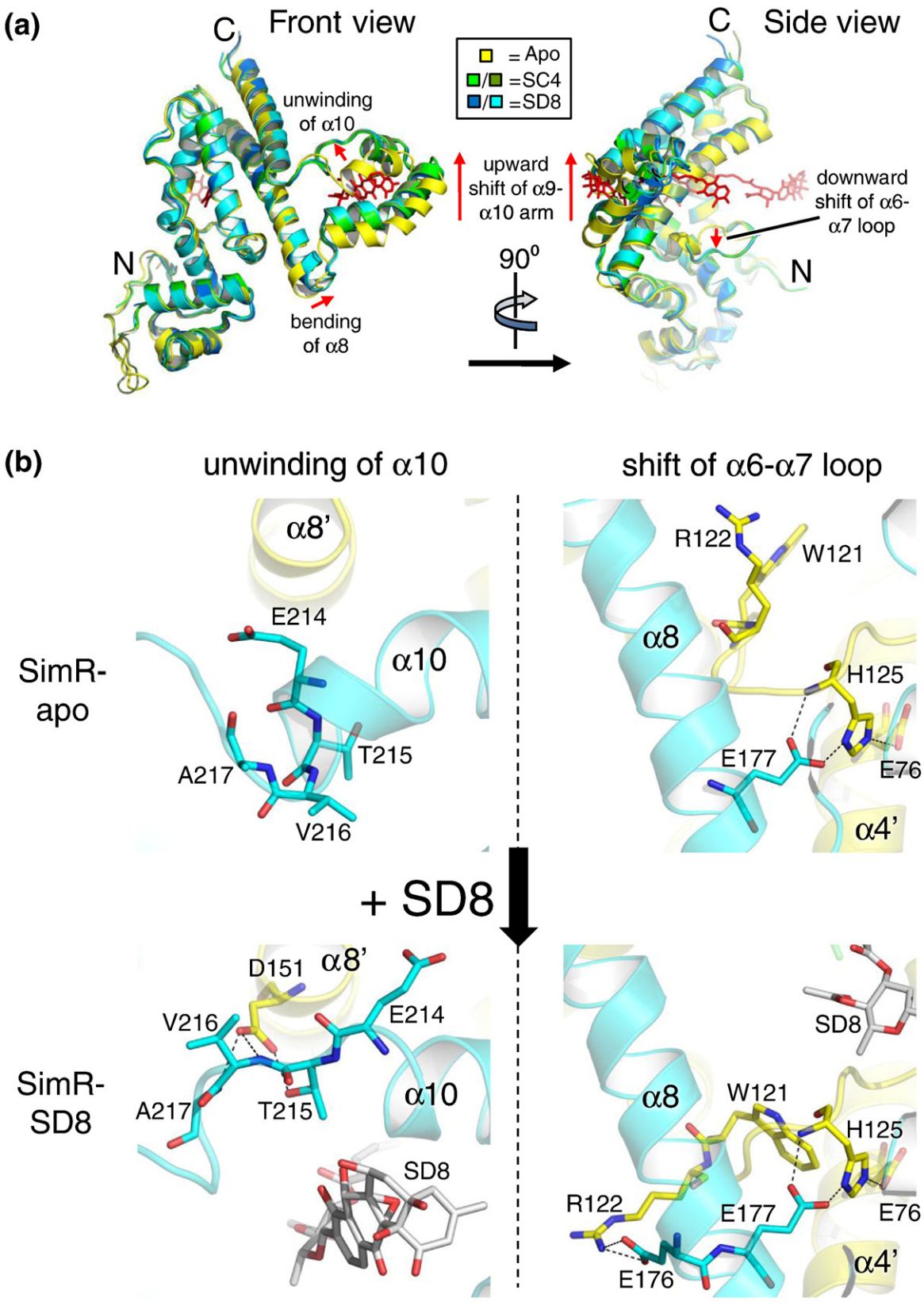


Fig. 5 (*legend on next page*)

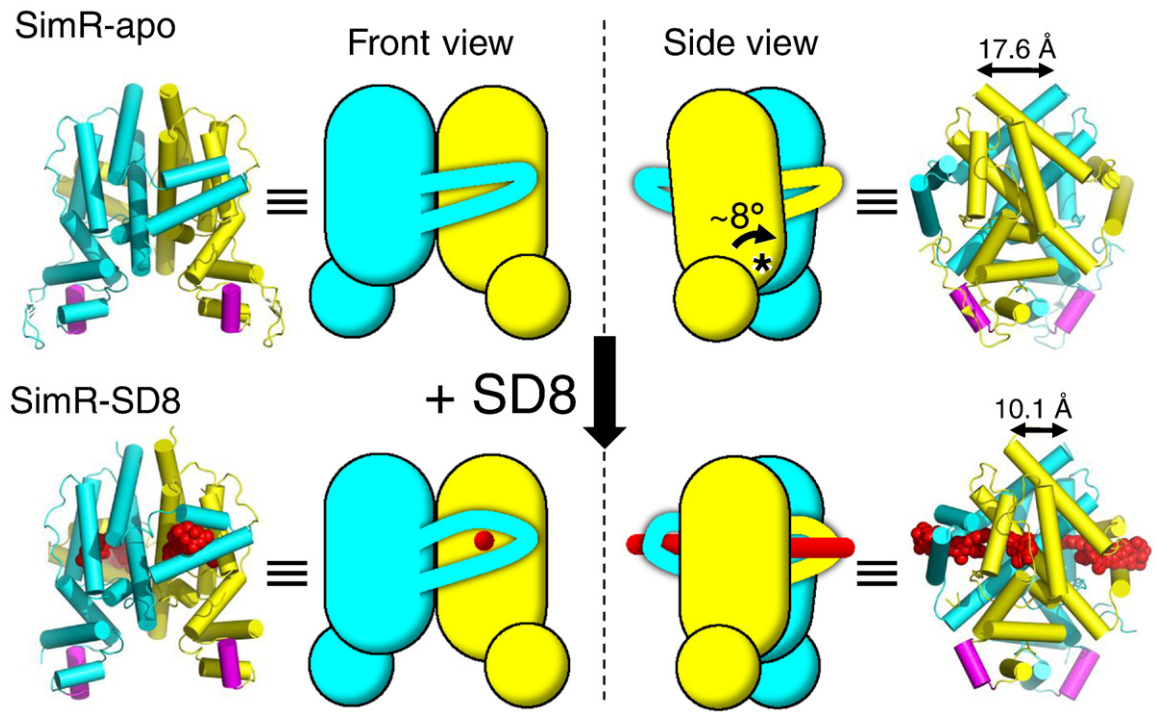


Fig. 6. Rotation of SimR subunits stabilised by ligand binding. Structures of SimR-apo and SimR–SD8 together with schematic representations illustrating the rotation of the subunits relative to one another. For this figure only, the 'front view' shows the dimer with the plane of the interface perpendicular to the paper (it differs from the 'front view' used elsewhere by a ~25° rotation around the vertical axis). In order to emphasize the subunit rotation, the cyan-coloured subunits are shown fixed in the same relative orientations. This can be clearly seen in the side view where the yellow subunit rotates by ~8° relative to the cyan subunit; the approximate pivot point is indicated by the asterisk. The distances separating the terminal residues of α11 (i.e., Arg243) in the two structures are marked.

The simocyclinone binding site in SimR

The SD8 and SC4 ligands bind to the same sites in their respective complexes with SimR, such that moieties common to both molecules occupy equivalent positions in the binding pocket (Fig. 4; Supplementary Fig. S3). Thus, only a detailed description of the protein–ligand interactions for SimR–SD8 is necessary. The SD8 molecule is bound in an extended conformation spanning almost 30 Å, making substantial contacts with both subunits of the homodimer (Fig. 7; Supplementary Fig. S3). The AC end is completely buried in the core of the SimR dimer, in a pocket formed by residues from α4, α5, α7, and α8 and several of the intervening loops from one SimR monomer. The α6–α7 loop adopts a different conformation in the ligand-bound structures, thereby avoiding a steric clash between Trp121 and the AC moiety. Otherwise, the AC pocket is essentially preformed at the protein backbone level in SimR-apo, only requiring a few side chains to reposition themselves, most notably, Gln135, Gln136, and Asn161. By contrast, the PK end is partially exposed to bulk solvent in a peripheral site that is formed almost exclusively by helices α9 and α10 (i.e., the extended arm) of the other monomer. The lower part of the PK binding site is present in SimR-apo and corresponds to the 'saddle' resulting from the kink in α9 (Fig. 2). The upper part of the site is completed in the ligand-bound form as the result of the conformational changes in the α9–α10 arm described below, principally the unwinding of α10. In between the AC and PK moieties, the cavity widens considerably

Fig. 5. Conformational differences between SimR-apo and SimR–SD8. (a) Superposition of all crystallographically independent subunits from all three SimR structures (five in total) based only on the core residues (i.e., not including the α9–α10 arms). Only the left-hand subunits from Fig. 2 are shown. (b) Close-up views illustrating the conformational changes in α10 and in the α6–α7 loop. Also shown in the right-hand panels is the Glu177–His125–Glu76 hydrogen-bonding network that is present in both structures. The SD8 molecules are shown in stick representation in red (a) or grey (b) carbons (see also Supplementary Fig. S4).

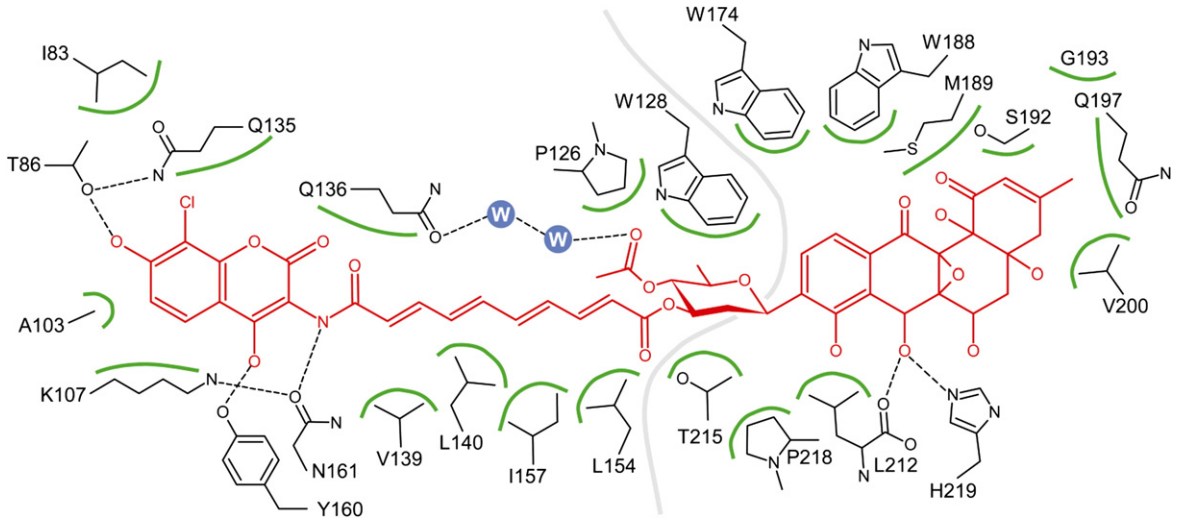


Fig. 7. Schematic representation of the SD8-binding pocket of SimR. The ligand is shown in red and all residues within 4 Å of it are shown in black. Dotted lines indicate hydrogen bonds, and green arcs indicate nonbonded interactions. The two water molecules (labelled 'W') that link Gln136 to the olivose moiety of SD8 are shaded blue. The grey line delineates the boundary between the two subunits. For clarity, all hydrogens have been omitted.

(Fig. 4). This is where the linker and the olivose sugar are accommodated. What is particularly striking about the binding of SD8 is that despite the ligand having 19 atoms that could either donate or accept a hydrogen bond, it makes a maximum of only five direct hydrogen bonds to SimR (Fig. 7; Supplementary Fig. S3). However, this is compensated by the extensive van der Waals contacts with the protein.

When comparing the SimR–SD8 and SimR–SC4 structures, we observed several minor differences due to the absence of the AC moiety in SC4. The side chains of Gln135 and Gln136 in SimR–SC4, which are displaced in SimR–SD8 due to the presence of the AC moiety, adopt conformations similar to those found in SimR-apo (Supplementary Fig. S3). Similarly, the α4–α5 surface loop differs slightly in SimR–SC4 due to the lack of an interaction between Thr86 and the ligand. Since these small differences appear to influence neither the dimer interface nor the DBD, they are unlikely to explain why SC4 is less effective than SD8 at derepressing SimR *in vitro*.[25] Rather, we speculate that this is simply a consequence of the fewer favourable interactions that SC4 makes with the protein due to the lack of an AC moiety, which is likely to result in a reduced binding affinity of SC4 for SimR.

SD8 is a potent inhibitor of DNA gyrase and functions by a novel mechanism, binding to the GyrA subunit and preventing DNA binding.[20] Recently, the crystal structure of the GyrA–SD8 complex was determined, revealing the structural basis of enzyme inhibition.[19] SD8 binds to GyrA in an extended conformation, with the PK and AC moieties occupying two separate binding pockets. As might be expected, there is no correspondence between the ligand-binding pockets in GyrA and SimR.

In free solution, the tetraene linker of SD8 would be predicted to be linear, due to the conjugated double bond system, as is observed in GyrA–SD8.[19] However, in SimR–SD8, the tetraene linker of SD8 is curved. This presumably occurs because the relative orientations of the AC and PK pockets are incompatible with the linker remaining linear. Even in SimR–SC4, the tetraene linker exhibits some curvature.

Comparison to other TetR family ligand-binding sites

The LBDs of TFRs are significantly more variable than their DBDs, reflecting the plethora of ligands that can be bound.[8] Moreover, the manner in which ligands are bound is highly variable. Even in comparisons between SimR–SD8 and the ligand-bound forms of the structurally closely related ActR and TetR, it is clear that there is little correspondence when the structures are superposed (Fig. 4b). The long axis of actinorhodin crosses that of SD8 at an angle of roughly 70°, while the majority of tetracycline aligns loosely with the tetraene linker, rather than with one of the ring systems of SD8. Where structures of TFRs in complex with their cognate ligands have been determined, the cognate ligands are bound predominantly by one subunit of the biological dimer: in ActR there is only one contact with the second subunit, while in TetR there is none (Fig. 3). By contrast, simocyclinone makes substantial contacts with both subunits of the dimer (Figs. 2, 3, 4, and 7; Supplementary Fig. S3), with the

majority of the contacts to the second subunit involving the α9–α10 arm. The extent of the SD8 binding pocket is also unusual; the only other example showing a pocket of comparable length is the palmitoyl–CoA complex of DesT (PDB code 3LSJ).[11] However, in this case, the pocket entrance is at the top of the subunit and the ligand lies roughly perpendicular to SD8, interacting mainly with a single subunit (Supplementary Fig. S2e).

Conformational changes in SimR captured upon ligand binding

Available evidence suggests that apo-TFRs sample a range of conformations and that ligand-binding captures one of these conformations, rather than inducing the conformational change.[8,34] As a consequence, we have assumed here that the same is true of SimR.

In a superposition of the subunits from the various SimR structures, the upward shift of the α9–α10 arm in the ligand-bound form is the most striking difference (Fig. 5a). Presumably, the α9–α10 arm is carried along with the neighbouring subunit when the two subunits rotate relative to one another as part of the normal dynamic behaviour of the system, and this conformational state is captured on ligand binding (Fig. 6). However, the arm does not move entirely as a rigid body: although α9 is essentially unchanged, the C-terminal end of α10 partially unwinds (it becomes two residues shorter), and the now extended α10–α11 loop moves even further upward, causing the remainder of α10 to tilt upward. The loop is stabilised in its new conformation by additional intersubunit hydrogen bonds involving Asp151 in α8 (Fig. 5b; Supplementary Fig. S4).

The net effect of these motions is that the channel connecting the central cavity to bulk solvent in SimR-apo closes, and a new one opens adjacent to it, which is wide enough to accommodate the PK moieties of SD8 and SC4 (Fig. 4). A more subtle effect of the upward arm motion is that the C-terminal end of α8 bends towards the opposing subunit, leading to a number of minor conformational changes in the α8–α9 loop.

In all three SimR structures, there are two symmetry-related hydrogen-bonding networks that lie on either side of the pivot and link Glu177 in α8 of one subunit to His125 in the α6–α7 loop of the other subunit, which in turn is hydrogen-bonded to Glu76 in α4 of the same subunit (Fig. 5; Supplementary Fig. S4). These networks may be important for tethering the subunits to one another and restraining intersubunit rotations.

