Downloaded from rstb.royalsocietypublishing.org

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Structure and Function of Escherichia coli met Repressor: Similarities and Contrasts with trp Repressor

Simon E. V. Phillips and Peter G. Stockley

Phil. Trans. R. Soc. Lond. B 1996 **351**, 527-535
doi: 10.1098/rstb.1996.0051

Email alerting service Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click **here**

To subscribe to *Phil. Trans. R. Soc. Lond. B* go to: **http://rstb.royalsocietypublishing.org/subscriptions**

This journal is © 1996 The Royal Society

Downloaded from rstb.royalsocietypublishing.org

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Structure and function of *Escherichia coli met* repressor: similarities and contrasts with *trp* repressor

SIMON E. V. PHILLIPS[1] AND PETER G. STOCKLEY[2]

[1] *Department of Biochemistry and Molecular Biology* and [2] *Department of Genetics, University of Leeds, Leeds LS2 9JT, U.K.*

SUMMARY

Transcription of genes encoding enzymes for the biosynthesis of methionine and trytophan in *Escherichia coli* is regulated by the ligand-activated *met* and *trp* repressors. X-ray crystallographic studies show how these two small proteins, although similar in size and function, have totally different three-dimensional structures and specifically recognize their respective DNA operator sequences in different ways. A common feature is that both repressors bind as cooperative arrays to tandem repeats of 8 base-pair 'Met' or 'Trp boxes' respectively, and the consensus sequences share the rare tetranucleotide CTAG. A series of structural and functional studies have shown how the two repressors discriminate between their operators, using a combination of direct contacts between side chains and bases, and indirect sensing of conformational properties of the DNA.

1. INTRODUCTION

Methionine is an important amino acid, in that it acts as the initiator of protein synthesis, as N-formyl methionine (Marcker & Sanger 1964; Adams & Capecci 1966), in protein elongation, and is the precursor of spermidine (Tabor *et al.* 1958). *S*-adenosylmethionine (SAM) is a universal methylating agent in the cell (Paik & Kim 1971; Soll 1971), and is formed by the addition of ATP to methionine (figure 1).

Studies of the regulation of methionine synthesis in bacteria, extensively reviewed by Saint-Girons *et al.* (1988) and Old *et al.* (1991), laid some of the early foundations for our understanding of regulation of gene expression. Cohn *et al.* (1953) showed that the presence of extracellular methionine in the medium represses the production of methionine synthase in *E. coli*, as well as that of other biosynthetic enzymes in the *met* pathway. Cohen & Jacob (1959) later showed a similar effect with tryptophan, whose presence repressed the synthesis of tryptophan synthetase. They went on to demonstrate that the effect was specific, so that tryptophan did not repress methionine synthase (or vice versa), and selected for mutant bacteria where this regulation was lost. It was deduced from the genetic data that some of these mutants contained lesions in genes governing production of specific repressors. These molecules are now well known as the *met* and *trp* repressor proteins (MetJ and TrpR). A series of parallels and contrasts between the structures and functions of these two repressors demonstrates differing evolutionary solutions to related problems of transcriptional regulation. In the description of *met* repressor below, comparisons are made to the equivalent features of *trp* repressor.

2. THE *MET* REPRESSOR PROTEIN

The *met* repressor is the product of the *metJ* gene, and has been cloned, sequenced and purified (Saint-Girons *et al.* 1984, 1986). It exists in dilute solution as a stable dimer (M_r 23 988) of identical 104 amino acid subunits. The free repressor (aporepressor) has a relatively low affinity for DNA, and for the amino acid methionine. It does, however, bind two molecules of SAM noncooperatively, with a K_d of 10^{-5} M, to form the active repressor (holorepressor) which has a high affinity for DNA. SAM is almost certainly the corepressor *in vivo*, and the evidence for this is reviewed in Old *et al.* (1991). An *E.coli* cell contains about 600 *met* repressor molecules which, in the presence of appropriate levels of SAM, are able to bind specifically to at least six independent DNA operators associated with genes controlling methionine and SAM biosynthesis that are widely scattered throughout the chromosome.

The *trp* repressor is a small, stable dimer of identical 107 amino acid subunits, which was cloned and sequenced by Gunsalus & Yanofsky (1980). The aporepressor binds two molecules of L-tryptophan

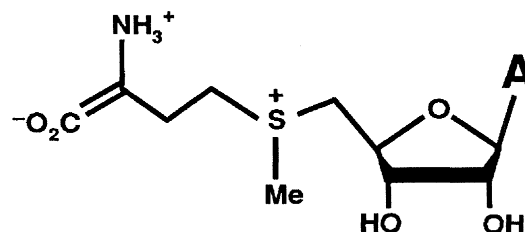

Figure 1. Structural formula for *S*-adenosylmethionine (SAM). (A = adenine).

© 1996 The Royal Society

The Royal Society is collaborating with JSTOR to digitize, preserve, and extend access to *Philosophical Transactions: Biological Sciences.* www.jstor.org

Downloaded from rstb.royalsocietypublishing.org

non-cooperatively, with a K_d of 10^{-5} M, to form the active holorepressor, which can bind specifically to three operators associated with genes for tryptophan biosynthesis (Joachimiak *et al.* 1983).

3. THE *MET* OPERATORS

The *met* operators in *E. coli* and *S. typhimurium* consist of tandem repeats of eight base pair (b.p.) sequences, homologous to a palindromic consensus AGACGTCT, known as 'met boxes' (Belfaiza *et al.* 1986). The known operator sequences are shown in figure 2.

They vary in length from 16–40 b.p. as defined by 50% sequence identity to consensus boxes, corresponding to two to five met boxes. The consensus sequence is highly symmetrical, with centres of inverted repeats (twofold axes of symmetry in three dimensions) at the centre of each met box, and at the junctions between them. The smallest unique sequence unit is therefore the first four bases (AGAC) of the box, the half-met box, from which the entire consensus can be generated by symmetry.

E. coli met operators do not contain a single example of a perfect consensus met box, and there are only two

Met operators

consensus:	AGACGTCT	AGACGTCT	AGACGTCT	AGACGTCT	AGACGTCT
metA Ec	AGctaTCT 63%	gGAtGTCT 75%	AaACGTaT 75%	AagCGTaT 63%	
metB Ec	AtACGcaa 50%	AGAaGTtT 75%	AGAtGTCc 75%	AGAtGTaT 75%	tGACGTCc 75%
metB St	AtACGcaa 50%	AGAaGTtT 75%	AGAtGTCc 75%	AGAtGTaT 75%	tGACGTCT 88%
metC Ec	AGACaTCc 75%	AGACGTaT 88%			
metC St	AGACaTCc 75%	AGACGgtT 75%			
metE Ec	gGAtGaaT 50%	AaACtTgc 50%	cGcCtTCc 50%		
metE St	gGAtGTgT 75%	AaACaTCc 62.5%	AGACGTCT 100%		
metF Ec	cttCaTCT 50%	ttACaTCT 63%	gGACGTCT 88%	AaACGgaT 63%	AGAtGTgc 63%
metF St	cGtCaTtT 50%	ttACaTCT 63%	gGACGTCT 88%	AaACGgaT 63%	AGAtGTtT 75%
metR Ec	AGgatTtT 50%	AGcCGTCc 75%	AGAtGTtT 75%	AcACaTCc 63%	
metR St	AGACGTCT 100%	gGAtGTtT 63%	AcACaTCc 63%	AtAaaTgT 50%	

Trp operators

consensus:	AGTTAACT	AGTTAACT −9	AGTTAACT	AGTTAACT +9	AGTTAACT
trpEDCBA	AATTAATC 63%	ATCGAACT 63%	AGTTAACT 100%	AGTACGCA 50%	AGTTCACG 75%
aroH	AGTCGCCG 50%	AATGTACT 63%	AGAGAACT 75%	AGTGCATT 63%	AGCTTATT 63%
trpR		ATCGTACT 50%	CTTTAGCG 50%	AGTACAAC 50%	

Figure 2. Alignment of known sequences for *met* operators in *E. coli* (*Ec*) and *S. typhimirium* (*St*), and *Ec trp* operators, compared to consensus met and trp box sequences. Upper case letters indicate bases identical to the consensus, and the figure below each individual box shows its identity to the consensus. The identity is generally higher in the shorter operators, and towards the centre of the longer ones. The minimum size of a viable *met* operator is two met boxes, as defined by sequence homology and *in vitro* binding assays, but lengths up to five boxes are observed. For the purposes of this discussion, bases are numbered from the first base of the box (figure 3). The *trpEDCBA* sequence additionally shows the standard numbering for *trp* operators relative to the centre. Bases −9 and +9 are shown, and will be referred to in the text as +9t and −9t. Arrows indicate the dyad axes coinciding with the central dyads of bound repressors. Sequences for *met* operators from Old *et al.* (1991) and references therein, and for *trp* from Bass *et al.* (1987).

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES
PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Downloaded from rstb.royalsocietypublishing.org

Met									
base position	1	2	3	4		5	6	7	8
consensus met-box	A	G	A	C	|	G	T	C	T
base conservation (%) (half site)	67	56	85	64					
replacement: frequency highest	G	T	C	T					
replacement: frequency middle	T	A	G	A					
consensus base type	R	R	R	Y	|	R	Y	Y	Y
conservation (%)	89	74	89	92					
Trp									
consensus trp-box	A	G	T	T	|	A	A	C	T
base conservation (%) (half site)	73	65	73	38					
replacement: frequency highest	C	A	C	G					
replacement: frequency middle	G	T	a/g	A					
consensus base type	R	R	Y	Y	|	R	R	Y	Y
conservation (%)	81	85	92	46					

Figure 3. Conservation of bases at each position in natural met and trp boxes. The percentage conservation for bases and base class (i.e. as pyrimidine (Y) or purine (R)) have been calculated for the half box by averaging over the dyad symmetry axis (shown by |). Note that 1 = 8,2 = 7 etc. and that this numbering differs from the standard *trp* operator numbering in figure 2.

in *S. Typhimurium*. They do, however, show interesting trends in sequence conservation. An analysis of base conservation in terms of the half-met box is shown in figure 3.

A_3 is the most strongly conserved base (85%) and is occasionally replaced by C or G but almost never T. G_2 is moderately conserved, and is most likely to be replaced by T. In these two positions, the class of the replacement base, purine or pyrimidine, is as expected from random chance. At the other two positions, however, A_1 and C_4, there is moderate conservation of the actual base, but preferred replacements are almost invariably of the same class, G and T respectively. Overall conservation of base class at each position is therefore uniformly high (figure 3), giving the sequence RRRY (R = purine, Y = pyrimidine) for the half-met box. Inspection of the natural operators reveals that whereas the centres of the boxes often contain exact copies of the consensus (ACGT), of the 31 junctions between boxes there is no case of an exact CTAG sequence. The dinucleotide step across the junction, however, is usually TA (42%) and almost always YR (77%).