The portion of the α6–α7 loop C-terminal to His125 remains unchanged, whereas the N-terminal portion moves downward in the ligand-bound structures. At the protein backbone level, the motion seems relatively minor. However, the side chains of two bulky residues, namely, Trp121 and Arg122, flip downward, with the tip of the latter moving by some 17 Å into the dimer interface, where it is salt-bridged to Glu176 in α8 of the opposing subunit (Fig. 5; Supplementary Fig. S4), thereby reciprocating the link between α8 and the α6–α7 loop, involving Glu177 and His125, described above. In SimR-apo, the two monomer cores (i.e., excluding the α9–α10 arms) present relatively flat surfaces to one another, and thus could potentially slide past each other without steric restraint. However, in the ligand-bound form, the two copies of Arg122 project across the dimer interface into pockets in the surfaces of the opposing monomers (Fig. 8). Thus, they may act as retaining pins, effectively locking the subunits together.

The mechanism of derepression in SimR

It is widely accepted that the derepression of TFRs involves allosteric mechanisms, since their ligand-binding sites are remote from their DBDs.[1,8] It may be significant that the majority of characterized TFR ligand-binding sites are contained almost entirely within individual subunits, and that for those TFRs for which ligand-bound and DNA-bound structures are available, the allosteric mechanism involves conformational changes transmitted largely within the same subunit.[6,7,9–11] The binding of ligands to these sites captures a conformational state in which the DBD (the recognition helix in particular) is repositioned relative to the LBD such that the dimer is prevented from binding to cognate DNA. Nevertheless, intersubunit communication appears to play a role in some systems (e.g., QacR[9] and CgmR)[10] as the binding of ligands by one subunit can render the other subunit incapable of doing the same. In SimR, the binding of SD8 stabilises a number of conformational changes, the most significant of which occur at the dimer interface. These changes have the effect of repositioning the two subunits of the dimer relative to one another and locking them in this new configuration. It seems likely that the side chain of Arg122 in the α6–α7 loop is important for this locking mechanism (Fig. 8). Moreover, it also seems likely that further rigidity is imparted on the system by the ligand, which threads through both subunits. Indeed, the induced curvature of the tetraene linker of bound SD8 may be indicative of the forces required to maintain the protein in the derepressed state.

In the derepression mechanism of TetR, the principal changes upon ligand binding are an upward movement of the α6–α7 loop (i.e., away from the DBD), a partial unwinding of α6, and a 'pendulum-like' swing of α4 around its C-terminal end that repositions the DBD relative to the LBD such that the two recognition helices move further apart to a configuration that would abrogate binding to DNA.[6,7,35–37] By contrast, in SimR, although there

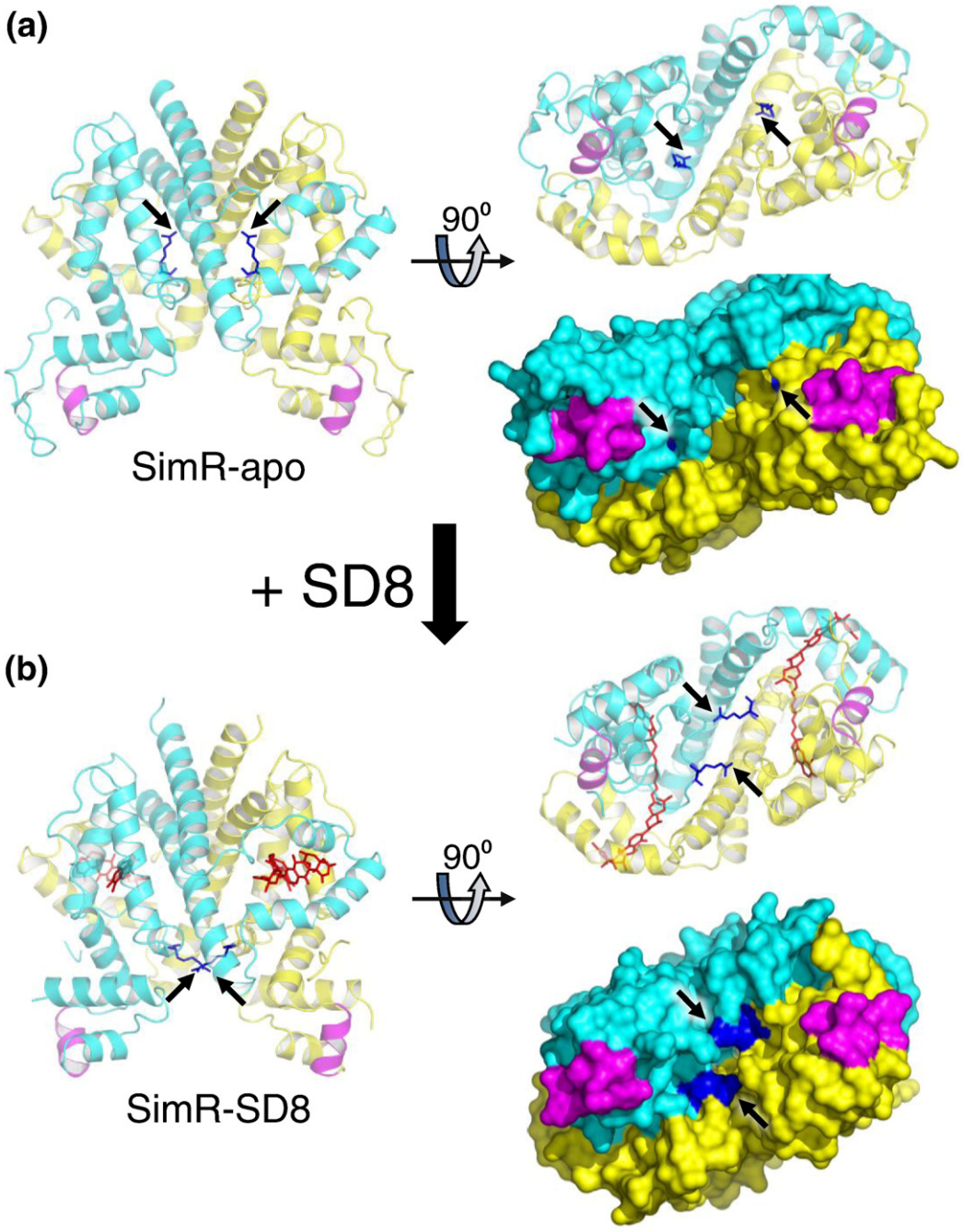


Fig. 8. Putative derepression mechanism of SimR. Semi-transparent cartoon and molecular surface representations of (a) SimR-apo and (b) SimR–SD8. The left-hand panels show the front view and the right-hand panels show the bottom view. The symmetry equivalent Arg122 residues are coloured blue (and shown in stick representation for the cartoons) and the ends of the side chains are indicated by the black arrows; SD8 molecules are shown as red sticks. In SimR–SD8, the interdigitation of the Arg122 side chains is most apparent in the right-hand panels.

are changes in the α6–α7 loop, it moves down instead of up, and there is no unwinding of α6. Furthermore, α4 does not swing at all, and thus the relative dispositions of the LBDs and DBDs within individual subunits of SimR remain essentially unchanged. Therefore, in both SimR-apo and SimR–SD8, the recognition helices are too far apart to allow SimR to bind DNA (Fig. 2). This is also true for the apo and ligand-bound states of TetR[37] and ActR.[13]

How does ligand binding prevent SimR from binding to DNA? It has been assumed that the

ligand-free form of SimR is sufficiently dynamic in solution that it can readily sample different conformational states. Davidson and colleagues performed equilibrium unfolding experiments on wild-type TetR and proposed that the thermodynamic coupling of TetR domains, in particular, the rigidification of the DBDs upon ligand binding, underlies the allosteric response in TetR.[34] For SimR, we speculate that a DNA-binding conformation could be achieved largely through rigid-body motions of the subunits about a pivot point towards the upper end of the LBD, without invoking significant repositioning of the DBDs relative to the LBDs. Such a conformational change would be disfavoured in the ligand-bound structures due to a combination of factors, including the rigidity imposed on the system by the ligand threading through both subunits and the increased number of favourable intersubunit contacts, in particular, the interdigitation of the Arg122 side chains (Fig. 8), together preventing DNA binding by SimR–SD8.

Conclusion and summary

SimR is a TetR-family transcriptional regulator that provides a mechanism that couples the biosynthesis of the antibiotic simocyclinone to its export in the producing organism *S. antibioticus*. We have determined the crystal structures of the protein in the absence of ligands and in complexes with the mature antibiotic SD8 and with the biosynthetic intermediate SC4. SimR displays the α-helical fold typical of TFRs and exists as a homodimer. In addition to the usual nine helices, the SimR monomer has an insertion of two extra helices, prior to the C-terminal helix, which together form an arm that wraps around the opposing subunit. This feature is only found in a subfamily of TFRs that includes ActR and TetR, which are close structural homologues of SimR.[8] Furthermore, SimR has an N-terminal, Arg-rich β-hairpin motif that is not observed in other TFR structures, which appears to be ideally placed to engage with phosphate groups in the minor groove of DNA.

The ligand-bound structures of SimR reveal an extended binding pocket that, unusually for TFRs, spans both subunits in the functional dimer. Interactions with one subunit almost exclusively involve the α-helical arm, which is reconfigured in the ligand-bound form. Through comparisons of these structures, we are able to postulate a derepression mechanism for SimR that involves rigid-body motions of the subunits relative to one another, coupled with a locking mechanism that most likely prevents further conformational change, holding the DBDs too far apart for DNA binding. SimR vividly illustrates the versatility of TFRs in their ability to recognise and respond to diverse ligands.

Materials and Methods

Protein overexpression and purification

The *simR* gene of *S. antibioticus* Tü 6040 encoding a 259-amino-acid protein was chemically synthesized with codon optimization (Genscript) for expression in *E. coli* and was subsequently engineered for expression with either an N-terminal or a C-terminal His_6-tag for nickel-affinity purification. Construction of the vector for expression of N-terminally tagged protein, pIJ10495, has been described.[38] This encodes a protein with an additional 26 amino acids at the N-terminus of the native sequence (with sequence MK-H_6-PMSDYDIPTTEN-LYFQGA), giving a total molecular weight of 32,222. For expression with a C-terminal tag, the gene was amplified by PCR using an upstream primer carrying an NdeI site [R2-Cter-shrt-F: 5′-GATCATATGAACGAAAAC-GAACCGGTGAG-3′] and a downstream primer carrying an XhoI site [R2-Cter-shrt-R: 5′-GATCTCGAGCGC-CAGCGCCGGGCGTTCGC-3′]. The amplified DNA fragment was 5′-phosphorylated, cloned into SmaI-cut pUC18, and verified by DNA sequencing. The *simR* coding sequence was excised by NdeI/XhoI digestion and cloned into NdeI–XhoI-cut pET20b, giving rise to the overexpression plasmid (pIJ10499). This encodes a protein with an additional 8 amino acids at the C-terminus of the native sequence (with sequence LE-H_6), giving a total molecular weight of 30,197.

For N-terminally His-tagged SimR expression, vector pIJ10495 was introduced into *E. coli* BL21(pLysS), and a 7-ml overnight culture was used to inoculate 500 ml of Luria-Bertani medium containing 50 mg kanamycin and 15 mg chloramphenicol. The cells were grown at 37 °C to OD_{600} of ~0.4. The culture was cooled to 20 °C, and 0.7 ml of dimethylsulfoxide (DMSO) was added before induction of protein expression by the addition of IPTG to a final concentration of 0.3 mM. The addition of DMSO was essential for a high yield of soluble SimR, presumably because it artificially induces the expression of chaperones that assist in the folding of SimR *in vivo*. The culture was incubated with shaking for 3 h at 30 °C. Harvested cells were resuspended in 50 mM Tris-HCl (pH 8.4), 300 mM NaCl, 5% (v/v) glycerol, containing a Complete EDTA-free protease inhibitor cocktail (Roche), and lysed by sonication (three cycles of 20 s with 40-s resting on ice in between each cycle). The cell debris was removed by centrifugation at 84,000*g* for 30 min, and the supernatant was filtered through a 0.45-μm membrane before being applied to a 1-ml Ni-loaded Hi-Trap Chelating HP column (GE Healthcare) that had been equilibrated with buffer A [50 mM Tris-HCl (pH 8.4), 300 mM NaCl, 50 mM imidazole]. Protein was eluted from the column using an increasing (50 to 500 mM) imidazole gradient in the same buffer. SimR fractions were identified using SDS-PAGE, pooled together, and concentrated to approximately 5–10 mg ml^{-1} using a Vivaspin 6 10-kDa cutoff concentrator (Vivascience). The concentrated protein was immediately exchanged into crystallization buffer [25 mM Tris-HCl (pH 8.4), 300 mM NaCl] using a Zeba desalting micro-column (Thermo Scientific).

Selenomethionine-(SeMet)-labelled SimR was obtained by the metabolic inhibition method.[39] *E. coli* BL21(pLysS) cells containing pIJ10495 were grown in 2 l of M9 medium

at 37 °C. When OD_{600} reached ~0.4, lysine, phenylalanine, and threonine were added to a final concentration of 100 mg l^{-1} each, and isoleucine, leucine, valine, and SeMet were added to a final concentration of 50 mg l^{-1} each. The culture was incubated for a further 30 min at 37 °C and then induced with 1 mM IPTG for 3 h. The purification of SeMet-labelled SimR followed the protocol described above except that all buffers were degassed and contained 0.5 mM DTT to prevent protein oxidation.

C-terminally His-tagged SimR was produced from the overexpression plasmid pIJ10499 using the same procedure as for the N-terminally His-tagged protein, except that the medium was supplemented with 25 mg carbenicillin in place of kanamycin.

Table 1. Summary of SimR X-ray data and model parameters

Data set[a]	Native apo (N-tag)	SeMet apo (N-tag)	Native apo (C-tag)	SD8 complex (C-tag)	SC4 complex (C-tag)
Data collection					
Space group	*H*32	*H*32	*H*32	*C*2	$P6_522$
Cell parameters (Å/°)	$a=b=116.62$, $c=110.58$	$a=b=116.11$, $c=110.35$	$a=b=116.06$, $c=109.38$	$a=147.37$, $b=83.91$, $c=45.47$, $\beta=90.94$	$a=b=86.57$, $c=349.86$
Solvent content (%)	45	45	48	47	61
Beamline[b]	I03	I03	I02	I04	I02
Wavelength (Å)	0.9763	0.9763	0.9700	0.9400	0.9700
Resolution range[c] (Å)	48.50–2.30 (2.42–2.30)	48.37–3.40 (3.58–3.40)	48.04–1.95 (2.06–1.95)	42.39–2.30 (2.42–2.30)	63.06–2.30 (2.42–2.30)
Unique reflections[c]	12,693 (1786)	4076 (585)	20,761 (3005)	23,951 (2988)	34,449 (4983)
Completeness[c] (%)	98.0 (95.2)	99.9 (100.0)	99.8 (100.0)	97.0 (83.4)	97.3 (98.5)
Redundancy[c]	5.0 (3.8)	15.2 (15.7)	7.8 (7.9)	2.9 (2.3)	8.4 (8.5)
R_{merge}[c,d]	0.093 (0.523)	0.158 (0.422)	0.070 (0.570)	0.077 (0.229)	0.109 (0.372)
Mean $I/\sigma(I)$[c]	10.2 (2.3)	14.4 (6.8)	20.6 (3.8)	10.3 (3.3)	15.4 (5.4)
Wilson B value ($Å^2$)	57.9	78.4	31.8	30.7	28.8
Refinement					
Reflections: working/free[e]			19,717/1044	22,723/1227	32,723/1726
R_{work}[f]	—	—	0.205	0.200	0.198
R_{free}[f]	—	—	0.234	0.270	0.246
Coordinate error[g] (Å)	—	—	0.149	0.265	0.210
Ramachandran favoured/allowed[h] (%)	—	—	97.2/98.8	98.2/99.8	99.0/100.0
Ramachandran outliers[h]	—	—	3	1	0
rmsd bond distances (Å)	—	—	0.018	0.017	0.017
rmsd bond angles (°)	—	—	1.57	1.46	1.58
Contents of model	—	—			
Protein residues in each chain (total no. of residues)	—	—	A: 7–244 (238)	A: 5–11, 27–246 (227)	A: 6–247 (242)
				B: 8–10, 28–247 (223)	B: 5–15, 26–246 (230)
SD8/SC4 molecules	—	—	0	2	2
PEG molecules	—	—	0	0	2
Calcium ions	—	—	0	0	1
Chloride ions	—	—	0	6	8
Water molecules	—	—	121	164	277
Temperature factors ($Å^2$)					
Main-chain atoms			36.9	24.4	19.3
Side-chain atoms			39.9	25.6	21.6
SD8/SC4 molecules			—	23.8	23.7
PEG molecules			—	—	37.2
Calcium ions			—	—	24.3
Chloride ions			—	60.0	37.9
Water molecules			20.4	22.8	25.1
Overall			37.3	24.9	21.0
PDB accession code	—	—	2Y2Z	2Y30	2Y31

[a] Protein was either N-terminally (N-tag) or C-terminally (C-tag) His tagged.
[b] I02, I03, I04: beamlines at the Diamond Light Source (Oxfordshire, UK).
[c] The figures in parentheses indicate the values for outer resolution shell.
[d] $R_{merge}=\sum_{hkl}\sum_i|I_i(hkl)-\langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the ith observation of reflection hkl and $\langle I(hkl)\rangle$ is the weighted average intensity for all observations i of reflection hkl.
[e] The data sets were split into 'working' and 'free' sets comprising 95% and 5% of the data, respectively. The free set was not used for refinement.
[f] The R-factors R_{work} and R_{free} are calculated as follows: $R=\sum(|F_{obs}-F_{calc}|)/\sum|F_{obs}|\times100$, where F_{obs} and F_{calc} are the observed and calculated structure factor amplitudes, respectively.
[g] Estimate of the overall coordinate errors calculated in REFMAC5 based on R_{free}.[48]
[h] Calculated using MOLPROBITY.[52]

Protein crystallization and cryoprotection

Crystallization screens and optimizations were performed as described,[38] using a protein concentration of approximately 5 mg ml^{-1} and a temperature of 20 °C throughout. Crystals of N-terminally His-tagged SimR-apo (both native and SeMet labelled) were obtained from 2% (w/v) polyethylene glycol (PEG)-10000, 0.2 M ammonium acetate in 0.1 M Bis-Tris (pH 5.5). Subsequently, C-terminally His-tagged protein was found to yield better-quality crystals from the same conditions. All SimR-apo crystals were cryoprotected by supplementing the crystallization solution with 30% (w/v) PEG-1500.