The three natural *trp* operator sequences (Bass *et al.* 1987), given in figure 2 show some relation to met boxes, but with important differences. They also contain 8 b.p. repeats homologous to a consensus sequence, AGT(t/g)(a/c)ACT (where (t/g) indicates 'T or G' etc.), which we will refer to as a 'trp box' for the purposes of this discussion. There is a similar trend of higher homologies in the central boxes. A symmetrical analysis of half-trp boxes (figure 3), shows that bases $A_1G_2T_3$ are as well conserved as in met boxes, but T_4 occurs with the same frequency as G_4. Inspection of the full sequences shows this to be due to a strong preference for G at position −9t and C at +9t(−9t and +9t in the *trp* numbering scheme correspond to position 4 in the half-met/trp box) (figure 2). Elsewhere the related half-trp box base tends to be T_4. Conservation of base class is also good for the first three positions ($R_1R_2Y_3$), with the fourth random. It is striking to note that the consensus trp box is 50% identical to a met box, as it shares the CTAG sequence, and would therefore qualify for inclusion in a table of *met* operators. In contrast to *met*, however, inspection of the *trp* operators shows exact copies of CTAG in 40% of the possible cases. For *E. coli* to regulate its amino acid metabolism correctly, the *met* and *trp* repressors must be able to distinguish between these operators despite their similarities.

4. REPRESSOR–OPERATOR BINDING

A series of experiments to examine *met* repressor function *in vivo*, using fusions of natural *metC* and *metF* operators with a *lacZ* reporter gene, showed that progressive mutation of non-consensus operator bases to consensus bases increased repression efficiency, and that single base insertions reduced it significantly (Davidson & Saint-Girons 1989). It was additionally demonstrated by footprinting that all 40 b.p. of the *metF* operator were protected in the repressor–operator complex, indicating that an array of repressors were bound. *In vitro* repressor binding experiments using 16 b.p. consensus synthetic operators cloned into various DNA fragment backgrounds, showed such targets were bound by holorepressor with $K_d \sim 10^{-8}$ M,

Downloaded from rstb.royalsocietypublishing.org

		[Met J] for half-maximal binding (nM)
consensus	**AGACGTCT AGACGTCT**	10
mutants		
1	gGACGTCc gGACGTCc	240
2	tGACGTCa tGACGTCa	760
3	cGACGTCg cGACGTCg	760
4	AcACGTgT AcACGTgT	960
5	AaACGTtT AaACGTtT	1200
6	AtACGTaT AtACGTaT	319
7	AGgCGCcCT AGgCGCcCT	119
8	AGtCGaCT AGtCGaCT	1200
9	AGcCGgCT AGcCGgCT	306
10	AGAgcTCT AGAgcTCT	59
11	AGAtaTCT AGAtaTCT	480
12	AGAatTCT AGAatTCT	480
control		960

Figure 4. *In vitro* repressor binding affinities for systematic mutations of a consensus two met box operator. Repressor concentrations required for 50% complex formation in gel retardation assays were measured for the sequences shown, inserted in a polylinker fragment. Lower case letters indicate bases differing from the consensus met box. The control was the polylinker sequence without the insert.

but that binding was 1000-fold weaker for aporepressor in the absence of SAM (Phillips *et al.* 1989). Footprinting again showed that arrays of repressors bound to the DNA, and that binding was strongly cooperative with respect to repressor concentration. Single met box 8 b.p. sites are not bound with high affinity, demonstrating a minimum requirement of 16 b.p. for a *met* operator.

This led to the proposal of a model for cooperative tandem binding of *met* repressor arrays to extended *met* operators (Phillips *et al.* 1989). Because the consensus operator contains two sets of equivalent dyad axes, a series of repressors placed on such an operator, such that each repressor is centred with its molecular dyad coincident with one of a set of equivalent local dyads, forms a regular array. The second set of local dyads in the operator additionally relate adjacent repressors in the array. Repressors lie on the central met box dyads (arrows in figure 2), rather than those between boxes, as demonstrated by binding assays and crystal structure determination of the repressor–operator complex (Somers & Phillips 1992). In three dimensions, the repressor array forms a left-handed superhelix wrapped around the duplex B-DNA, with a relative rise and rotation between molecules of 8 b.p. (ca. 27 Å) and 90°. Protein–protein contacts between adjacent repressors in the array account for the observed cooperativity in the system. Site-directed mutagenesis (He *et al* 1992; Davidson & Saint-Girons 1989), confirmed this as the origin of the cooperativity, and showed that it is essential to repressor function *in vivo*.

Although the symmetry is less perfect, similar *trp* repressor arrays were proposed for the two longer *trp* operators by Kumamoto *et al.* (1987), and have been subsequently observed in a crystal structure (Lawson & Carey 1993). The similarities of the *met* and *trp* arrays were discussed by Phillips & Stockley (1994).

Two approaches have been taken to determine the relative importance of each base in a *met* operator with respect to repressor recognition. In the first, the symmetry of the consensus sequence was exploited to generate the 12 possible symmetrical variants with single base changes (Phillips *et al.* 1993; Wild *et al.* 1996). Their sequences are shown in figure 4, together with their relative repressor affinities in a gel retardation assay.

All mutations result in reduced affinities, but some bases are clearly more sensitive than others. In another study, He *et al.* (1996) used *in vitro* selection to generate DNA sequences showing tight binding to *met* repressor in a gel mobility assay, from a pool of random sequence 20-mer oligonucleotides. The top 75 tight binding sequences found contained many examples of perfect consensus met boxes (figure 5). The base frequencies in the evolved met boxes largely mimic those observed in natural operators, with some important differences. A_3 is strongly conserved as in natural operators, and the conservation of G_2 is even higher. This is consistent with the results of systematic mutation (figure 4) where changes at these two positions are the most deleterious. Exact copies of the CTAG sequence occur in 80% of the cases, while it is never observed in natural operators. This implies that other selective pressures in addition to binding affinity are operating *in vivo*. In an *in vitro* evolution experiment for *trp* repressor sites, Czernik *et al.* (1994) used a selection based on *trp* repressors bound to a chromatography column, effectively precluding the formation of tandem arrays. The evolved consensus for their tight-binding sequences is also shown in figure 5, from which they concluded that exact CTAG sequences are the hallmark of a *trp* repressor site. They also observe strong conservation of G at position −9t (*trp* operator numbering). This is in accord with the natural operator sequences. Why then do the two repressors select identical CTAG sequences *in vitro* but not *in vivo*, and how does *E. coli* avoid crosstalk between the *met* and *trp* control systems? To answer such questions we need to consider the three-dimensional structures of the respective repressor–operator complexes.

Downloaded from rstb.royalsocietypublishing.org

In Vitro evolution of _met_ and _trp_ repressor binding sites

Met sites (75 sequences)

selected consensus

Trp sites (56 sequences)

selected consensus

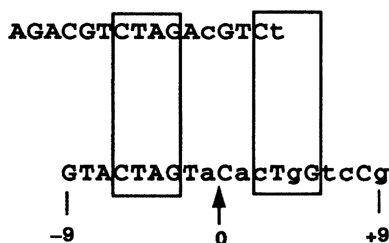

Figure 5. Consensus binding sequences produced by *in vitro* evolution of DNA targets for *met* (He, *et al* 1996) and *trp* (Czernik *et al.* 1994). Bases in upper case are more than 70% conserved, and CTAG regions are boxed. (the *trp* operator numbering scheme is shown for the *trp* sequence).

5. THREE-DIMENSIONAL STRUCTURE OF *MET* REPRESSOR AND ITS COMPLEXES

Crystal structures have been determined for *met* repressor in several crystal forms. The aporepressor and holorepressor have each been solved in two different crystal lattices, and their structures refined to crystallographic R factors under 20% at resolutions of 1.8–2.2 Å (Rafferty *et al.* 1989; Phillips 1992; Phillips *et al.* 1993; Strathdee 1993). The crystal structure of a complex of repressor with a two met box consensus operator fragment has also been determined at 2.8 Å resolution, and refined to R = 22% (Somers & Phillips 1992).

The overall structure of the holorepressor (Rafferty *et al.* 1989) is shown in figure 6, and has an unusual fold, which it shares only with the related Arc and Mnt repressors from bacteriophage P22 (Breg *et al.* 1990; Raumann *et al.* 1994). It is a symmetrical dimer with intertwined subunits, which can not be separated without substantial unfolding. Each subunit contains flexible β-hairpin (residues 12–20) labelled 'loop' in figure 6, leading into a β-strand (20–29). This strand pairs with the symmetry-related strand of the other subunit to form a two stranded antiparallel β-sheet, or β-ribbon. The rest of the subunit consists of three α-helices A (30–45), B (52–66) and C (86–94) linked by loops of various lengths. SAM binds at two independent symmetry-related sites on the opposite face of the repressor to the β-ribbon. The purine ring is inserted into a pocket next to the B helix, and the methionine moiety lies on the protein surface. The positively charged methylated sulphur atom lies at the C-terminal end of the B helix, presumably stabilized by the helix dipole (Hol *et al.* 1978) and partly screened from solvent by the methyl group itself. Apart from the

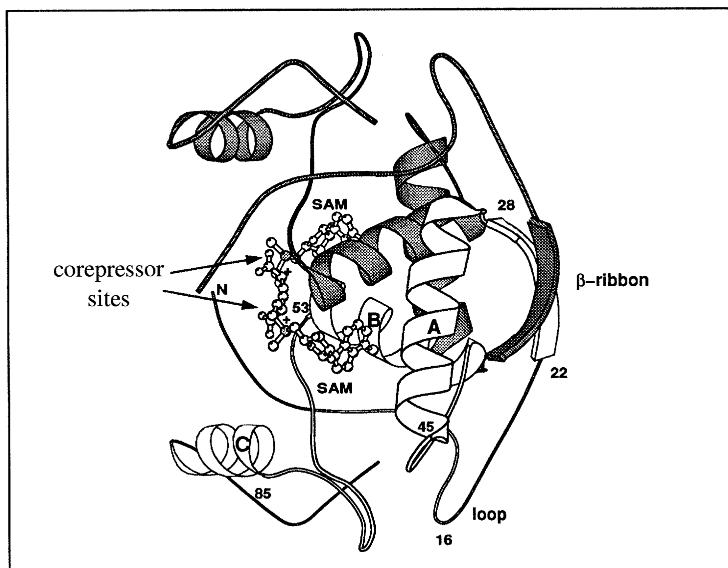

Figure 6. Overall structure of the *met* holorepressor, with the molecular dyad horizontal and in the plane of the page. The protein is shown in a ribbon representation and the SAM molecules as ball and stick. One subunit is shaded. The β-ribbon DNA binding motif is on the right. The white subunit has sequence numbers, and elements of secondary structure labelled. (Drawn using the program MOLSCRIPT Kraulis 1991). Redrawn from Phillips *et al.* 1993 with permission.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES
PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Downloaded from rstb.royalsocietypublishing.org

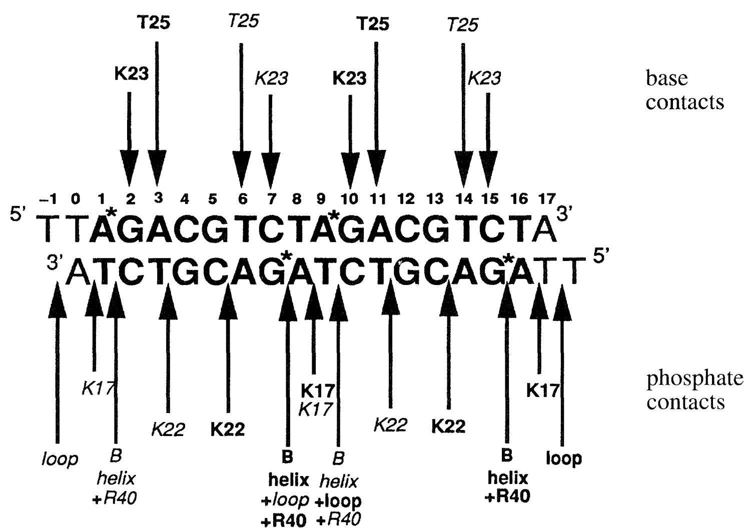

Figure 7. Sequence and numbering for the synthetic oligonucleotide used in crystallization of the *met* repressor–operator complex. The lower strand is related to the upper one by dyad symmetry, and is exactly equivalent in the structure. Contacts to the repressor are shown schematically, with direct base contacts above the line, and contacts to phosphates below. Major contacts to the phosphodiester backbone are made by the B helix and 12–20 loop. Italicized labels indicate contacts to the opposite strand. Redrawn from Phillips *et al.* 1993 with permission.