Simocyclinones D8 and C4 were isolated from *S. antibioticus* Tü 6040 as described.[21,22] Both compounds were dissolved in DMSO and then diluted in crystallization buffer [25 mM Tris-HCl (pH 8.4), 300 mM NaCl] to a concentration of 500 μM prior to mixing with C-terminally His-tagged SimR: an equal volume of 10 mg ml^{-1} protein was slowly titrated into the above solution, with thorough mixing, to give a final concentration of 5 mg ml^{-1} SimR. The final concentration of DMSO did not exceed 1% (v/v). Crystals of the SimR–SD8 complex grew from 25% (w/v) PEG-1000 in 0.1 M sodium acetate (pH 4.6) and were cryoprotected by supplementing the crystallization solution with 32% (w/v) PEG-1500. Crystals of the SimR–SC4 complex grew from 28% (v/v) PEG-400, 0.2 M $CaCl_2$ in 0.1 M Hepes (pH 7.5). Due to the presence of PEG-400 in the mother liquor, no further cryoprotection was required for these crystals.

Structure determination and refinement

All crystals were flash-cooled by plunging into liquid nitrogen and then mounted in Unipuck cassettes before being robotically mounted onto the goniostat on station I02, I03, or I04 at the Diamond Light Source (Oxford, UK) and were maintained at −173 °C with a Cryojet cryocooler (Oxford Instruments). Diffraction data were recorded using an ADSC Quantum 315 CCD detector. The resultant data were integrated using MOSFLM[40] and subsequently scaled by SCALA.[41] Native data were collected from an apo crystal of the N-terminally His-tagged SimR to 2.3-Å resolution in space group *H*32, as described.[38] Solvent content estimation gave a value of 45% based on a single copy of the monomer in the ASU. A single apo SeMet-labelled crystal of N-terminally His-tagged SimR was used to collect a selenium anomalous peak data set at 3.4-Å resolution with the wavelength set to 0.9763 Å. Experimental phases were determined by the single-wavelength anomalous dispersion method using the SHELX suite[42] implemented through the HKL2MAP graphical user interface.[43] Substructure solution was performed with SHELXD, and seven of the expected nine selenium sites in the native sequence were located. This solution gave a figure-of-merit of 0.541- to 3.4-Å resolution after phasing in SHELXE. After density modification with PARROT,[44] a partial backbone trace was obtained with BUCCANEER.[45] At the same time, an approximate molecular replacement solution was obtained from the 2.3-Å-resolution native data set using the BALBES pipeline[46] (with a molecular replacement score of 2.63), which selected the structure of Sco6599 as a template (PDB accession code 2HXI; sequence identity with SimR, 25.4%). The BUCCANEER model and the BALBES solution were then brought to the same origin after visual inspection in COOT.[47] The BALBES model was subsequently rebuilt against the experimentally phased electron density map at 3.4-Å resolution and then refined against the 2.3-Å-resolution native data set using REFMAC5.[48] At this stage, the model contained only 118 residues and had R_{work} and R_{free} values of 50.6% and 51.7%, respectively. The refined model phases were then combined with the experimental phase information using the program SIGMAA[49] to give improved phases. Further iterations of rebuilding, refinement, and phase combination were performed until a model consisting of 165 residues was obtained, with corresponding R_{work} and R_{free} values of 36.3% and 40.0%, respectively. Thereafter, the model was refined against a new data set (but for C-terminally His-tagged SimR-apo instead) collected to 1.95-Å resolution, whereupon the electron density maps were significantly improved. Additional cycles of rebuilding and refinement yielded a model consisting of 238 residues, with corresponding R_{work} and R_{free} values of 20.5% and 23.4%, respectively.

The SD8 and SC4 complexes with SimR each yielded new crystal forms containing two SimR protomers per ASU. In both cases, the structures were solved by molecular replacement using the structure of the C-terminally His-tagged SimR-apo monomer as the search model. The SimR–SD8 structure was solved using MOLREP,[50] and the SimR–SC4 structure using PHASER.[51] The two complex structures were then rebuilt and refined as for SimR-apo. All X-ray data collection and refinement statistics are summarized in Table 1.

Difference electron density 'omit' maps were generated for the SD8 and SC4 ligands using phases calculated from the final model minus the ligand coordinates after simulated annealing refinement (Fig. 4; Supplementary Fig. S3). This was performed from a starting temperature of 5000 K after applying random shifts to the model ('shake' term set to 0.3) using PHENIX.[53] Structural figures were generated using PyMOL.[54]

Accession numbers

Coordinates and structure factors for the three SimR structures described herein have been deposited in the Protein Data Bank with accession numbers 2Y2Z, 2Y30, and 2Y31.

Supplementary materials related to this article can be found online at doi:10.1016/j.jmb.2011.02.035

Acknowledgements

We are grateful to Mike Merrick and Ray Dixon for their comments on the manuscript, to Justin Nodwell for his advice and encouragement, and to beamline scientists at the Diamond Light Source for assistance with X-ray data collection. This work was supported by a John Innes Centre Rotation Studentship (to T.B.K.L.), by BBSRC grant BB/I002197/1 (to M.J.B. and D.M.L.), and by the BBSRC Core Strategic Grant to the John Innes Centre.

References

1. Ramos, J. L., Martinez-Bueno, M., Molina-Henares, A. J., Teran, W., Watanabe, K., Zhang, X. *et al.* (2005). The TetR family of transcriptional repressors. *Microbiol. Mol. Biol. Rev.* **69**, 326–356.
2. Hillen, W. & Berens, C. (1994). Mechanisms underlying expression of Tn10 encoded tetracycline resistance. *Annu. Rev. Microbiol.* **48**, 345–369.
3. Bertrand, K. P., Postle, K., Wray, L. V., Jr. & Reznikoff, W. S. (1983). Overlapping divergent promoters control expression of Tn10 tetracycline resistance. *Gene*, **23**, 149–156.
4. Meier, I., Wray, L. V. & Hillen, W. (1988). Differential regulation of the Tn10-encoded tetracycline resistance genes *tetA* and *tetR* by the tandem *tet* operators O1 and O2. *EMBO J.* **7**, 567–572.
5. Lederer, T., Takahashi, M. & Hillen, W. (1995). Thermodynamic analysis of tetracycline-mediated induction of Tet repressor by a quantitative methylation protection assay. *Anal. Biochem.* **232**, 190–196.
6. Orth, P., Saenger, W. & Hinrichs, W. (1999). Tetracycline-chelated Mg^{2+} ion initiates helix unwinding in Tet repressor induction. *Biochemistry*, **38**, 191–198.
7. Orth, P., Schnappinger, D., Hillen, W., Saenger, W. & Hinrichs, W. (2000). Structural basis of gene regulation by the tetracycline inducible Tet repressor-operator system. *Nat. Struct. Biol.* **7**, 215–219.
8. Yu, Z., Reichheld, S. E., Savchenko, A., Parkinson, J. & Davidson, A. R. (2010). A comprehensive analysis of structural and sequence conservation in the TetR family transcriptional regulators. *J. Mol. Biol.* **400**, 847–864.
9. Schumacher, M. A., Miller, M. C., Grkovic, S., Brown, M. H., Skurray, R. A. & Brennan, R. G. (2001). Structural mechanisms of QacR induction and multidrug recognition. *Science*, **294**, 2158–2163.
10. Itou, H., Watanabe, N., Yao, M., Shirakihara, Y. & Tanaka, I. (2010). Crystal structures of the multidrug binding repressor *Corynebacterium glutamicum* CgmR in complex with inducers and with an operator. *J. Mol. Biol.* **403**, 174–184.
11. Miller, D. J., Zhang, Y. M., Subramanian, C., Rock, C. O. & White, S. W. (2010). Structural basis for the transcriptional regulation of membrane lipid homeostasis. *Nat. Struct. Mol. Biol.* **17**, 971–975.
12. Alguel, Y., Meng, C., Teran, W., Krell, T., Ramos, J. L., Gallegos, M. T. & Zhang, X. (2007). Crystal structures of multidrug binding protein TtgR in complex with antibiotics and plant antimicrobials. *J. Mol. Biol.* **369**, 829–840.
13. Willems, A. R., Tahlan, K., Taguchi, T., Zhang, K., Lee, Z. Z., Ichinose, K. *et al.* (2008). Crystal structures of the *Streptomyces coelicolor* TetR-like protein ActR alone and in complex with actinorhodin or the actinorhodin biosynthetic precursor (S)-DNPA. *J. Mol. Biol.* **376**, 1377–1387.
14. Tahlan, K., Ahn, S. K., Sing, A., Bodnaruk, T. D., Willems, A. R., Davidson, A. R. & Nodwell, J. R. (2007). Initiation of actinorhodin export in *Streptomyces coelicolor*. *Mol. Microbiol.* **63**, 951–961.
15. Baulard, A. R., Betts, J. C., Engohang-Ndong, J., Quan, S., McAdam, R. A., Brennan, P. J. *et al.* (2000). Activation of the pro-drug ethionamide is regulated in mycobacteria. *J. Biol. Chem.* **275**, 28326–28331.
16. DeBarber, A. E., Mdluli, K., Bosman, M., Bekker, L. G. & Barry, C. E., 3rd (2000). Ethionamide activation and sensitivity in multidrug-resistant *Mycobacterium tuberculosis*. *Proc. Natl Acad. Sci. USA*, **97**, 9677–9682.
17. Frenois, F., Engohang-Ndong, J., Locht, C., Baulard, A. R. & Villeret, V. (2004). Structure of EthR in a ligand bound conformation reveals therapeutic perspectives against tuberculosis. *Mol. Cell*, **16**, 301–307.
18. Zheng, J., Sagar, V., Smolinsky, A., Bourke, C., LaRonde-LeBlanc, N. & Cropp, T. A. (2009). Structure and function of the macrolide biosensor protein, MphR(A), with and without erythromycin. *J. Mol. Biol.* **387**, 1250–1260.
19. Edwards, M. J., Flatman, R. H., Mitchenall, L. A., Stevenson, C. E., Le, T. B., Clarke, T. A. *et al.* (2009). A crystal structure of the bifunctional antibiotic simocyclinone D8, bound to DNA gyrase. *Science*, **326**, 1415–1418.
20. Flatman, R. H., Howells, A. J., Heide, L., Fiedler, H. P. & Maxwell, A. (2005). Simocyclinone D8, an inhibitor of DNA gyrase with a novel mode of action. *Antimicrob. Agents Chemother.* **49**, 1093–1100.
21. Schimana, J., Fiedler, H. P., Groth, I., Sussmuth, R., Beil, W., Walker, M. & Zeeck, A. (2000). Simocyclinones, novel cytostatic angucyclinone antibiotics produced by *Streptomyces antibioticus* Tü 6040. I. Taxonomy, fermentation, isolation and biological activities. *J. Antibiot. (Tokyo)*, **53**, 779–787.
22. Holzenkampfer, M., Walker, M., Zeeck, A., Schimana, J. & Fiedler, H. P. (2002). Simocyclinones, novel cytostatic angucyclinone antibiotics produced by *Streptomyces antibioticus* Tü 6040 II. Structure elucidation and biosynthesis. *J. Antibiot. (Tokyo)*, **55**, 301–307.
23. Galm, U., Schimana, J., Fiedler, H. P., Schmidt, J., Li, S. M. & Heide, L. (2002). Cloning and analysis of the simocyclinone biosynthetic gene cluster of *Streptomyces antibioticus* Tü 6040. *Arch. Microbiol.* **178**, 102–114.
24. Trefzer, A., Pelzer, S., Schimana, J., Stockert, S., Bihlmaier, C., Fiedler, H. P. *et al.* (2002). Biosynthetic gene cluster of simocyclinone, a natural multihybrid antibiotic. *Antimicrob. Agents. Chemother.* **46**, 1174–1182.
25. Le, T. B., Fiedler, H. P., den Hengst, C. D., Ahn, S. K., Maxwell, A. & Buttner, M. J. (2009). Coupling of the biosynthesis and export of the DNA gyrase inhibitor simocyclinone in *Streptomyces antibioticus*. *Mol. Microbiol.* **72**, 1462–1474.
26. Krissinel, E. & Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **60**, 2256–2268.
27. Gouet, P., Courcelle, E., Stuart, D. I. & Metoz, F. (1999). ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics*, **15**, 305–308.
28. Krissinel, E. & Henrick, K. (2005). Detection of protein assemblies in crystals. In *Computational Life Sciences* (Berthold, M. R., Glen, R. C., Diederichs, K., Kohlbacher, O. & Fischer, I., eds), Vol. 3695, pp. 163–174. Springer-Verlag, Berlin Heidelberg.
29. Holm, L. & Sander, C. (1995). DALI: a network tool for protein structure comparison. *Trends. Biochem. Sci.* **20**, 478–480.
30. Itou, H., Okada, U., Suzuki, H., Yao, M., Wachi, M., Watanabe, N. & Tanaka, I. (2005). The CGL2612 protein from *Corynebacterium glutamicum* is a drug

resistance-related transcriptional repressor: structural and functional analysis of a newly identified transcription factor from genomic DNA analysis. *J. Biol. Chem.* **280**, 38711–38719.
31. Schumacher, M. A., Miller, M. C., Grkovic, S., Brown, M. H., Skurray, R. A. & Brennan, R. G. (2002). Structural basis for cooperative DNA binding by two dimers of the multidrug-binding protein QacR. *EMBO J.* **21**, 1210–1218.
32. Grkovic, S., Brown, M. H., Schumacher, M. A., Brennan, R. G. & Skurray, R. A. (2001). The staphylococcal QacR multidrug regulator binds a correctly spaced operator as a pair of dimers. *J. Bacteriol.* **183**, 7102–7109.
33. Laskowski, R. A. (2001). PDBsum: summaries and analyses of PDB structures. *Nucleic Acids Res.* **29**, 221–222.
34. Reichheld, S. E., Yu, Z. & Davidson, A. R. (2009). The induction of folding cooperativity by ligand binding drives the allosteric response of tetracycline repressor. *Proc. Natl Acad. Sci. USA*, **106**, 22263–22268.
35. Hinrichs, W., Kisker, C., Duvel, M., Muller, A., Tovar, K., Hillen, W. & Saenger, W. (1994). Structure of the Tet repressor-tetracycline complex and regulation of antibiotic resistance. *Science*, **264**, 418–420.
36. Kisker, C., Hinrichs, W., Tovar, K., Hillen, W. & Saenger, W. (1995). The complex formed between Tet repressor and tetracycline-Mg^{2+} reveals mechanism of antibiotic resistance. *J. Mol. Biol.* **247**, 260–280.
37. Orth, P., Cordes, F., Schnappinger, D., Hillen, W., Saenger, W. & Hinrichs, W. (1998). Conformational changes of the Tet repressor induced by tetracycline trapping. *J. Mol. Biol.* **279**, 439–447.
38. Le, T. B., Stevenson, C. E. M., Buttner, M. J. & Lawson, D. M. (2011). Crystallization and preliminary X-ray analysis of the TetR-like efflux pump regulator SimR. *Acta Crystallogr. Sect. F*, **67**, 307–309.
39. Doublié, S. (1997). Preparation of selenomethionyl proteins for phase determination. *Methods Enzymol.* **276**, 523–530.
40. Leslie, A. G. (2006). The integration of macromolecular diffraction data. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **62**, 48–57.
41. Evans, P. (2006). Scaling and assessment of data quality. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **62**, 72–82.
42. Sheldrick, G. M. (2008). A short history of SHELX. *Acta Crystallogr., Sect. A*, **64**, 112–122.
43. Pape, T. & Schneider, T. R. (2004). HKL2MAP: a graphical user interface for phasing with SHELX programs. *J. Appl. Crystallogr.* **37**, 843–844.
44. Cowtan, K. (2010). Recent developments in classical density modification. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **66**, 470–478.
45. Cowtan, K. (2006). The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **62**, 1002–1011.
46. Long, F., Vagin, A. A., Young, P. & Murshudov, G. N. (2008). BALBES: a molecular-replacement pipeline. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **64**, 125–132.
47. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **60**, 2126–2132.
48. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **53**, 240–255.
49. Read, R. J. (1986). Improved fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallogr., Sect. A*, **42**, 140–149.
50. Vagin, A. & Teplyakov, A. (2000). An approach to multi-copy search in molecular replacement. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **56**, 1622–1624.
51. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674.
52. Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X. *et al.* (2007). MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**, W375–W383.
53. Adams, P. D., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W. *et al.* (2002). PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **58**, 1948–1954.
54. DeLano, W. L. (2002). *The PyMOL User's Manual.* DeLano Scientific, San Carlos, CA, USA.