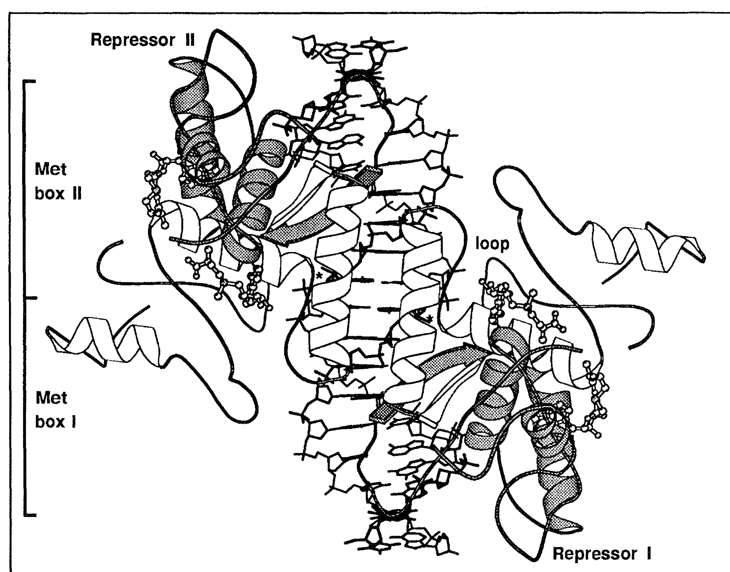

Figure 8. The structure of the repressor–operator complex. Two repressors, I and II, bind to the DNA fragment, at the lower right and upper left respectively, related by a dyad axis passing through the centre of the complex between met boxes. The local intramolecular dyads of the repressors coincide with local dyads in the met box centres. Repressor β-ribbons occupy the DNA major groove at lower right and upper left. The A helices form a long, antiparallel protein–protein contact above the minor groove in the centre of the diagram. The phosphate marked (*) is adjacent to the overwound central TA step, and is displaced from its expected position. The four SAM molecules lie on the outer surface of the complex, remote from the DNA. Redrawn from Phillips *et al.* 1993 with permission.

flexible loop (12–20), and one or two residues at the N-terminus, the repressor structure is essentially identical in all crystal forms, regardless of the presence or absence of SAM or DNA.

The structure of the complex of *met* repressor with a synthetic 19-mer oligonucleotide containing two consensus met boxes (see figure 7 for sequence), shows two repressors bound to the DNA (figure 8), each lying with its intramolecular dyad axis coincident with the central dyad of a met box i.e. between bases 4–5 and 12–13, from the major groove side. The two-stranded β-ribbon in each case is inserted into the DNA major groove, where its amino acid side-chains are able to contact the DNA bases directly. The antiparallel A helices of adjacent repressors mediate a protein–protein contact across the central dyad axis of the complex between the two met boxes. This contact is responsible for most, if not all, the cooperativity of the system, and its disruption by site-directed mutagenesis of the repressor results in greatly reduced DNA binding affinity *in vitro* and total loss of repression *in vivo* (He *et al.* 1992). The interface formed by a single repressor with the DNA buries only 648 Å^2 of solvent accessible surface area, insufficient to stabilize a high affinity

Downloaded from rstb.royalsocietypublishing.org

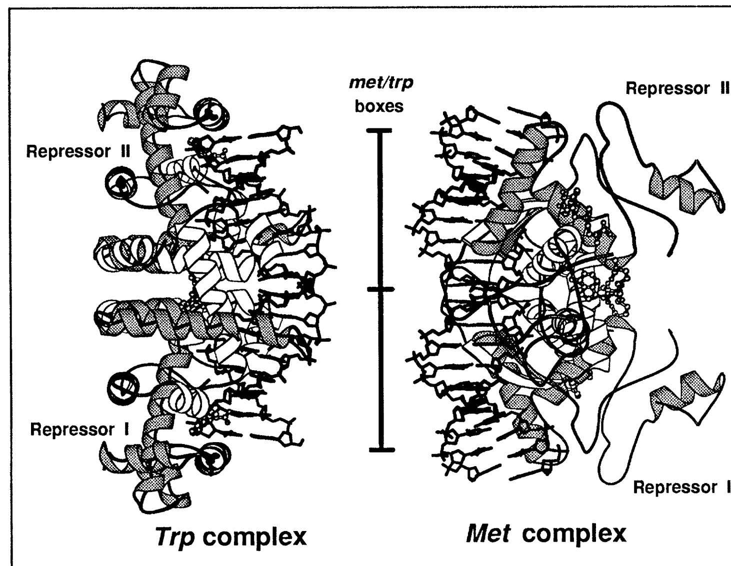

Figure 9. Tandem *trp* and *met* complexes viewed with the DNA in the same orientation. The *met* complex is as in figure 8, but rotated 90 ° about the vertical axis. The extent of the two *met*/*trp* boxes is shown by the vertical bar. The central CTAG sequence is in the minor groove facing *met* repressor and in the major groove facing the *trp* repressor reading heads. Corepressors are drawn as ball and stick in both cases.

complex (Janin 1995), but the additional protein–protein interface raises the total to 1732 Å^2 for the tandem complex, which is in the expected range. This shows why a minimum of two repressors are needed for complex formation. The corepressor lies on the outside surface of the complex, distant from the DNA. Calculations (Phillips & Phillips 1994) and binding studies (Parsons *et al.* 1995) are consistent with activation of DNA binding being achieved by long-range electrostatic interactions between the negatively charged phosphates, and the positive charges present on the sulphur atoms of the corepressors.

Three regions of the repressor make extensive contacts with the DNA bases and phosphates, namely the 12–20 loop, β-ribbon and the N-terminus of the B helix. The only direct contacts to base pairs are made by two side-chains from each strand of the β-ribbon. Lys23 Nζ donates hydrogen bonds to O_6 and N_7 of the G paired to C_7 (G_{10} of the bottom strand in figure 7), and Thr25 $O_{\gamma 1}$ donates a hydrogen bond to N_7 of A_3. The symmetry of the system generates equivalent contacts to bases G_2 and A_3 of all half-met boxes on both strands (figure 7). Model building alternative bases in these two positions using computer graphics readily yields consistent qualitative explanations of the observed affinities in the systematic operator mutation experiment (figure 4). The most striking observation is that it is impossible to replace A_3 with T, as its 5-methyl group would cause a steric clash with Thr25 too severe to be relieved without major disruption of the protein–DNA interface. Although there are no direct contacts from side-chains to bases C_4 of the half boxes, there is sufficient space for ordered solvent molecules in the groove which, although not observed at this resolution, might mediate specific contacts. No direct contacts are made to the bases in the central T_8A_9 dinucleotide step, corresponding to A_1 of the half-met box, which are exposed in the major groove at the back of the complex, away from the protein. This step is heavily overwound (helical twist 44 °), as are the TA steps in poly(AT) (Klug *et al.* 1979). Associated with this overwinding is a 2 Å shift of the adjacent phosphate (marked with * in figures 7,8) from its expected position in regular B-DNA. This phosphate forms a tight interaction with the N-terminus of the B helix, that would not be possible without this shift. As such overwinding is more energetically favourable in YR, and especially TA, steps where stacking energies are weak, it provides indirect recognition of base type at this position in the operator.

In contrast to *met*, *trp* repressor (figure 9) is an all α-helical protein dimer with tightly interlocked subunits (Schevitz *et al.* 1985; Zhang *et al* 1987; Lawson & Sigler 1988; Luisi & Sigler 1990). The first three helices of each subunit (A,B,C) form a rigid core whose structure is undisturbed by binding corepressor, but the remaining helices (D,E,F) form two flexible 'reading heads' that protrude from the structure and contain 'helix-turn-helix' DNA-binding motifs formed from helices D and E. The distance between the reading heads varies from 25 Å in the aporepressor to 34 Å in one of the holorepressor structures, and is affected both by the presence of corepressor and crystal contacts. The larger separation is ideal for inserting of the heads into successive turns of the major groove of B-DNA, an arrangement confirmed by the subsequent structure determination of a repressor–operator complex (Otwinowski *et al.* 1988). The corepressor binds on the DNA-binding face, wedged between helix E of the reading head and the central core of the repressor, tending to push the heads apart and stabilize the active conformation both structurally and dynamically.

In the crystal structure of a single *trp* repressor–operator complex (Otwinowski *et al.* 1988), where the DNA target is closely related to the central 18 b.p. of the *trpEDCBA* operator, the repressor dyad is co-

Downloaded from rstb.royalsocietypublishing.org

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Table 1. *Comparison of* met *and* trp *repressors*

	met	*trp*
operators	many	few
consensus sequence	AGACGTCT	AGTTAACT
tandem binding	essential	possible
repressor	2 × 104 aa dimer	2 × 107aa dimer
structural class	$\alpha+\beta$	α
DNA binding motif	β-ribbon	helix-turn-helix
base contacts	direct	water mediated
indirect readout	yes	yes
operator sequence features	CTAG (imperfect)	CTAG (perfect)
	A_3 never T_3	T_3
corepressor	SAM	Trp
activation by corepressor	electrostatic	structural

incident with the central dyad of the trp box, but approaches from the minor groove side i.e. opposite the equivalent *met* repressor site. This places the two DNA reading heads in the adjacent turns of the major groove, where the bases in the CTAG sequences are accessible. Bases A_1 and G_2 are recognized by water mediated, but specific, contacts to protein side-chains on the reading heads. An additional direct contact is made to G_{-9t} (*trp* operator numbering). A large number of contacts to DNA phosphate groups allows recognition of the conformational preferences of the operator. For instance, the weak TA step in this case shows strong roll rather than overwinding. The corepressor not only stabilizes the active conformation, but binds in the protein–DNA interface, orienting side-chains that contact the DNA, and itself making a direct contact to a phosphate. Corepressor activation is thus readily explained in direct structural terms. The solvent accessible area buried in the complex is 2900 Å^2, sufficient for high affinity complex formation in the absence of tandem binding. Tandem binding is, however, observed in the repressor–operator complex reported by Lawson & Carey (1993). Contacts between adjacent repressors are stabilized in part by interactions of ordered N-terminal arms. These arms are disordered in free repressor and the single repressor–operator complex. Truncation of the N-terminal arms, however, results in repressors defective in repression *in vivo*, and in *trp* operator binding *in vitro* (Hurlbert & Yanofsky 1992), implying a functional role for cooperative tandem binding. Repressor–operator contacts are similar in single and tandem complexes. Figure 9 shows the tandem *met* and *trp* complexes in similar orientations.