Nucleic Acids Research Advance Access published August 10, 2011

Nucleic Acids Research, 2011, 1–15
doi:10.1093/nar/gkr640

The crystal structure of the TetR family transcriptional repressor SimR bound to DNA and the role of a flexible N-terminal extension in minor groove binding

Tung B. K. Le[1,2,*], Maria A. Schumacher[2], David M. Lawson[3], Richard G. Brennan[2] and Mark J. Buttner[1]

[1]Department of Molecular Microbiology, John Innes Centre, Norwich Research Park, Norwich NR4 7UH, UK, [2]Department of Biochemistry and Molecular Biology, University of Texas M.D. Anderson Cancer Center, Houston, TX 77030-4009, USA and [3]Department of Biological Chemistry, John Innes Centre, Norwich Research Park, Norwich NR4 7UH, UK

Received June 13, 2011; Revised and Accepted July 21, 2011

ABSTRACT

SimR, a TetR-family transcriptional regulator (TFR), controls the export of simocyclinone, a potent DNA gyrase inhibitor made by *Streptomyces antibioticus*. Simocyclinone is exported by a specific efflux pump, SimX and the transcription of *simX* is repressed by SimR, which binds to two operators in the *simR-simX* intergenic region. The DNA-binding domain of SimR has a classical helix-turn-helix motif, but it also carries an arginine-rich N-terminal extension. Previous structural studies showed that the N-terminal extension is disordered in the absence of DNA. Here, we show that the N-terminal extension is sensitive to protease cleavage, but becomes protease resistant upon binding DNA. We demonstrate by deletion analysis that the extension contributes to DNA binding, and describe the crystal structure of SimR bound to its operator sequence, revealing that the N-terminal extension binds in the minor groove. In addition, SimR makes a number of sequence-specific contacts to the major groove via its helix-turn-helix motif. Bioinformatic analysis shows that an N-terminal extension rich in positively charged residues is a feature of the majority of TFRs. Comparison of the SimR–DNA and SimR–simocyclinone complexes reveals that the conformational changes associated with ligand-mediated derepression result primarily from rigid-body rotation of the subunits about the dimer interface.

INTRODUCTION

The genus *Streptomyces* accounts for the production of approximately two-thirds of the known antibiotics (1,2). By producing and expelling these compounds into their environment, these bacteria likely acquire a competitive advantage over other organisms inhabiting the same ecological niche. One such antibiotic, simocyclinone, a potent inhibitor of DNA gyrase produced by *Streptomyces antibioticus* Tü 6040 (3–5), consists of a chlorinated aminocoumarin connected to an angucyclic polyketide at the other end via a tetraene linker and a D-olivose sugar (6,7). Because antibiotics are often potentially lethal to the producing organism, there must be mechanisms to ensure that the machinery responsible for export of the mature antibiotic is in place at the time of biosynthesis. In the case of simocyclinone, such a mechanism is specified by two genes, *simR* and *simX*, embedded within the simocyclinone (*sim*) biosynthetic gene cluster (8–10). The SimR/SimX pair resembles the TetR/TetA repressor–efflux pump pair that confers resistance to clinically important tetracyclines in several human pathogens (11). Simocyclinone is exported from the producing organism by the SimX efflux pump, a member of the major facilitator superfamily. The transcription of *simX* is repressed by SimR,

*To whom correspondence should be addressed. Tel: +44 1603 450757; Fax: +44 1603 450778; Email: tung.le@jic.ac.uk
Present addresses:
Maria A. Schumacher, Department of Biochemistry, Duke University School of Medicine, Box 3711, DUMC, Durham, NC 27710, USA.
Richard G. Brennan, Department of Biochemistry, Duke University School of Medicine, Box 3711, DUMC, Durham, NC 27710, USA.

a TetR-family transcriptional regulator (TFR) that binds to two distinct operators in the intergenic region between the divergently transcribed *simR* and *simX* genes (9). Simocyclinone abolishes DNA binding by SimR, inducing expression of the SimX efflux pump, providing a mechanism that couples the biosynthesis of simocyclinone to its export (9).

TFRs are one of the major families of transcriptional regulators in bacteria (12,13). They function as homodimers, with each subunit consisting of two domains: an N-terminal DNA-binding domain (DBD) containing a helix-turn-helix (HTH) motif, and a C-terminal ligand-binding domain (LBD) (12,13). While the LBDs are diverse in amino acid sequence, reflecting the wide range of molecules to which different TFRs respond, the HTH DNA-binding motif is conserved and readily predicted bioinformatically. To date, the structures of only four TFRs bound to cognate DNA have been determined (TetR, DesT, CgmR and QacR), and it is clear that the mode of operator recognition differs from one member of the TFR family to another (14–17). For example, the tetracycline efflux pump repressor, TetR, binds as a dimer to a 15-bp operator and deforms the binding site by 17°, bending away from the protein in order to optimize the position of its HTH for specific base pair interaction (16). In contrast, the multidrug efflux pump repressor from *Staphylococcus aureus*, QacR, binds its cognate DNA site as a dimer of dimers and bends its operator by just 3°, but widens the major groove to create an optimal DNA environment for a second QacR dimer to bind cooperatively nearby (17).

Recently, we determined the structures of apo (unliganded) SimR and SimR in complex with either simocyclinone D8 (SD8) or its biosynthetic intermediate simocyclinone C4 (SC4) (18). These structures revealed the same overall domain architecture for SimR as for other TFRs, including a classical HTH motif. However, SimR possesses an additional arginine-rich N-terminal extension that precedes the core DBD, which is significantly longer than those present in the four TFRs for which protein–DNA crystal structures are available (TetR, DesT, CgmR and QacR) (Figure 1). With the exception of three residues, this 28 amino acid residue extension is disordered in both subunits in the SimR–SD8 structure, and it is only partially ordered in one subunit in the SimR–SC4 structure (18). Consistent with this, the N-terminal extension of SimR is predicted to be disordered in solution.

Here, we show by deletion analysis that the flexible N-terminal extension of SimR plays an important role in DNA binding, and we present the crystal structure of SimR bound to its operator sequence, which shows that this extension binds in the minor groove adjacent to the major groove occupied by the classical HTH motif. Although the N-terminal extension is hypersensitive to proteolysis *in vitro*, it becomes protease resistant upon binding cognate DNA. Together these data suggest that the N-terminal extension transitions from a disordered to an ordered state upon DNA binding. Bioinformatic analysis of the entire TetR family shows that an N-terminal extension rich in positively charged residues is a feature of the majority of TFRs. Finally, comparison of the SimR–DNA and SimR–SD8 complexes reveals the conformational changes required to interchange between DNA- and ligand-bound states, which largely involve rigid-body motions of the subunits relative to one another.

MATERIALS AND METHODS

Protein overexpression and purification

The *simR* gene of *Streptomyces antibioticus* Tü 6040 encoding a 259 amino acid protein was chemically synthesized with codon optimization (Genscript) for expression in *Escherichia coli* and was subsequently engineered for expression with a C-terminal hexa-histidine (His_6) tag for nickel affinity purification. Construction of the vector for expression of C-terminally tagged protein, pIJ10499, has been described previously (18). This results in a purified protein with an additional 8 amino acid residues at the C-terminus of the native sequence (with sequence LEHHHHHH), giving a total molecular weight of 30 197 Da.

For expression of N-terminally truncated SimR (lacking 10, 15, 22 or 25 amino acid residues from the N-terminus), the gene was amplified by PCR using a downstream primer carrying a XhoI site [R2-full-CtagHis-R: 5′-GAT CTCGAGCGCCAGCGCCGGGCGTTCGC-3′] and an upstream primer carrying an NdeI site

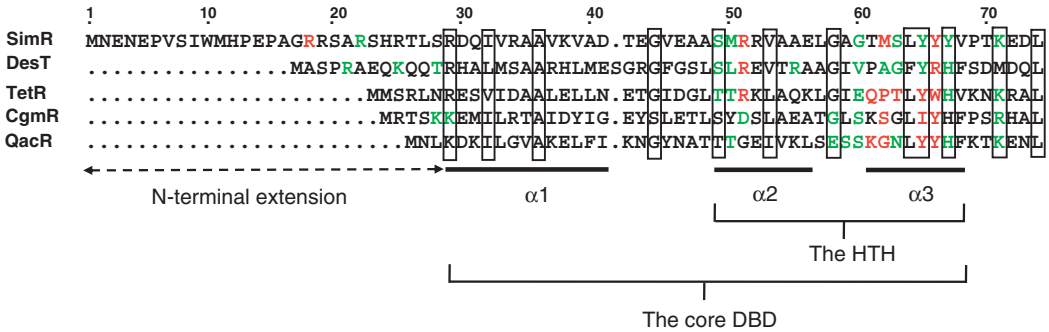


Figure 1. Alignment of the amino acid sequence of SimR with the four other TFRs for which protein–DNA crystal structures are available (DesT, TetR, CgmR and QacR), showing the HTH motif, the core DBD and the N-terminal extension present in SimR, herein termed the TFR arm. For each TFR, amino acid residues that interact with the bases of the cognate DNA operator are highlighted in red, and those that contact the phosphate backbone are highlighted in green. Conserved residues are boxed.

[R2-M10-trunc-F-NdeI (for SimR-Δ10): GCCCATATG ATGCATCCGGAACCGGCCGG; R2-A15-trunc-F-NdeI (for SimR-Δ15): GCCCATATGGCCGGTCGT CGCAGCGCGCG; R2-S22-trunc-F-NdeI (for SimR-Δ22): GCCCATATGAGCCACCGTACCCTGA GCCG; R2-T25-trunc-F-NdeI (for SimR-Δ25): GCCCA TATGACCCTGAGCCGCGATCAGATTG]. The amplified DNA fragment was 5′-phosphorylated, cloned into SmaI-cut pUC18 and verified by DNA sequencing. The *simR* alleles were excised by NdeI/XhoI digestion and cloned into NdeI–XhoI-cut pET20b, giving rise to the overexpression plasmids pIJ10500 (Δ10), pIJ10501 (Δ15), pIJ10502 (Δ22) and pIJ10503 (Δ25). All derivatives of SimR were C-terminally His-tagged and purified as described for wild-type SimR (18).

Protein crystallization and cryoprotection

Directly after nickel-affinity purification, fractions of full-length SimR were pooled and concentrated using a Vivaspin 6 30-kDa cut-off concentrator (Vivascience) to 10–12 mg ml^{-1} (~200 μM SimR dimer). The concentrated protein was exchanged into crystallization buffer [25 mM Tris–HCl (pH 8.4), 300 mM NaCl] using a Zeba desalting micro-column (Thermo Scientific). Complementary pairs of DNA oligonucleotides with different lengths (16–21 bp) and ends (blunt or sticky ends) were ordered from Oligos etc® and DNA duplexes were reconstituted by annealing oligonucleotide pairs overnight in crystallization buffer at a final concentration of 2 mM.

SimR and annealed oligonucleotides were mixed together in the ratio 1 SimR dimer to 1.2 double-stranded oligonucleotide and incubated at 20°C for 10 min before crystallization screening. Crystallization trials of SimR–DNA were set-up in hanging-drop vapour diffusion format with 48-well VDX plates (Hampton Research) using a variety of commercially available screens (Emerald BioSystems and Hampton Research) at a constant temperature of 20°C. Drops consisted of 1 μl SimR–DNA complex solution mixed with 1 μl precipitant solution and the reservoir volume was 150 μl. Improved crystals were subsequently obtained by refining the successful conditions in a hanging-drop format using 24-well VDX plates (Molecular Dimensions) over a reservoir volume of 1 ml.

SimR–DNA crystals were obtained under several different screening conditions, but only with the blunt-ended 17-mer DNA. The best crystals were obtained from solutions containing 10% (w/v) polyethylene glycol 8000, 0.2 M potassium chloride, 0.1 M magnesium acetate in 0.05 M sodium cacodylate (pH 6.5) 2 weeks after set-up. The crystals belonged to the orthorhombic space group $P2_12_12_1$. The SimR–17-mer crystals were cryoprotected by a three-step transfer process in which ethylene glycol was added to the drop to a final concentration of 20%.

Structure determination and refinement

All crystals were flash-cooled by plunging into liquid nitrogen and then mounted onto the goniostat at beamline 8.3.1 at the Advanced Light Source (Berkeley, CA, USA). The resultant data were integrated using MOSFLM (19) and subsequently scaled by SCALA (20). Native intensity data were collected from a SimR–17-mer crystal to 2.99 Å resolution. The reflections used to calculate the R-free value were selected in thin resolution shells to avoid bias resulting from the use of non-crystallographic symmetry (NCS) restraints in refinement. The structure of the complex was solved by molecular replacement using the structure of a subunit of C-terminally hexa-histidine-tagged apo SimR (PDB: 2Y2Z) and an idealized B-DNA of the correct sequence as the search models in PHASER (21). SimR–17mer crystals contained two SimR dimer–DNA complexes in the asymmetric unit. The structure of the complex was then rebuilt in COOT (22) and refined using REFMAC5 (23) and PHENIX (24) with NCS restraints. In the final stages, TLS refinement was used with a total of 20 TLS domains, which were defined using the TLS motion determination server (http://skuld.bmsc.washington.edu/~tlsmd/) (25). X-ray data collection and refinement statistics are summarized in Table 1.

Structural figures were generated using PyMOL (26). The local DNA helical parameters were calculated using Curves+ (27).

Table 1. Selected crystallographic data

Data set	SimR–DNA (17-mer)
Data collection	
Space group	$P2_12_12_1$
Cell parameters (Å/°)	a = 85.8, b = 112.6, c = 163.7
Solvent content (%)	62.5
Wavelength (Å)	1.11
Resolution range[a] (Å)	92.78–2.99 (3.15–2.99)
Unique reflections[a] (#)	31 030 (4547)
Completeness[a] (%)	95.2 (96.5)
Redundancy[a]	3.1 (3.1)
R_{merge}[b] (%)	10.0 (59.1)
$<I>/<\sigma I>$	8.3 (1.8)
Wilson B value (Å^2)	53.4
Refinement	
R_{cryst}[c] (based on 95% of data)	20.9
R_{free}[c] (based on 5% of data)	25.1
Coordinate error[d] (Å)	0.420
Ramachandran favoured[e] (%)	98.0
Ramachandran outliers[e] (%)	0.22
RMSD bond distances (Å)	0.010
RMSD bond angles (°)	1.22
Mean B-value for protein (Å^2)	57.6
Mean B-value for the DNA (Å^2)	54.6
Contents of model	
Protein residues in each chain (totals in brackets)	A: 7–241 B: 7–15 and 26–242 C: 7–242 D: 7–15 and 26–242
DNA nucleotides	E and F, G and H : 1–17
PDB accession code	3ZQL

[a]The figures in brackets indicate the values for highest resolution shell.
[b]$R_{merge} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ is the ith observation of reflection hkl and $\langle I(hkl)\rangle$ is the weighted average intensity for all observations i of reflection hkl.
[c]The R-factors R_{cryst} and R_{free} are calculated as follows: $R = \sum(|F_{obs} - F_{calc}|)/\sum|F_{obs}| \times 100$, where F_{obs} and F_{calc} are the observed and calculated structure factor amplitudes, respectively.
[d]Estimate of the overall coordinate errors calculated in REFMAC5 based on R_{free} (23).
[e]As calculated using MOLPROBITY (45).

Electrophoretic mobility shift assays

The electrophoretic mobility shift assay (EMSA) DNA probe spanning the entire *simR–simX* intergenic region (138 bp), containing both the O_X and O_R operators, was amplified by PCR and 5′-end labelled using [γ^{32}-P] ATP and T4 polynucleotide kinase (New England Biolabs). Binding of wild-type or mutated SimR to DNA was carried out in 20 µl EMSA Buffer [20 mM Tris (pH 8.0), 1 µg calf-thymus DNA, 100 mM NaCl, 8% (v/v) glycerol] containing 0.1 nM radiolabelled DNA (~8000 cpm) and varying amounts of SimR. After incubation at 22°C for 10 min, the binding reaction mixtures were loaded on 5% (w/v) native polyacrylamide gels and run in TBE buffer at 100 V for 45 min. EMSA data were collected and analysed on a PhosphoImager (FujiFilm) using Multi Gauge image analysis software (FujiFilm).

DNase I footprinting

Templates for DNase I footprinting were amplified by PCR using one unlabelled primer and one primer 5′-end labelled using [γ^{32}-P] ATP and T4 polynucleotide kinase (New England Biolabs). The primers were the same pair used to generate the *simR–simX* intergenic region probe for the EMSA experiments. DNase I footprinting assays were performed in 40 µl EMSA buffer containing ~180 000 cpm radiolabelled DNA and varying amounts of SimR. After incubation at 22°C for 10 min, 10 µl DNase I solution (10 U in 10 mM $CaCl_2$) was added and the incubation was continued for a further 60 s. Reactions were stopped by adding 140 µl DNase I stop solution (200 mM unbuffered sodium acetate, 30 mM EDTA, 0.15% SDS and 0.1 mg ml^{-1} yeast tRNA), the samples were precipitated with ethanol and the pellets were dried and dissolved in 5 µl Sequencing Loading Dye [80% (v/v) formamide, 10 mM NaOH, 1 mM EDTA, 0.1% (w/v) xylene cyanol and 0.1% (w/v) bromophenol blue]. After heating at 80°C for 3 min and cooling on ice, the samples were run on a 6% (w/v) polyacrylamide/8 M urea sequencing gel, which was dried and analysed using a PhosphoImager (FujiFilm). A G+A sequencing ladder was generated from the template DNA by chemical sequencing (28).

Limited proteolysis and protease protection assays

For limited proteolysis assays, 1 nmol of wild-type SimR was incubated with 1 pmol bovine trypsin (Sigma) in a total volume of 100 µl buffer [50 mM Tris (pH 8.0), 20 mM $CaCl_2$ and 150 mM NaCl] at 4°C. For protease protection assays, 1 nmol wild-type SimR was incubated with equimolar amounts of 15, 25 or 31-mer double-stranded oligonucleotide containing the SimR O_X operator in a total reaction volume of 100 µl for 5 min at 4°C before addition of 1 pmol bovine trypsin. The 20 µl samples were then taken at 5-min time intervals. Reactions were stopped by adding SDS–PAGE loading buffer, boiled for 5 min, and analysed using SDS–PAGE. Proteins were transferred to PVDF membrane by electroblotting, stained with Coomasie blue and proteolytically resistant species were identified by N-terminal sequencing at the Protein & Nucleic Acid Chemistry Facility, University of Cambridge.

Global bioinformatic analysis of TFRs

We searched the PFAM database (http://pfam.sanger.ac.uk) for proteins that match the Hidden Markov Model profile PF00440, identifying 23 137 TFR candidates. Protein sequences longer than 300 amino acid residues were removed to eliminate false positives, and highly similar orthologous TFRs were removed using Jalview with a threshold of 99% identity (29), resulting in a non-redundant set of 12 715 TFRs.

The non-redundant set of TFRs was divided into clusters of 200 sequences using USEARCH and UCLUST (30). The amino acid sequences of the TFRs in each cluster were then aligned using MUSCLE (31) to identify their N-terminal extensions, which were defined as the amino-acid sequences preceding the conserved core DBDs (Figure 1). The globular body of the TFRs was defined by excluding the N-terminal extension from the whole protein sequence. In-house Perl scripts were used to quantify the length of the N-terminal extension and the fractions of R+K or D+E residues within these extensions. The sequences of the N-terminal extensions were concatenated together and submitted to the Regional Order Neural Network (RONN) programme (32) to predict the disorder probability for each residue. QtiPlot (http://soft.proindependent.com/qtiplot.html) was used to produce histograms.

RESULTS AND DISCUSSION

N-terminally truncated SimR derivatives bind DNA with reduced affinity

SimR possesses a 28-residue N-terminal extension that precedes the core DBD, herein termed the TFR arm (Figure 1), which carries four arginine residues at positions 18, 19, 22 and 25. This TFR arm is significantly longer than those in DesT, TetR, CgmR and QacR (Figure 1), the four TFRs for which DNA–protein crystal structures are available (14–17). To determine if the TFR arm of SimR is involved in DNA binding, we made C-terminally His-tagged SimR derivatives with progressively shorter N-terminal extensions and tested them for binding to the *simR–simX* intergenic region by EMSA. Wild-type SimR and SimR derivatives with 10, 15, 22 or 25 amino acid residues deleted from the N-terminus were overexpressed and purified (Supplementary Figure S1). Increasing concentrations of protein were incubated with a DNA probe spanning the *simR–simX* intergenic region and the complexes were resolved on native polyacrylamide gels (Figure 2). The *simR–simX* intergenic region contains two SimR operators: O_R closer to *simR*, and a higher affinity binding site, O_X, closer to *simX* (9). The lower and upper sets of shifted protein–DNA complexes seen in Figure 2 correspond, respectively, to single and double occupancy of these two SimR-binding sites (9). SimR DNA-binding affinity was reduced ~30-fold when 10 or 15 amino acid residues were deleted from the N-terminus, and was reduced by at least 120-fold when

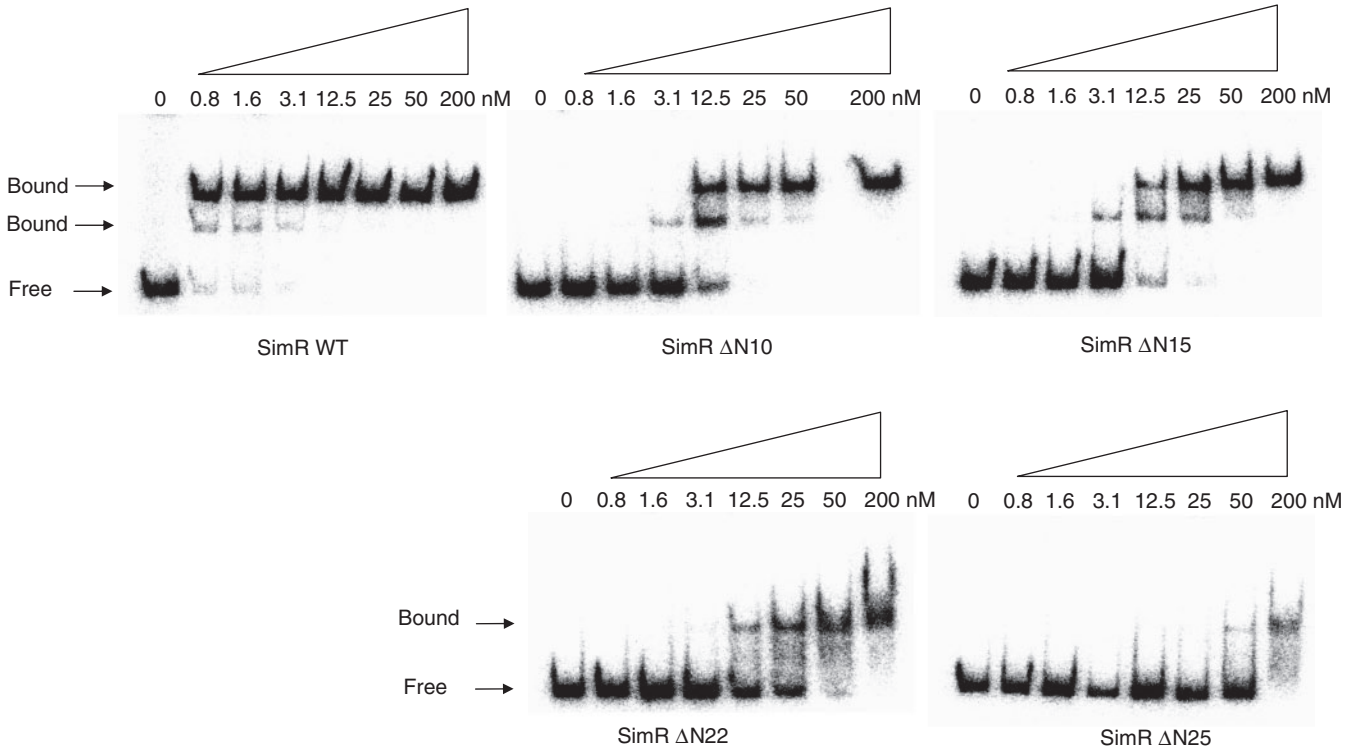


Figure 2. Electrophoretic mobility shift assay (EMSA) showing the binding of purified wild-type and N-terminally truncated derivatives of SimR to the *simR–simX* intergenic region. Bands correspond to SimR–DNA complexes (Bound) and free DNA (Free) are indicated. The final concentration of SimR is indicated above each lane.

22 or 25 amino acid residues were removed (Figure 2). These results suggested the TFR arm plays a role in DNA binding.

The TFR arm becomes protease resistant upon DNA binding

The 28-amino acid TFR arm of SimR has a high proportion of disorder-promoting amino acids and is predicted by the Proteins Disorder Prediction System (PrDOS; http://prdos.hgc.jp/cgi-bin/top.cgi) and by the Regional Order Neural Network (RONN; http://www.strubi.ox.ac.uk/RONN) servers to be disordered in solution (Supplementary Figure S2). Additionally, with the exception of three residues (residues 8–10, here termed the anchor string), this extension is disordered in both monomers in the SimR–SD8 structure, and it is only partially ordered in one monomer in the SimR–SC4 structure (18). The TFR arm is ordered in the SimR–apo structure, but its structure is the likely result of crystal packing (Supplementary Figure S3).

Because disordered regions are often hypersensitive to proteolysis (33), we examined the sensitivity of SimR to trypsin. The TFR arm was rapidly digested, leaving a much more stable product with a N-terminus at either residue Ser20 or Ser23 (Figure 3 and Supplementary Figure S4). Taken together, these observations suggest that the TFR arm is solvent exposed and displays conformational flexibility in solution in the absence of cognate DNA.

Since many unstructured regions exhibit increased resistance to proteolysis on binding of a partner (33,34), we determined the effect of DNA binding on the sensitivity of the TFR arm to trypsin. Addition of 25- or 31-bp DNA duplexes spanning the O_X operator substantially decreased the rate of SimR proteolysis, suggesting that DNA binding renders the TFR arm more resistant to trypsin (Figure 3). Consistent with this interpretation, proteolysis was not inhibited when a 15-bp O_X DNA duplex that is unable to bind to SimR was incubated with SimR (Figure 3 and Supplementary Figure S5A). In total, these experiments suggest that the TFR arm transitions from a disordered or conformationally flexible state to a more ordered, rigid state upon DNA binding.

The structure of SimR bound to its DNA operator

To understand how SimR binds to its operator sequence and to shed light on the role of the TFR arm in DNA binding, we crystallized SimR in complex with DNA. We tested DNA duplexes from 17 to 21 bp in length and found that only the minimal, blunt-ended 17-bp duplex crystallized in complex with SimR. The 17-bp DNA duplex used was the O_X operator (5′-TTCGTACGGTGTATGAA-3′), but carrying 2 bp changes to generate a near perfect inverted repeat (5′-TTCGTACGGCGTACGAA-3′), which bound SimR at least as tightly as the wild-type 17-bp DNA duplex (Supplementary Figure S5B). We solved the structure of full-length SimR (residues 1–259) in complex with this 17-bp DNA duplex

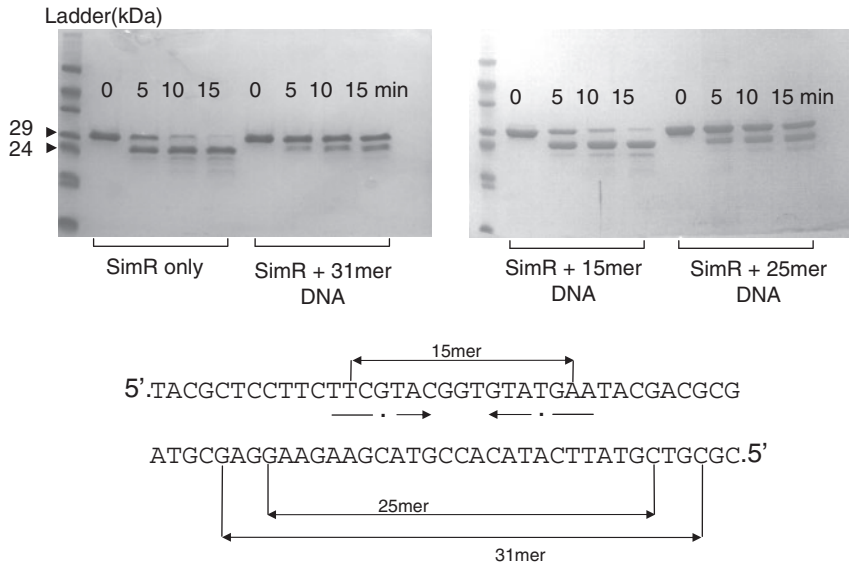


Figure 3. Limited tryptic proteolysis of SimR in the presence or absence of DNA. SimR was incubated either alone or with the O_X operator DNA duplexes indicated, before the addition of trypsin. Note that the 15-mer DNA duplex does not bind SimR (Supplementary Figure S5A). After SDS–PAGE, SimR species were visualized by Coomassie blue staining. The major product of tryptic digestion (arrowed) was shown by Edman sequencing to have an N-terminus corresponding to Ser20 or Ser23 of wild-type SimR.

to 2.99 Å resolution (Figure 4A). X-ray data collection and refinement statistics are summarized in Table 1.

The asymmetric unit contained two SimR dimers, each bound to a 17-bp DNA duplex. The two SimR dimer–DNA complexes are essentially identical [root mean square deviation (RMSD) between complexes for the Cα backbone = 0.15 Å], and thus only one complex is discussed throughout (Figure 4A). The conformation of bound DNA is mostly regular B-form but is bent away from the SimR dimer by ~15° (see below and Supplementary Figure S10A). The bases at the end of adjacent DNA duplexes stack and interact to form a pseudo-continuous double-helical DNA filament running through the crystal (Figure 4B and Supplementary Figure S6).

Interactions between the HTH motif and the major groove

The core DBD is composed of helices α1–α3 (residues 29–67). Helix α2 (residues 49–58) and the recognition helix α3 (residues 61–67) form the HTH motif which packs against α1 for stabilization (Figure 4A). Surprisingly, the recognition helix makes no canonical hydrogen bonds with the bases. However, the side chain of Met62 makes a series of contacts to three different bases including van der Waals to C3 (Cβ to C^5), and an uncommon electrostatic interaction between the S atom and the face of the base of T12, which is analogous to S stacking over the aromatic side chains of tryptophan, histidine and phenylalanine (35) (Figure 5). This interaction is buttressed by van der Waals contacts to the C^7 methyl group of T12. The S atom of Met62 also accepts a hydrogen bond from the N^6 hydrogen bond donor of A13. Another key interaction involved in the DNA sequence recognition mechanism of SimR is the stacking of the side chain of residue Tyr66 with the C^7 exocyclic methyl groups of T1 and T2. This interaction explains in great part why SimR has a higher affinity for the O_X operator, which has this pair of thymines, than for O_R, which has a pair of guanines at these positions (9). The dominant recognition helix interactions are with the phosphate backbone. For each operator half-site, there are hydrogen bonds between the hydroxyl group of Ser63 and the phosphate group of C3, between the hydroxyl group of Tyr65 and the phosphate group of T12 and between Tyr67 and the phosphate group of T2 (Figure 5). Just outside helix α3, the backbone NH group of Gly60 hydrogen bonds with the phosphate group of C3. On binding DNA, the recognition helix adopts a 3_{10} helical conformation, in contrast to the canonical α-helical conformation seen in the structures of SimR–apo and SimR–simocyclinone complexes (9). This conformational alteration in the recognition helix on DNA binding is also observed in TetR, and is believed to facilitate intimate interaction with the DNA (16).

Three residues in helix α2 contribute to DNA binding, with the side chain hydroxyl group of Ser49 forming a hydrogen bond with the phosphate backbone of C10 and the backbone NH group of Met50 forming a hydrogen bond with the phosphate backbone of G11 (Figure 5). The guanidinium group of Arg51 is involved in direct base recognition by bifurcated hydrogen bonds from the $N\eta^2$ atom to the O^6 and N^7 acceptors of G11. Other interactions between SimR and the major groove are hydrogen bonds between the amino group of Lys71 and the phosphate group of G11, and between the backbone NH group of Lys71 and the phosphate group of T12. Lys71 lies at the N-terminus of helix α4 at the very beginning of the LBD, just outside the core HTH motif of the DBD. This residue is highly conserved among TFRs and the equivalent lysine in TetR also forms a hydrogen bond with the phosphate backbone (16).

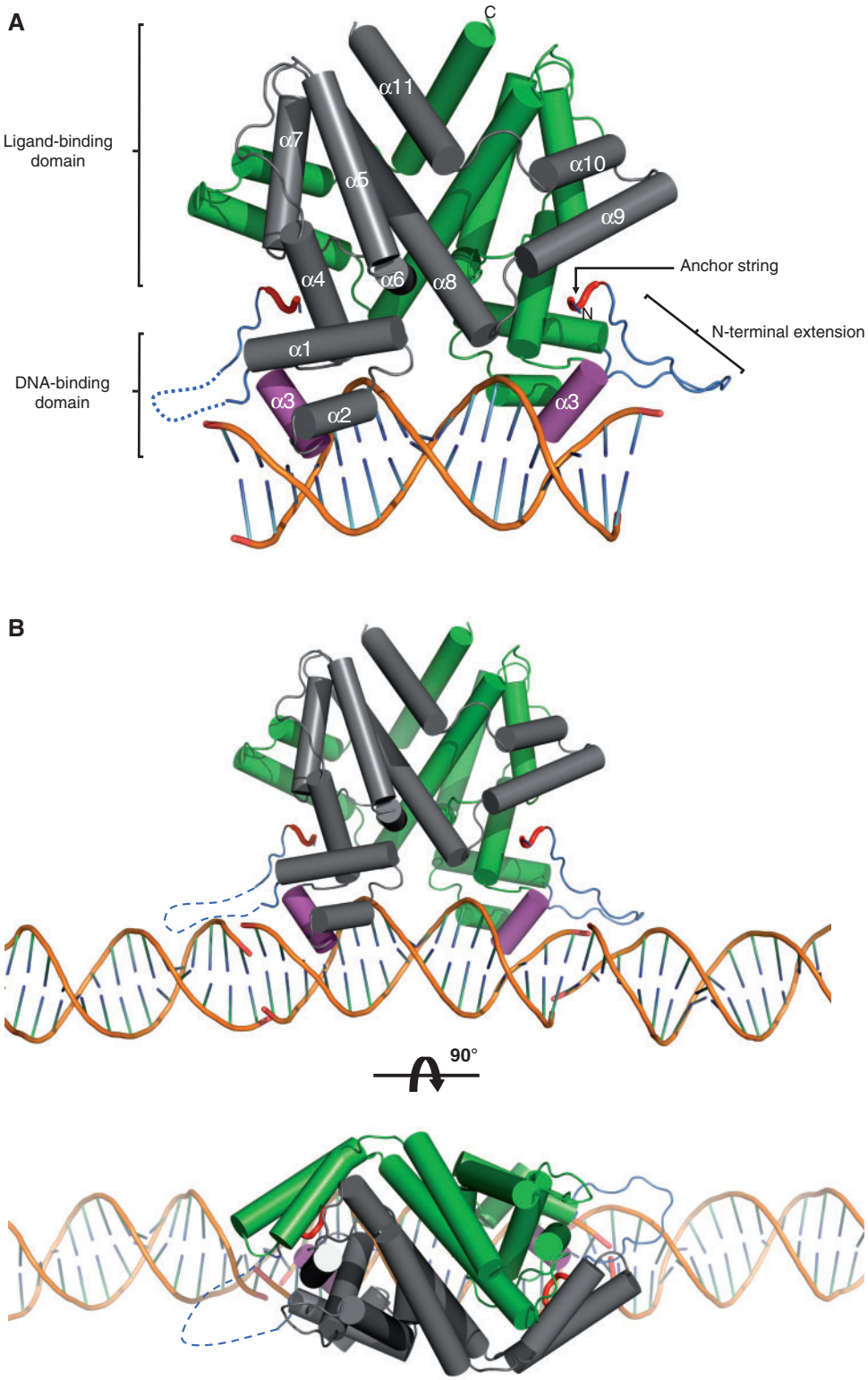


Figure 4. Structure of the SimR–17-mer complex (**A**) in isolation or (**B**) showing the adjacent DNA duplexes in the crystal. A cylindrical helix representation is used to highlight the secondary structure of SimR with key features labelled in (A). One subunit of the biological-relevant dimer is shown in grey and one in green. The recognition helix α3 is shown in magenta, the TFR arm is shown in blue and the N- and C-termini are labelled. The anchor string of the TFR arm (residues 8–11) is shown as a red tube cartoon. The dotted blue line represents the disordered TFR arm in the left-hand SimR subunit. In (**B**) only the DNA components of the adjacent symmetry complexes are shown in order to highlight the pseudo-continuous DNA filament running through the crystal (See also Supplementary Figure S6 and Figure 7).

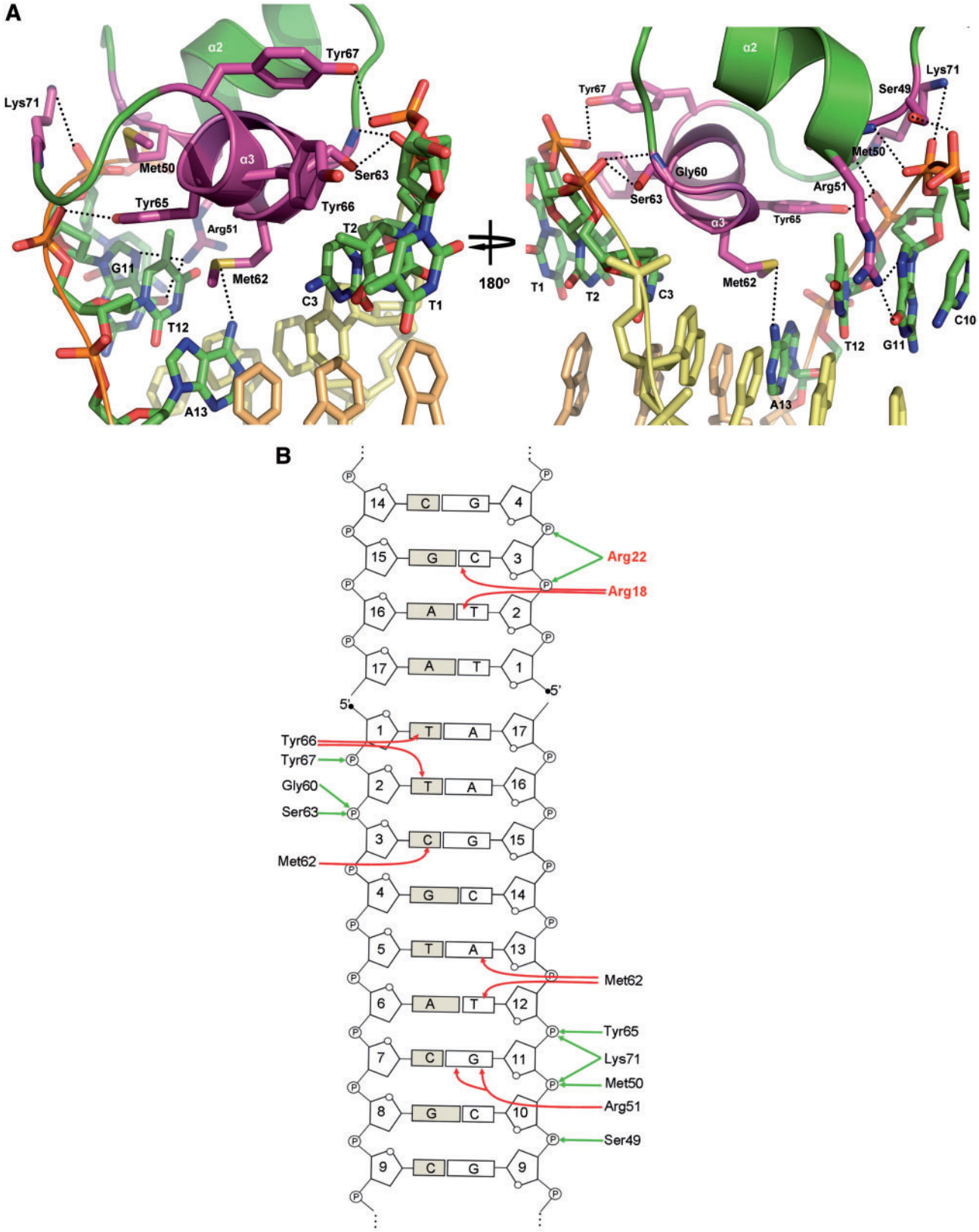


Figure 5. (**A**) Interactions between the HTH motif and the major groove. Stick representations of the interacting residues are shown in magenta. The Cα backbone of recognition helix α3 is shown in magenta and that of helix α2 is shown in green. Hydrogen bonds are represented by dotted black lines. The interacting bases are labelled and only the ring frames are shown for non-interacting bases. (**B**) Schematic representation of SimR–DNA contacts. For simplicity, only a recognition half-site and the first 4 bp of an adjacent duplex are shown. Interactions between amino acid residues and the bases of the cognate DNA operator are indicated by red arrows, and those between amino acid residues and the phosphate backbone are represented by green arrows. Amino acid residues belonging to the TFR arm are shown in red.

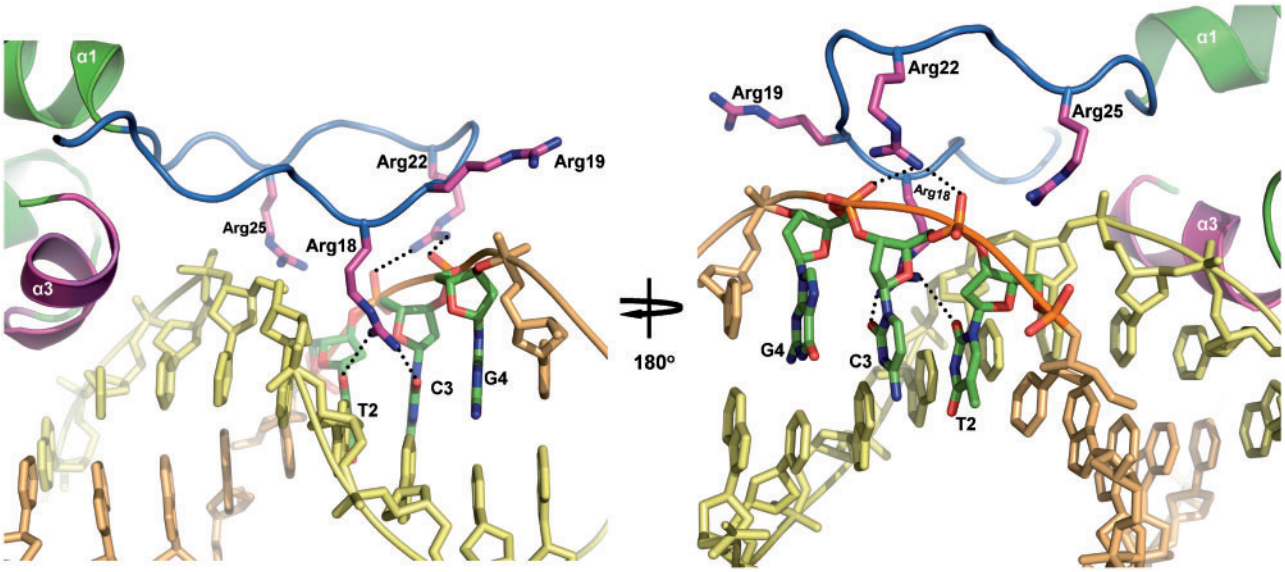


Figure 6. Interactions between the TFR arm and the minor groove. The Cα backbone of the TFR arm is shown in blue and stick representations of arginine residues Arg18, Arg19, Arg22 and Arg25 are shown in magenta. Hydrogen bonds are represented by dotted black lines. The interacting bases are labelled and only the ring frames are shown for non-interacting bases.

TFRs frequently rely on phosphate backbone contacts to mediate interaction with the DNA. In an extreme case, the DesT–DNA interface involves 11 phosphate backbone contacts but only two specific interactions with a pair of guanine bases within each half site (15). In contrast, TetR and QacR make extensive direct hydrogen bond contacts with the bases (16,17). In this sense, SimR is perhaps more similar to DesT than to TetR or QacR in its DNA sequence recognition mechanism. Thus, although the overall structure of the DBD in TFRs is conserved, it is clear that the mode of operator recognition differs from one member of the TFR family to another (14–17). TFRs recycle conserved residues and inventively employ non-conserved ones within the DBD for either base-specific hydrogen bond formation or for phosphate backbone contacts (Figure 1). It seems that there is no deterministic set of rules for TFR–DNA recognition.

Interactions between the TFR arm and the minor groove

If the structure of a single SimR–DNA complex is viewed in isolation, it can be seen that the TFR arm does not make contact with the cognate DNA duplex (Figure 4A). Instead, the TFR arm binds the minor groove of the adjacent DNA duplex in the pseudo-filament (Figure 4B). This binding to the minor groove is mediated through arginine residues that sit at the tip of the TFR arm (Figures 5B and 6). Specifically, the $N\eta^2$ atom of the guanidinium group of Arg18 forms a hydrogen bond with the O^2 of C3, while the $N\eta^1$ atom interacts with the O^2 of T2. In addition, the guanidinium group of Arg22 forms two salt bridges to the phosphate backbone of C3 and G4 (Figures 5B and 6). The electro-positive side chain of Arg18 is deeply buried in this minor groove (Figure 6), where the electronegative potential of the phosphate backbone is focused (36,37). This helps anchor the tip of the TFR arm in the minor groove. A third arginine in the flexible TFR arm, Arg19, does not contact DNA in the structure reported here (Figure 6). However, given the non-covalent nature of the DNA pseudo-filament, we considered the possibility that Arg19 might be involved in DNA binding in truly continuous double-stranded DNA. To examine this possibility, we mutagenized Arg19 to alanine and assayed the resulting protein for its ability to bind to the *simR–simX* intergenic region by EMSA. SimR R19A-bound DNA with an affinity equal to that of wild-type SimR (Supplementary Figure S7), suggesting Arg19 does not contribute to DNA binding. In contrast, when we constructed SimR R18A and SimR R22A variants, we found that each exhibited an approximate 15-fold reduction in binding affinity (Supplementary Figure S7), consistent with roles for R18 and R22 in DNA binding, as suggested by the structure of the SimR–DNA complex.

Initially, it was difficult to understand why SimR variants lacking just 10 or 15 N-terminal amino acid residues should have reduced DNA-binding affinity, given that they retain the interacting arginine residues. In the previously solved structures of apo–SimR and SimR–ligand complexes, although the TFR arm is mostly disordered, residues 8–10, herein termed the anchor string, are always visible in electron density maps (18), probably because this string of amino acid residues is stabilized by van der Waals interactions with the cleft between the LBD and the DBD. It therefore seems likely that this short segment, highlighted in red in Figure 4, serves as an anchoring point for the TFR arm to loop back onto the body of SimR. This arrangement may be important for restricting the flexibility of the TFR arm, so that it is poised appropriately to interact with the minor groove. Deleting 10 or 15 amino acids from the N-terminus would remove this anchor point, destabilizing loop formation and reducing DNA-binding affinity.

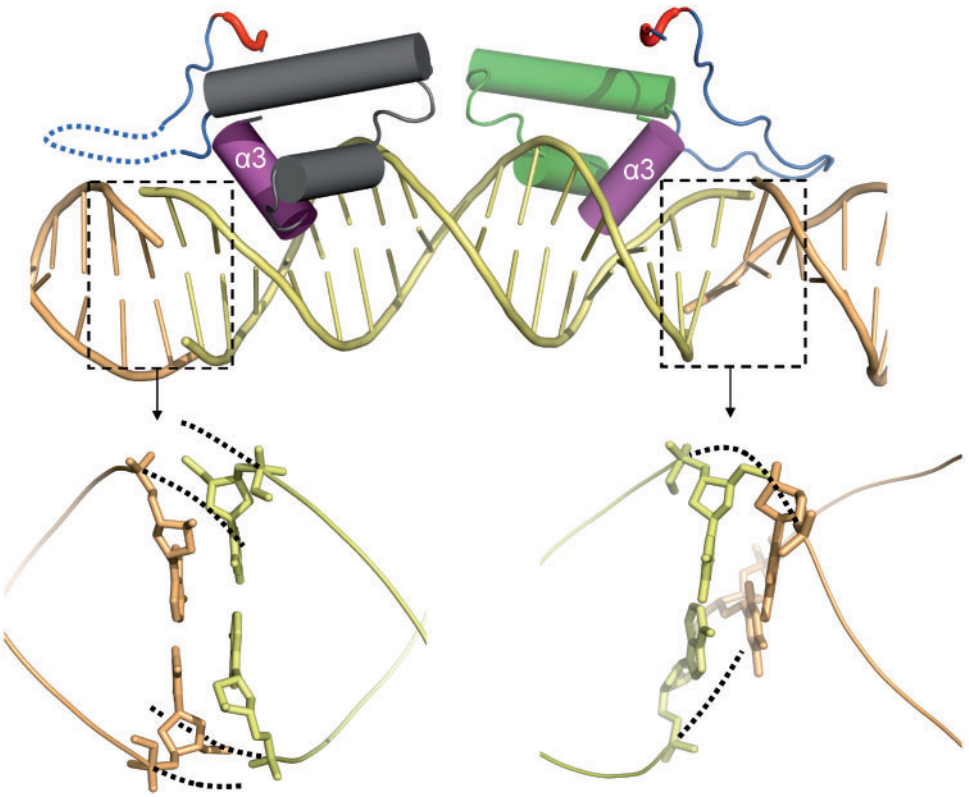


Figure 7. Non-equivalent stacking between adjacent DNA duplexes in the crystal pseudo-filament creates two different minor grooves. Only the DBD of SimR is shown. At the right-hand end of the central DNA duplex the base stacking allows the DNA phosphate backbone to transit smoothly (dotted lines) between adjacent duplexes, creating a relatively normal minor groove. At the left-hand end of the central DNA duplex the base stacking causes the phosphate backbone to veer away to avoid a steric clash (dotted lines), producing an abnormal minor groove. Adjacent DNA duplexes are shown in contrasting colours.

The more severe deletions, removing 22 or 25 amino acids, further reduce binding affinity because they remove the interacting arginine residues themselves.

In the crystal structure of the SimR–DNA complex, the TFR arm is seen in one SimR subunit but is disordered in the other subunit (Figure 4). From an inspection of the end-to-end base stacking between adjacent DNA duplexes within the crystal, it is clear that the two ends are not equivalent. The stacking at the right-hand end (as viewed in Figure 7) allows the neighbouring DNA strands to transit smoothly across the gap, producing a relatively normal minor groove. However, on the left-hand end the strands veer away to avoid a steric clash while maintaining base pair stacking, producing a much wider minor groove (Figure 7). It seems likely that the TFR arm is unable to interact with this 'abnormal' minor groove and is therefore disordered in the crystal. In the structure of the SimR–17-mer duplex, apart from the interaction of the anchor string with the body of SimR, the only contacts made by the TFR arm are with the minor groove of DNA (Figure 4 and Supplementary Figure S6). Based on the crystal structure of the SimR–DNA complex and the results of the proteolysis protection assays, we propose that the TFR arm transitions from a disordered or conformationally flexible state to a more ordered state upon binding to its cognate DNA.

N-terminally truncated SimR derivatives have a smaller footprint on DNA than wild-type SimR

We used DNase I protection to compare the footprints of wild-type SimR and the N-terminally truncated SimRs on the O_X and O_R operators in the *simR–simX* intergenic region (Figure 8A). In each case, saturating amounts of SimR protein were used to ensure complete protection of the binding sites. The footprint for wild-type SimR was comparable with that reported previously (9). In contrast, in the footprints generated using the N-terminally truncated SimR proteins, the edge of the protected region retracted at both ends of the footprint when compared to the footprint of full-length SimR (Figure 8). Specifically, when N-terminally truncated proteins were used, on the upper DNA strand the O_R footprint retracted by two base pairs at the left edge and by one base pair at the right edge (Figure 8). No retraction of the O_R footprint was apparent on the lower DNA strand. When N-terminally truncated proteins were used, on the upper DNA strand the left edge of the O_X footprint retracted by 1 bp, while no retraction was apparent at the right edge (Figure 8). On the lower DNA strand, the O_X footprint receded by 1 bp at both ends. These observations indicate that the TFR arm sterically hinders DNase I, protecting additional phosphodiester bonds from cleavage by the nuclease. Each SimR mutant protein

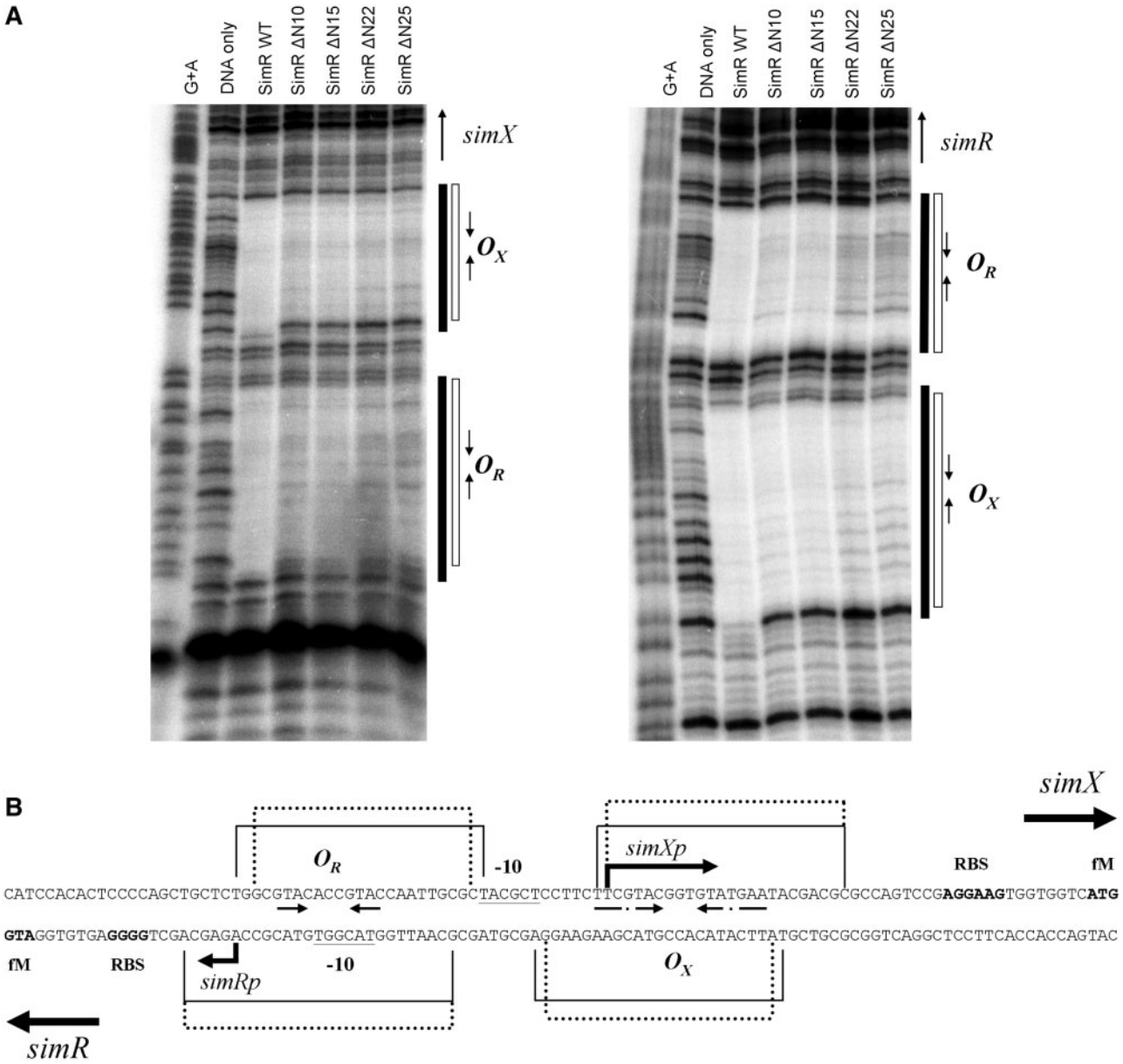


Figure 8. (**A**) DNase I footprinting analysis of the binding of wild-type and N-terminally truncated derivatives of SimR to the *simR–simX* intergenic region. A DNA fragment containing the *simR–simX* intergenic region, 5′-end labelled on either the upper strand (left panel) or the lower strand (right panel), was exposed to DNase I in the presence of saturating concentrations of SimR protein (200 nM for wild-type SimR, SimRΔN10 and SimRΔN15; 400 nM for SimRΔN22 and SimRΔN25). The sequencing ladders were generated by subjecting the probes to Maxam-Gilbert G+A chemical sequencing. Regions protected from DNase I cleavage (operators O_X and O_R) by wild-type SimR are indicated by solid vertical bars, and those protected by the N-terminally truncated SimR derivatives are indicated by open bars. Inverted repeats within the DNase I protected regions are indicated by convergent arrows. (**B**) Sequence of the *simR–simX* intergenic region summarizing the DNase I footprinting data. Regions protected by wild-type SimR are indicated by solid lines, and those protected by the N-terminally truncated SimR derivatives are indicated by dotted lines. Also indicated are the *simRp* and *simXp* transcription start points and putative −10 sequences, the *simR* and *simX* ribosome-binding sites (RBS), and the imperfect inverted repeats within the footprints.

produced the same footprint, regardless of whether 10, 15, 22 or 25 amino acids had been deleted from the N-terminus, consistent with the idea that residues 8–10, (i.e. the anchor string), are needed for the TFR arm to be fully functional, as discussed above. Note that the retraction of the footprint occurs at both ends of the operator, suggesting that the TFR arms of both monomers in the SimR dimer function in solution.

We also performed a complementary experiment to determine the binding affinity of wild-type SimR to three DNA duplexes of different lengths (15, 17 and 23 bp) spanning the O_X inverted repeat sequence. The 23-bp duplex bound SimR more strongly than the minimal 17-bp duplex, showing that DNA flanking the core 17-bp inverted repeat contributes to SimR binding (Supplementary Figure S5A). The 15-bp duplex failed to

bind SimR (Supplementary Figure S5A). In addition, although the minimal 17-bp duplex binds to SimR relatively well (Supplementary Figure S5A), it is unable to protect the TFR arm of SimR from tryptic digestion, while a 23-mer reduced the rate of proteolysis considerably (Supplementary Figure S8). Taken together, these observations suggest that, in solution, the TFR arm interacts with DNA outside the core 17-bp O_X operator, consistent with the SimR–DNA structure, which shows dimer–DNA interactions spanning 21 bp.

Among the five TFRs for which protein–DNA crystal structures are available (DesT, TetR, CgmR, QacR and SimR; Figure 1), only SimR possesses a flexible TFR arm that undergoes a transition to an ordered state upon DNA binding. DesT has a 12 amino acid residue N-terminal extension (Figure 1) but it is not disordered, instead forming part of an extended helix α1. Residues Arg5 and Lys9 of this short N-terminal extension in DesT nevertheless contribute to DNA binding (15), which is unusual because the main role of helix α1 is in stabilizing the HTH motif (α2– α3). Residues N-terminal to the core DBD in two other TFRs, *Neisseria gonorrhoeae* MtrR (11 amino acids) and *Streptomyces coelicolor* ActR (32 amino acids) have also been suggested to be involved in DNA binding (38,39), implying a possible common role for TFR N-terminal extensions when present (see also the global TFR bioinformatic analysis presented below). Similar kinds of extensions have been identified in at least two other families of DNA-binding proteins. For example, members of the eukaryotic Hox family recognize nearly identical major groove sequences through the recognition helix of their homeodomain but use an extended arm to insert into the minor groove to enhance binding specificity (40). A related example is phage lambda repressor, which has a conventional HTH motif and an additional N-terminal extension that promotes DNA binding, in this case by interacting with the major groove (41). A comprehensive analysis of all available protein–DNA structures has shown that the binding of arginine residues to narrow minor grooves is a widely used mechanism in protein–DNA recognition. This readout mechanism exploits the fact that narrow minor grooves, often associated with A-tracts, strongly enhance the negative electrostatic potential of the DNA (36,37). However, it should be noted that the minor groove bound by the TFR arm of SimR is not associated with an A-tract, and has a slightly enlarged width with respect to canonical B DNA (Supplementary Figure S10C).

The arginine- and lysine-rich TFR arm is likely to be a common feature of TetR family members

We searched the PFAM database (http://pfam.sanger.ac.uk/) for proteins that match the Hidden Markov Model profile PF00440, identifying 12 715 non-redundant TFRs (see Materials and methods for further details). The amino acid sequences of these TFRs were then aligned using MUSCLE (31) to identify the core DBD and any N-terminal extension. Twenty-eight per cent had N-terminal extensions of less than 10 amino acids, 44% had N-terminal extensions of 11–20 amino acids, 17% had N-terminal extensions of 21–30 amino acids and 11% had N-terminal extensions >31 amino acids. Further, the fraction of Arg and Lys residues in these N-terminal extensions (mean value = 20.5%) was almost double the frequency found in the globular body of the TFRs (mean value = 11.4%) (Supplementary Figure S9A). Finally, the RONN server predicts that the majority of these N-terminal extensions are likely to be disordered in solution (Supplementary Figure S9B). It therefore seems likely that a conformationally malleable, DNA-binding N-terminal extension is a common feature of TFRs.

DNA bending induced by SimR binding

DNA helical parameters were analysed using the Curves+ programme (27). The overall conformation of the 17-bp duplex is B-DNA, with an average helical twist of 33.7° (compared to a helical twist value of 36.0° for an idealized B-form DNA). It should be noted that individual steps might show significant deviations from the average value. The global bending of DNA is ∼15° (Supplementary Figure S10A). Since bending is most affected by the base step roll and twist angles (42), we plotted the roll and twist angles against the base steps to pinpoint the source of bending (Supplementary Figure S10B). There are two significant positive rolls (10–10.7°) centred around base steps 6–7–8 in the operator half-site and symmetrically around steps 9–10–11 of the opposite half-site (Supplementary Figure S10B). The increase in roll angle coincides with the decrease in twist angle (26.7– 26.9°) (Supplementary Figure S10B). The average global roll and twist angles are 2.9° and 33.4°, respectively. Thus local kinks around those base steps produce a global bend in the DNA, rather than a smooth bending. Moreover, there is a significant increase in the width of the minor groove from base step 6 through to base step 12, while the major groove width is just below the value for an idealized B-form DNA (Supplementary Figure S10C). Since the average distance between the two recognition helices in the SimR–DNA complex is 36.8 Å [assessed as the distance between the Cα atom of Tyr65 in each subunit (13)], greater than the distance between two consecutive major grooves in idealized B-DNA (34 Å), it is likely that the bending and the unwinding of the central DNA steps might be necessary for optimal positioning of the HTH motifs in adjacent major grooves. Lastly, although the sequence of the 17-bp duplex used in this study is a perfect inverted repeat with the exception of the central GC base pair, the groove width and roll parameters are not symmetrical across this central base pair. This reflects the non-equivalent end-to-end interactions between neighbouring DNA duplexes described above (Figure 7).

Comparison of the SimR–DNA and SimR–simocyclinone complexes suggests the mechanism of derepression

In a previous report, we speculated about the mechanism of simocyclinone-mediated derepression, based on a comparison of the structures of SimR–apo and the SimR–SD8 complex (18). However, it was apparent that SimR–apo had not crystallized in its DNA-binding form, since the distance between its recognition helices was 42.3 Å,

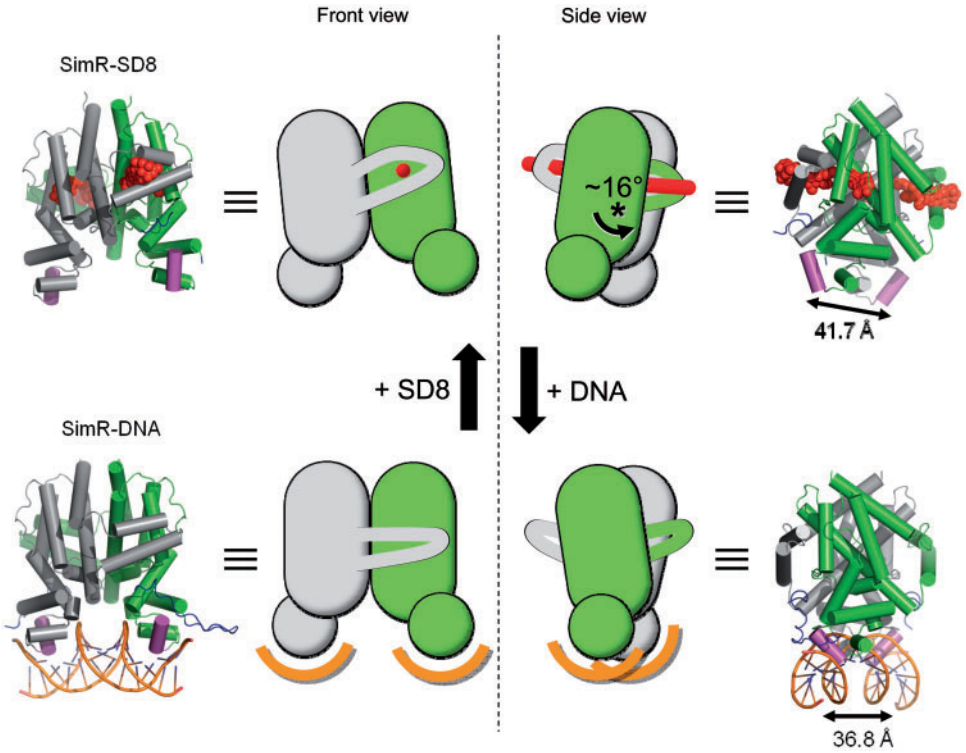


Figure 9. Structures of SimR–simocyclinone and SimR–DNA together with schematic representations illustrating the rigid-body rotation of the subunits relative to one another. In order to emphasize the subunit rotation, the grey coloured subunits are shown fixed in the same relative orientations. This can be clearly seen in the side view where the green subunit rotates by ~16° relative to the grey subunit; the approximate pivot point is indicated by the asterisk (see also Supplementary Figures S11 and S12). The distances separating the recognition helices α3 and α3′ in the two structures are indicated.

a spacing incompatible with binding to two consecutive major grooves (18). Moreover, this spacing was comparable to the corresponding value of 41.7 Å obtained for the SimR–SD8 complex. Indeed, TFR apo-proteins in general do not crystallize in their DNA-binding form (13). The helix separation obtained for SimR–DNA was significantly shorter at 36.8 Å (averaged over the two complexes in the asymmetric unit), this value lying within the range of 34.7–38.8 Å observed in other TFR–DNA complexes (13,15). The major structural differences between the repressed, DNA-bound conformation of SimR and the derepressed, SD8-bound conformation, result from a 16° rotation of the subunits relative to one another roughly about the centre of the dimer interface (Figure 9 and Supplementary Figure S11). This re-defines many of the inter-subunit contacts, although the interface areas remain similar at 2795 and 2640 $Å^2$ for SimR–SD8 and SimR–DNA, respectively [as calculated by the Protein Interactions, Surfaces and Assemblies server (PISA, http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html) (43)]. However, five reciprocated inter-subunit hydrogen bonds (i.e. 10 in total) are preserved between the two conformational states. These link the C-terminal end of α8 and the α9–α10 wrapping arm to the LBD of the adjacent subunit. As a consequence, when the subunits rotate, the α9–α10 wrapping arm moves with the adjacent subunit and the C-terminal end of α8 bends (Supplementary Figure S11). Pair-wise superpositions of individual subunits taken from the SimR–SD8 and SimR–DNA structures based on the subunit cores (i.e., inclusive of residues 29–168 plus 222–247 and exclusive of the TFR arm, the C-terminal end of α8 and the α9–α10 wrapping arm) gave RMSD values in the range 0.85–0.96 Å, indicating that the cores move essentially as rigid bodies at the protein backbone level and, importantly, there is no significant re-orientation of the DBD with respect to the LBD, in contrast to the 'pendulum-like' motion seen in TetR (Supplementary Figure S12) (12,16). Nevertheless, the crystal structures do not convey the dynamic behaviour of the system and, as has been illustrated for other TFRs (13,44), in the absence of ligands or DNA, the protein is generally highly flexible and capable of sampling a variety of conformations, presumably including those akin to both the ligand- and DNA-bound states. The binding of SD8, a relatively hydrophobic molecule, contributes to the hydrophobic core of the SimR dimer; this will have a stabilizing effect on the overall structure, locking it into a relatively rigid, low-energy state. Moreover, the combination of the threading of the ligand through both subunits and the projection of the side chain of Arg122 into the opposing subunit contribute to the rigidification of the system (18). The flexibility of the apo form is important to enable the TFR arms and the recognition helices to engage optimally with the DNA. The resulting favourable protein–DNA interactions will have a stabilizing effect on this conformation of SimR. Moreover, in the DNA binding

conformation, the repositioning of the C-terminal end of helix α8 appropriately places it to make salt bridges to the DBD of the same subunit and to that of the opposing subunit, specifically between Arg179 and Glu46, and between Arg180 and Glu72, respectively. These interactions, not present in the SD8-bound form, will further stabilize the DNA-bound conformation of SimR.

ACCESSION NUMBER

Coordinates and structure factors for the SimR–DNA structure described herein have been deposited in the Protein Data Bank with accession number 3ZQL.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors are grateful to Clare Stevenson for assistance with X-ray crystallography, to Richard Little and Mahmoud Al-Bassam for advice on limited proteolysis experiments and to beamline scientists at the Advanced Light Source (Berkeley, CA, USA) and the Diamond Light Source (Harwell, UK) for assistance with X-ray data collection.

FUNDING

John Innes Centre Rotation Studentship (to T.B.K.L.); Short Term EMBO Fellowship ASTF: 290-2010 (to T.B.K.L.); Korner Travelling Fellowship (to T.B.K.L.); BBSRC grant BB/I002197/1 (to M.J.B. and D.M.L.); BBSRC Core Strategic Grant to the John Innes Centre; R.A. Welch Foundation grant G-0040 (to R.G.B.); M.D. Anderson Trust Fellowship (to M.A.S.). Funding for open access charge: BBSRC grant BB/I002197/1 (to M.J.B. and D.M.L.).

Conflict of interest statement. None declared.

REFERENCES

1. Berdy,J. (2005) Bioactive microbial metabolites. *J. Antibiot.*, **58**, 1–26.
2. Walsh,C. (2003) *Antibiotics: Actions, Origins, Resistance*. ASM Press, Washington, DC.
3. Edwards,M.J., Flatman,R.H., Mitchenall,L.A., Stevenson,C.E., Le,T.B., Clarke,T.A., McKay,A.R., Fiedler,H.P., Buttner,M.J., Lawson,D.M. *et al.* (2009) A crystal structure of the bifunctional antibiotic simocyclinone D8, bound to DNA gyrase. *Science*, **326**, 1415–1418.
4. Flatman,R.H., Howells,A.J., Heide,L., Fiedler,H.P. and Maxwell,A. (2005) Simocyclinone D8, an inhibitor of DNA gyrase with a novel mode of action. *Antimicrob. Agents Chemother.*, **49**, 1093–1100.
5. Oppegard,L.M., Hamann,B.L., Streck,K.R., Ellis,K.C., Fiedler,H.P., Khodursky,A.B. and Hiasa,H. (2009) In vivo and in vitro patterns of the activity of simocyclinone D8, an angucyclinone antibiotic from Streptomyces antibioticus. *Antimicrob. Agents Chemother.*, **53**, 2110–2119.
6. Holzenkämpfer,M., Walker,M., Zeeck,A., Schimana,J. and Fiedler,H.P. (2002) Simocyclinones, novel cytostatic angucyclinone antibiotics produced by Streptomyces antibioticus Tu 6040 II. Structure elucidation and biosynthesis. *J. Antibiot.*, **55**, 301–307.
7. Schimana,J., Fiedler,H.P., Groth,I., Süssmuth,R., Beil,W., Walker,M. and Zeeck,A. (2000) Simocyclinones, novel cytostatic angucyclinone antibiotics produced by Streptomyces antibioticus Tu 6040. I. Taxonomy, fermentation, isolation and biological activities. *J. Antibiot.*, **53**, 779–787.
8. Galm,U., Schimana,J., Fiedler,H.P., Schmidt,J., Li,S.M. and Heide,L. (2002) Cloning and analysis of the simocyclinone biosynthetic gene cluster of Streptomyces antibioticus Tu 6040. *Arch. Microbiol.*, **178**, 102–114.
9. Le,T.B., Fiedler,H.P., den Hengst,C.D., Ahn,S.K., Maxwell,A. and Buttner,M.J. (2009) Coupling of the biosynthesis and export of the DNA gyrase inhibitor simocyclinone in Streptomyces antibioticus. *Mol. Microbiol.*, **72**, 1462–1474.
10. Trefzer,A., Pelzer,S., Schimana,J., Stockert,S., Bihlmaier,C., Fiedler,H.P., Welzel,K., Vente,A. and Bechthold,A. (2002) Biosynthetic gene cluster of simocyclinone, a natural multihybrid antibiotic. *Antimicrob. Agents Chemother.*, **46**, 1174–1182.
11. Chopra,I. and Roberts,M. (2001) Tetracycline antibiotics: mode of action, applications, molecular biology, and epidemiology of bacterial resistance. *Microbiol. Mol. Biol. Rev.*, **65**, 232–260.
12. Ramos,J.L., Martinez-Bueno,M., Molina-Henares,A.J., Teran,W., Watanabe,K., Zhang,X., Gallegos,M.T., Brennan,R. and Tobes,R. (2005) The TetR family of transcriptional repressors. *Microbiol. Mol. Biol. Rev.*, **69**, 326–356.
13. Yu,Z., Reichheld,S.E., Savchenko,A., Parkinson,J. and Davidson,A.R. (2010) A comprehensive analysis of structural and sequence conservation in the TetR family transcriptional regulators. *J. Mol. Biol.*, **400**, 847–864.
14. Itou,H., Watanabe,N., Yao,M., Shirakihara,Y. and Tanaka,I. (2010) Crystal structures of the multidrug binding repressor Corynebacterium glutamicum CgmR in complex with inducers and with an operator. *J. Mol. Biol.*, **403**, 174–184.
15. Miller,D.J., Zhang,Y.M., Subramanian,C., Rock,C.O. and White,S.W. (2010) Structural basis for the transcriptional regulation of membrane lipid homeostasis. *Nat. Struct. Mol. Biol.*, **17**, 971–975.
16. Orth,P., Schnappinger,D., Hillen,W., Saenger,W. and Hinrichs,W. (2000) Structural basis of gene regulation by the tetracycline inducible Tet repressor-operator system. *Nat. Struct. Biol.*, **7**, 215–219.
17. Schumacher,M.A., Miller,M.C., Grkovic,S., Brown,M.H., Skurray,R.A. and Brennan,R.G. (2002) Structural basis for cooperative DNA binding by two dimers of the multidrug-binding protein QacR. *EMBO J.*, **21**, 1210–1218.
18. Le,T.B., Stevenson,C.E., Fiedler,H.P., Maxwell,A., Lawson,D.M. and Buttner,M.J. (2011) Structures of the TetR-like simocyclinone efflux pump repressor, SimR, and the mechanism of ligand-mediated derepression. *J. Mol. Biol.*, **408**, 40–56.
19. Leslie,A.G. (2006) The integration of macromolecular diffraction data. *Acta Crystallogr. D Biol. Crystallogr.*, **62**, 48–57.
20. Evans,P. (2006) Scaling and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr.*, **62**, 72–82.
21. McCoy,A.J., Grosse-Kunstleve,R.W., Adams,P.D., Winn,M.D., Storoni,L.C. and Read,R.J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.*, **40**, 658–674.
22. Emsley,P. and Cowtan,K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 2126–2132.
23. Murshudov,G.N., Vagin,A.A. and Dodson,E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.*, **53**, 240–255.
24. Adams,P.D., Grosse-Kunstleve,R.W., Hung,L.W., Ioerger,T.R., McCoy,A.J., Moriarty,N.W., Read,R.J., Sacchettini,J.C., Sauter,N.K. and Terwilliger,T.C. (2002) PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D Biol. Crystallogr.*, **58**, 1948–1954.
25. Painter,J. and Merritt,E.A. (2006) Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr. D Biol. Crystallogr.*, **62**, 439–450.
26. DeLano,W.L. (2002) *The PyMol User's Manual*. DeLano Scientific, San Carlos, CA, USA.

27. Lavery,R., Moakher,M., Maddocks,J.H., Petkeviciute,D. and Zakrzewska,K. (2009) Conformational analysis of nucleic acids revisited: curves+. *Nucleic Acids Res.*, **37**, 5917–5929.
28. Maxam,A.M. and Gilbert,W. (1980) Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol.*, **65**, 499–560.
29. Waterhouse,A.M., Procter,J.B., Martin,D.M., Clamp,M. and Barton,G.J. (2009) Jalview version 2–a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, **25**, 1189–1191.
30. Edgar,R.C. (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, **26**, 2460–2461.
31. Edgar,R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.*, **32**, 1792–1797.
32. Yang,Z.R., Thomson,R., McNeil,P. and Esnouf,R.M. (2005) RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics*, **21**, 3369–3376.
33. Receveur-Brechot,V., Bourhis,J.M., Uversky,V.N., Canard,B. and Longhi,S. (2006) Assessing protein disorder and induced folding. *Proteins*, **62**, 24–45.
34. Dyson,H.J. and Wright,P.E. (2005) Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.*, **6**, 197–208.
35. Pal,D. and Chakrabarti,P. (2001) Non-hydrogen bond interactions involving the methionine sulfur atom. *J. Biomol. Struct. Dyn.*, **19**, 115–128.
36. Rohs,R., Jin,X., West,S.M., Joshi,R., Honig,B. and Mann,R.S. (2010) Origins of specificity in protein-DNA recognition. *Annu. Rev. Biochem.*, **79**, 233–269.
37. Rohs,R., West,S.M., Sosinsky,A., Liu,P., Mann,R.S. and Honig,B. (2009) The role of DNA shape in protein-DNA recognition. *Nature*, **461**, 1248–1253.
38. Hoffmann,K.M., Williams,D., Shafer,W.M. and Brennan,R.G. (2005) Characterization of the multiple transferable resistance repressor, MtrR, from Neisseria gonorrhoeae. *J. Bacteriol.*, **187**, 5008–5012.
39. Willems,A.R., Tahlan,K., Taguchi,T., Zhang,K., Lee,Z.Z., Ichinose,K., Junop,M.S. and Nodwell,J.R. (2008) Crystal structures of the Streptomyces coelicolor TetR-like protein ActR alone and in complex with actinorhodin or the actinorhodin biosynthetic precursor (S)-DNPA. *J. Mol. Biol.*, **376**, 1377–1387.
40. Joshi,R., Passner,J.M., Rohs,R., Jain,R., Sosinsky,A., Crickmore,M.A., Jacob,V., Aggarwal,A.K., Honig,B. and Mann,R.S. (2007) Functional specificity of a Hox protein mediated by the recognition of minor groove structure. *Cell*, **131**, 530–543.
41. Beamer,L.J. and Pabo,C.O. (1992) Refined 1.8 A crystal structure of the lambda repressor-operator complex. *J. Mol. Biol.*, **227**, 177–196.
42. Dickerson,R.E. (1998) DNA bending: the prevalence of kinkiness and the virtues of normality. *Nucleic Acids Res.*, **26**, 1906–1926.
43. Krissinel,E. and Henrick,K. (2005) Detection of protein assemblies in crystals. In Berthold,M.R., Glen,R.C., Diederichs,K., Kohlbacher,O. and Fischer,I. (eds), *Computational Life Sciences*, Vol. 3695. Springer-Verlag, Berlin Heidelberg, pp. 163–174.
44. Reichheld,S.E., Yu,Z. and Davidson,A.R. (2009) The induction of folding cooperativity by ligand binding drives the allosteric response of tetracycline repressor. *Proc. Natl Acad. Sci. USA*, **106**, 22263–22268.
45. Chen,V.B., Arendall,W.B. 3rd, Headd,J.J., Keedy,D.A., Immormino,R.M., Kapral,G.J., Murray,L.W., Richardson,J.S. and Richardson,D.C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 12–21.