6. DISCRIMINATION BETWEEN *MET* AND *TRP* OPERATORS

The complexity of the two systems renders a complete description of operator discrimination extremely difficult. They have evolved together, and represent different, but related solutions to similar problems of regulation. A number of their contrasting features are summarized in table 1. The two repressors approach their respective operators at equivalent sites, but from opposite sides of the DNA. *Met* repressor places its β-ribbon in the major groove at the middle of the box, and makes direct base contacts to read the DNA sequence in this region. It specifically prefers A_3, and cannot tolerate T_3. At the junctions between boxes it detects conformational properties of its preferred CTAG sequences, but in the absence of direct base contacts it can tolerate related sequences with similar properties. *Trp* repressor, however, approaches the DNA from the minor groove side, and places its α-helical DNA reading heads in the major groove at either side, to make water mediated contacts to bases at the junctions between boxes. It recognizes exact CTAG sequences at the junctions, and tolerates few deviations. It also prefers T_3 to other bases at this position, but is less specific for the central bases in the operator. Under evolutionary pressure, natural *trp* operators therefore contain exact CTAG sequences, and maintain the third position as T_3: Natural *met* operators avoid exact CTAG sequences, even though they form good *met* repressor binding sites, to inhibit *trp* repressor binding, and prefer A_3 but never T_3. This illustrates one of the differences between *in vitro* evolution techniques, that select for repressor binding, and natural evolution that selects for viability of the organism. The two systems have coevolved a delicate balance, and still have much to teach us about the subtleties of regulation.

We thank the many members of our groups and collaborators who have worked on the *met* system over the years, including C. W.G. Boys, Y.-Y. He, I. Manfield, T. McNally, O. Navratil, I.G. Old, I.D. Parsons, K. Phillips, J.B. Rafferty, W.S. Somers, S.D. Strathdee, I. Saint-Girons, C. Wild. We thank the BBSRC and the University of Leeds for financial support. S. E. V. P. is an International Research Scholar of the Howard Hughes Medical Institute.

REFERENCES

Adams, J. M. & Capecci, M. R. 1966 N-formylmethionine-*s*-RNA as the initiator of protein synthesis. *Proc. natn. Acad. Sci. U.S.A.* **55**, 147–155.

Bass, S., Sugiono, P., Arvidson, D. N., Gunsalus, R. P. & Youderian, P. 1987 DNA specificity determinants of *Escherichia coli* tryptophan repressor binding. *Genes Dev.* **1**, 565–573.

Belfaiza, J., Parsot, C., Martel, A., Bouthier de la Tour, C., Margarita, D., Cohen, G. N. & Saint-Girons, I. 1986 Evolution in biosynthetic pathways: Two enzymes catalyzing consecutive steps in methionine biosynthesis originate from a common ancestor and possess a similar regulatory region. *Proc. natn. Acad. Sci. U.S.A.* **83**, 867–871.

Downloaded from rstb.royalsocietypublishing.org

Breg, J. N., van Opheusden, J. H. J., Burgering, M. J. M., Boelens, R. & Kaptein, R. 1990 Structure of Arc repressor in solution: evidence for a family of β-sheet DNA binding proteins. *Nature, Lond.* **346**, 586–589.

Cohen, G. N. & Jacob, F. 1959 Sur la répression de la synthèse des enzymes intervenant dans la formation du tryptophane chez *E. coli. CR Acad. Sci. Paris* **248**, 3490–3492.

Cohn, M., Cohen, G. N. & Monod, J. 1953 L'effet inhibiteur spécifique de la méthionine dans la formation de la méthionine synthase chez *Escherichia coli. C. R. Acad. Sci. Paris* **236**, 746–750.

Czernik, P. J., Shin, D. S. & Hurlburt, B. K. 1994 Functional selection and characterization of DNA binding sites for *trp* repressor of *Escherichia coli. J. biol. Chem.* **269**, 27869–27875.

Davidson, B. E. & Saint-Girons, I. 1989 The *Escherichia coli* regulatory protein MetJ binds to five 8 base pair repeats in the metF operator. *Molec. Microbiol.* **3**, 1639–1648.

Gunsalus, R. P. & Yanofsky, C. 1980 Nucleotide sequence and expression of *Escherichia coli TrpR*, the structural gene for the *trp* aporepressor. *Proc. natn. Acad. Sci. U.S.A.* **77**, 7117–7121.

He, Y.-Y., McNally, T., Manfield, I., Navratil, O., Old, I. G., Phillips, S. E. V., Saint-Girons, I. & Stockley, P. G. 1992 Probing *met* repressor-operator interactions in solution. *Nature, Lond.* **359**, 431–433.

He, Y.-Y., Stockley, P. G. & Gold, L. 1996 *In vitro* evolution of the DNA binding sites of *Escherichia coli* methionine repressor, MetJ. *J. Molec. Biol.* **255**, 55–66.

Hol, W. G. J., van Duijnen, P. T. & Berendsen, H. J. C. 1978 The α-helix dipole and the properties of proteins. *Nature, Lond.* **273**, 443–446.

Hurlburt, B. K. & Yanofsky, C. 1992 The NH_2 terminal arms of *trp* repressor participate in repressor/operator binding. *Nucl. Acids Res.* **20**, 337–341.

Janin, J. 1995 Elusive affinities. *Proteins: Struct. Func. Genet.* **21**, 30–39.

Joachimiak, A. J., Kelley, R. L., Gunsalus, R. P., Yanofsky, C. & Sigler, P. B. 1983 Purification and characterization of *trp* aporepressor. *Proc. natn. Acad. Sci. U.S.A.* **80**, 668–672.

Klug, A., Jack, A., Viswamitra, M. A., Kennard, O., Shakked, Z. & Steitz, T. A. 1979 A hypothesis on a specific sequence-dependent conformation of DNA and its relation to the binding of the lac-repressor protein. *J. molec. Biol.* **131**, 669–680.

Kraulis, P. J. 1991 MOLSCRIPT - a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.* **24**, 946–950.

Kumamoto, A. A., Miller, W. G. & Gunsalus, R. P. 1987 *Escherichia coli* tryptophan repressor binds multiple sites within the *AroH* and *trp* operators. *Genes Dev.* **1**, 556–564.

Lawson, C. L. & Carey, J. 1993 Tandem binding in crystals of a *trp* repressor/operator half-site complex. *Nature, Lond.* **366**, 178–182.

Lawson, C. L. & Sigler, P. B. 1988 The structure of *trp* pseudorepressor at 1.65 Å shows why indole propionate acts as a *trp* 'inducer'. *Nature, Lond.* **333**, 869–871.

Luisi, B. F. & Sigler, P. B. 1990 The stereochemistry and biochemistry of the *trp* repressor-operator complex. *Biochim. Biophys. Acta* **1048**, 113–126.

Marcker, K. & Sanger, F. 1964 *N*-formylmethionyl-*S*-RNA. *J. molec. Biol.* **8**, 835–840.

Old I. G., Phillips S. E. V., Stockley P. G. & Saint-Girons I. 1991 Regulation of methionine biosynthesis in the enterobacteriaceae. *Prog. Biophys. Molec. Biol.* **56**, 145–184.

Otwinowski, Z., Schevitz, R. W., Zhang, R.-g., Lawson, C. L., Joachimiak, A., Marmorstein, R. Q., Luisi, B. F. & Sigler, P. B. 1988 Crystal structure of *trp* repressor/operator complex at atomic resolution. *Nature, Lond.* **335**, 321–329.

Paik, W. K. & Kim, S. 1971 Protein methylation. *Science, Wash.* **174**, 114–119.

Parsons, I. D., Persson, B., Mekhalfia, A., Blackburn, G. M. & Stockley, P. G. 1995 Probing the molecular mechanism of action of co-repressor in the *E.coli* methionine repressor-operator complex using surface plasmon resonance (SPR) *Nucl. Acids Res.* **23**, 211–216.

Phillips, K. 1992 Structural studies of *E. coli met* repressor and its complexes. PhD Thesis, University of Leeds.

Phillips, K. & Phillips, S. E. V. 1994 Electrostatic activation of *E. coli* methionine repressor. *Structure* **2**, 309–316.

Phillips, S. E. V., Manfield, I., Parsons, I., Davidson, B. E., Rafferty, J. B., Somers, W. S., Margarita, D., Cohen, G. N., Saint-Girons, I. & Stockley, P. G. 1989 Cooperative tandem binding of *met* repressor of *Escherichia coli. Nature, Lond.* **341**, 711–715.

Phillips, S. E. V., Boys, C. W. G., He, Y.-Y., Manfield, I., McNally, T., Navratil, O., Old, I. G., Phillips, K., Rafferty, J. B., Somers, W. S., Strathdee, S. D., Saint-Girons, I. & Stockley, P. G. 1993 *E.coli met* repressor: DNA recognition by *β*-strands. In *Nucleic acids and molecular biology* (ed. F. Eckstein & D. M. J. Lilley), pp. 28–46. Berlin: Springer-Verlag.

Phillips, S. E. V. & Stockley, P. G. 1994 Similarity of *E. coli met* and *trp* repressors. *Nature, Lond.* **368**, 106.

Rafferty, J. B., Somers, W. S., Saint-Girons, I. & Phillips, S. E. V. 1989 Three dimensional crystal structures of the *Escherichia coli Met* repressor with and without corepressor. *Nature, Lond.* **342**, 705–710.

Raumann, B. E., Rould, M. A., Pabo, C. O. & Sauer, R. T. 1994 DNA recognition by *β*-sheets in the Arc repressor-operator structure. *Nature, Lond.* **367**, 754–757.

Saint-Girons, I., Duchange, N., Cohen, G. N. & Zakin, M. M. 1984 Structure and autoregulation of the *metJ* regulatory gene in *E. coli. J. biol. Chem.* **259**, 14282–14285.

Saint-Girons, I., Belfaiza, J., Guillou, Y., Perrin, D., Guiso, N., Bârzu, O. & Cohen, G. N. 1986 Interactions of the *E. coli* methionine repressor with the *metF* operator and with its corepressor *S*-adenosylmethionine. *J. biol. Chem.* **261**, 10,936–10,940.

Saint-Girons I., Parsot C., Zakin M. M., Bârzu O. & Cohen G. N. 1988 Methionine biosynthesis in Enterobacteriaceae: Biochemical, regulatory and evolutionary aspects. *CRC Crit. Rev. Biochem.* **23**, 1–42. (Suppl.)

Soll, D. 1971 Enzymatic modification of transfer RNA. *Science, Wash.* **173**, 293–299.

Somers, W. S. & Phillips, S. E. V. 1992 Crystal structure of the *met* repressor-operator complex at 2.8 Å resolution reveals DNA recognition by β-strands. *Nature, Lond.* **359**, 387–393.

Schevitz, R. W., Otwinowski, Z., Joachimiak, A., Lawson, C. L. & Sigler, P. B. 1985. The three-dimensional structure of *trp* repressor. *Nature, Lond.* **317** 782–786.

Strathdee, S. D. 1993 Structural studies of proteins involved in the regulation of gene expression. PhD Thesis, University of Leeds.

Tabor, H., Rosenthal, S. M. & Tabor, C. W. 1958 The biosynthesis of spermidine and spermine from putrescine and methionine. *J. biol. Chem.* **233**, 907–914.

Zhang, R.-g., Joachimiak, A., Lawson, C. L., Schevitz, R. W., Otwinowski, Z. & Sigler, P. B. 1987 The crystal structure of *trp* aporepressor at 1.8 Å shows how binding tryptophan enhances DNA affinity. *Nature, Lond.* **327**, 591–597.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES