Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Specificity of TRAP–RNA interactions: crystal structures of two complexes with different RNA sequences

Nicholas H. Hopcroft,[a] Alice L. Wendt,[b] Paul Gollnick[b] and Alfred A. Antson[a]*

[a]York Structural Biology Laboratory, Department of Chemistry, University of York, York Y010 5DD, England, and [b]Department of Biological Sciences, State University of New York, Buffalo, New York 14260, USA

Correspondence e-mail: fred@ysbl.york.ac.uk

Received 5 October 2001
Accepted 18 February 2002

PDB References: TRAP–GAGUU complex, 1gtf, r1gtfsf; TRAP–GAGCC complex, 1gtn, r1gtnsf.

The *trp* RNA-binding attenuation protein (TRAP) regulates expression of the tryptophan biosynthetic genes in bacilli by binding to the leader region of the nascent *trp* operon mRNA. When activated by binding tryptophan, the 11-subunit circular TRAP molecule binds to a target sequence consisting of 11 (G/U)AG repeats, separated by two or three variable 'spacer' nucleotides. Reported here are two crystal structures of TRAP bound to RNAs containing 11 GAG repeats separated by UU and CC spacer nucleotides, determined at 1.75 and 2.50 Å resolution, respectively. These show the spacer regions of the RNA molecules to be highly flexible, making no direct hydrogen-bonding contacts with the protein. Comparison of these structures with the previous structure of TRAP bound to $(GAGAU)_{10}GAG$ RNA, in which the spacer nucleotides stack with each other close to the protein surface, shows that the RNA can adopt different conformations depending on the sequence of the spacer regions. This gives insight into the structural basis of the specificity of TRAP and into the mechanism of binding.

1. Introduction

In *Bacillus subtilis* and several other bacilli, expression of the tryptophan biosynthetic genes is regulated by the *trp* RNA-binding attenuation protein TRAP (Gollnick, 1994; Babitzke, 1997). The *trpEDCFBA* operon mRNA has a 203 base-pair 5′ leader sequence preceding the first structural gene, which contains inverted repeats capable of forming two alternative hairpin structures. TRAP regulates transcription of these genes by an attenuation mechanism (Shimotsu *et al.*, 1986). When activated by high intracellular levels of L-tryptophan, TRAP binds to the leader transcript, preventing the formation of the antiterminator hairpin and thus allowing the overlapping terminator hairpin to form, which inhibits continued transcription. When levels of L-tryptophan are lower, TRAP is not activated and the transcription-permitting antiterminator structure forms.

TRAP also regulates translation of several genes, by two different mechanisms. Binding of TRAP to the leader region of *trp* readthrough transcripts induces the RNA to form a downstream hairpin that sequesters the ribosome-binding site of *trpE* (Kuroda *et al.*, 1988; Merino *et al.*, 1995; Du & Babitzke, 1998). For several other genes including *trpG* (Yang *et al.*, 1995; Du *et al.*, 1997) and *yhaG* (Sarsero *et al.*, 2000), TRAP exerts translational control by competing directly for a binding site that overlaps the ribosome-binding site. Recently, it has been shown that TRAP's regulatory functions are inhibited by the anti-TRAP (AT) protein, which is expressed in response to uncharged $tRNA^{Trp}$ (Valbuzzi & Yanofsky, 2001).

```
 1     2     3     4     5     6     7     8     9     10    11    12
GAGcaaGAGcuGAGaacGAGugCAGggUAGauGAGaauGAGcGAGuuUAGcuGAGguGAG  B. stearothermophilus
GAGcgaGAGcaGAGagcGAGcgUAGggUAGauGAGaauGAGcGAGuuUAGcuGAGguGAG  B. caldotenax
UAGaauGAGuuGAGuuaGAGaaUAGggUAGcaGAGaauGAG....uuUAGuuGAGcuGAG  B. subtilis
UAGag.AAGauGAGaauGAGacUAGggUAGuuGAG.auGAG....uaGAGuuGAGcaGAG  B. pumilus
```

Figure 1
Multiple sequence alignment of the TRAP-binding site in the *trp* leader mRNA from various bacilli. The (G/U)AG triplets are numbered at the top of the figure and are shown in bold capital letters.

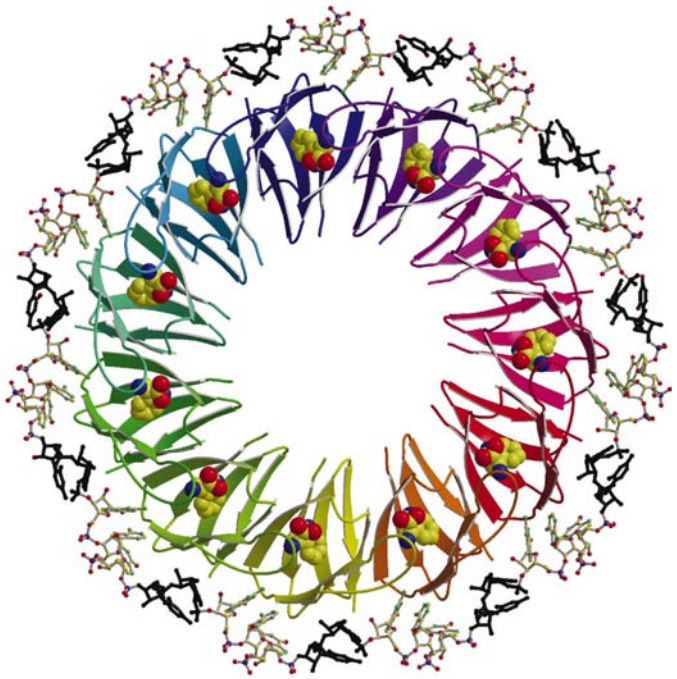

Figure 2
Diagram showing the TRAP 11-mer bound to RNA. Each protein subunit is shown in ribbon format and the 11 bound L-tryptophan ligands are shown in space-filling representation. The RNA, shown in ball-and-stick format, wraps around the edge of the protein ring. The spacer regions are highlighted in black. Figs. 2, 3 and 4 were generated using the program *MOLSCRIPT* (Kraulis, 1991).

The TRAP-binding site in the *trp* leader of *B. subtilis* consists of 11 (G/U)AG repeats, each separated by two or three more variable spacer nucleotides (Babitzke *et al.*, 1994). This sequence of trinucleotide repeats is conserved across several bacilli, although there are 12 such triplets in the thermophiles *B. stearothermophilus* and *B. caldotenax* (Fig. 1) and the binding site in the *trp* leader of *B. halodurans* is more divergent. The TRAP-binding sites in *trpG* and *yhaG* each contain eight (G/U)AG repeats and one AAG triplet and the spacer regions are more variable in length. The sub-optimal binding site in *trpG* may be a consequence of the fact that the product of this gene is also involved in folate biosynthesis and so some expression is required even in the presence of high tryptophan levels.

The structure of tryptophan-activated TRAP has been determined to 1.8 Å resolution (Antson *et al.*, 1995). TRAP forms an 11-mer of identical subunits arranged in a ring, with a total molecular mass of ∼91 kDa. The secondary structure of the protein is made up of anti-parallel β-sheets, β-turns and loop regions. The overall structure consists of 11 seven-stranded β-sheets, each containing four β-strands from one subunit and three β-strands from the next. The structure of TRAP bound to a 53-base RNA molecule consisting of 11 GAG triplets separated by AU spacers shows that the RNA wraps around the outer edge of the TRAP ring, with each repeat in the RNA sequence extending from one subunit to the equivalent point on the next (Fig. 2; Antson *et al.*, 1999). Only the GAG repeats make specific contacts with the protein, *via* hydrogen bonds to residues Glu36, Lys37, Asp39, Lys56 and Arg58 as well as van der Waals interactions. This confirms alanine-scanning mutagenesis results, which implicated Lys37, Lys56 and Arg58 in RNA binding (Yang *et al.*, 1997), and substitutions with nucleoside analogues, which showed the adenine and guanine in the second and third positions to be most important for binding (Elliot *et al.*, 1999). A model for the mechanism of RNA binding to TRAP has recently been proposed (Elliot *et al.*, 2001) in which a small subset of the 11 GAG repeats in the target sequence bind to TRAP first, facilitating cooperative binding of the rest of the RNA.

The importance of the spacer nucleotides (shown in black in Fig. 2) in TRAP-binding sites is unclear. The spacer nucleotides of natural TRAP-binding sites are predominately U and A (Baumann *et al.*, 1996). Moreover, binding studies suggest that TRAP has higher affinity for RNAs with U and A spacers (Babitzke *et al.*, 1996). The structure of the TRAP–GAGAU RNA complex shows that the AU spacer bases stack on each other (Antson *et al.*, 1999). However, *in vitro* selection experiments (SELEX) found that RNAs with pyrimidine spacer nucleotides are preferred, leading to the suggestion that TRAP favours RNA whose bases unstack easily (Baumann *et al.*, 1997). To investigate the role of the spacer bases in TRAP–RNA complexes, we obtained two crystal structures of TRAP bound to RNAs with GAG repeats separated by UU and CC spacers. The structures presented here show that when two pyrimidines are present, the sugar–phosphate backbone of the spacer region can have a flexible geometry. When compared with the TRAP–GAGAU RNA complex, they show that the spacer nucleotides can accommodate different positions on the surface of TRAP depending on sequence.

2. Materials and methods

2.1. Protein purification and RNA synthesis

TRAP from *B. stearothermophilus* was overexpressed in *Escherichia coli* and purified as previously described (Chen *et al.*, 1999). $(GAGUU)_{10}GAG$ and $(GAGCC)_{11}G$ RNAs were synthesized *in vitro* as previously for the complex with $(GAGAU)_{10}GAG$ RNA (Ziehler & Engelke, 1996; Antson *et al.*, 1999) and gel purified (Baumann *et al.*, 1996).

Table 1
X-ray data and refinement statistics.

Values in parentheses are for the outer resolution shell.

X-ray data		
TRAP–RNA complex	$(GAGUU)_{10}GAG$	$(GAGCC)_{11}G$
Space group	*C*2	
Unit-cell parameters		
a (Å)	142.1	145.8
b (Å)	111.5	111.7
c (Å)	138.3	138.7
β (°)	117.3	117.8
Resolution (Å)	50.0–1.75 (1.78–1.75)	20.0–2.5 (2.59–2.5)
Unique reflections	184884 (7925)	65753 (5337)
Redundancy†	3.0 (2.2)	1.8 (1.6)
Completeness (%)	97.6 (82.6)	94.6 (77.0)
$\langle I/\sigma(I)\rangle$‡	14.8 (1.7)	13.5 (2.4)
$R_{\rm merge}$§	0.073 (0.450)	0.050 (0.265)
Refinement and model correlation		
Resolution range (Å)	50.0–1.75	20.0–2.50
No. of reflections used in refinement	182643	63405
R factor¶ (%)	19.5	23.5
No. of reflections used for $R_{\rm free}$	1848	1312
$R_{\rm free}$¶ (%)	24.5	27.3
No. of protein atoms	12182	12200
No. of RNA atoms	968	880
No. of water molecules	1460	73
Average *B* factors (Å^2)		
Protein main chain	27.6	52.5
Protein side chain	32.2	53.3
RNA atoms	51.5	76.6
Water molecules	42.1	44.2
R.m.s. deviations from ideal geometry (targets in square brackets)		
Bond lengths (Å)	0.021 [0.020]	0.019 [0.020]
Bond angles (°)	2.0 [2.0]	2.0 [2.0]
Non-crystallographic symmetry†† (Å)		
Protein main chain	0.02 [0.10]	0.02 [0.10]
RNA atoms	0.18 [0.30]	0.18 [0.30]

† The average number of observations of the same reflection. ‡ The value of $\langle I/\sigma(I)\rangle$ crossed 3.0 within the 1.89–1.85 Å shell of the GAGUU data and the 2.59–2.50 Å shell of the GAGCC data. § The value of the merging *R* factor between equivalent measurements of the same reflection, $R_I = \sum|I - \langle I\rangle|/\sum I$. The $R_{\rm merge}$ value crossed 20% in the 1.93–1.89 Å shell of the GAGUU data and the 2.81–2.60 Å shell of the GAGCC data. ¶ Crystallographic *R* factor ($R_{\rm free}$) = $\sum||F_o| - |F_c||/\sum|F_o|$. †† R.m.s. deviation of atoms of the 11 subunits from the average structure.

2.2. Crystallization and data collection

Protein–RNA complex formation was monitored by protein mobility shift in non-denaturing 7.5% polyacrylamide gels. For crystallization, TRAP was mixed with the corresponding RNA molecule in a 2:1 molar ratio and crystallized as described previously (Antson *et al.*, 1999). The best crystals were obtained by the hanging-drop method with the protein–RNA complex (~15 mg ml^{-1}) in a solution containing 70 m*M* potassium phosphate pH 7.8 and 10 m*M* L-tryptophan. The reservoir contained 0.2 *M* potassium glutamate, 50 m*M* triethanolamine pH 8.0, 10 m*M* $MgCl_2$ and 8–11% monomethyl ether PEG 2000. To accelerate crystallization, a further gradient was induced by adding 0.4 *M* KCl to the reservoir after 1.5 µl protein solution was mixed with an equal volume of the reservoir solution. Crystals grew within 3 d and were vitrified and stored in liquid nitrogen immediately after growth, as the crystal's diffracting quality rapidly deteriorated in a further few days. The cryosolution contained 12% monomethyl ether PEG 2000, 30 m*M* triethanolamine pH 8.0, 6 m*M* L-tryptophan, 0.1 *M* potassium glutamate, 35 m*M* potassium phosphate pH 7.8, 5 m*M* $MgCl_2$ and 25% MPD. Crystals of both protein–RNA complexes belong to the space group *C*2, with similar unit-cell parameters (Table 1). X-ray data were collected at a wavelength of 0.946 Å at 120 K using synchrotron radiation at the ESRF ID14-4 beamline and an ADSC Quantum4 CCD detector. Data were processed using the programs *DENZO* and *SCALEPACK* (Otwinowski & Minor, 1997; Table 1).

2.3. Structure determination and refinement

Calculations were performed using the *CCP*4 suite of programs (Collaborative Computational Project, Number 4, 1994) with molecular replacement conducted by *AMoRe* (Navaza, 1994) and refinement performed with *REFMAC* (Murshudov *et al.*, 1997). Both structures were determined by molecular replacement using the structure of TRAP complexed with $(GAGAU)_{10}GAG$ RNA (Antson *et al.*, 1999; PDB code 1c9s). Model building was performed using the program *X-AUTOFIT* (Oldfield, 1996) implemented in *QUANTA* (Molecular Simulations Inc.). The validity of the refinement and the model-rebuilding process were monitored using $R_{\rm free}$ (Brünger, 1992). During the refinement, non-crystallographic symmetry restraints were imposed on the protein main-chain atoms of the two 11-subunit molecules and on the RNA atoms of the 11 repeating units.

3. Results and discussion

3.1. Overall structure description

The overall structure of the protein was the same for each new complex as it was for the previous complex with GAGAU RNA (Antson *et al.*, 1999). In each structure, all aminoacyl residues were placed except for the two at the N-terminus and the two at the C-terminus of each subunit. The electron density for each of the GAG triplets was clear, with no significant difference between individual repeats. The electron-density map of the complex with GAGUU RNA allowed the placing of the first of the two spacer nucleotides (Fig. 3*a*), although averaging was required to achieve this completely in some of the repeats. However, even the averaged map did not give clear electron density for the second spacer nucleotide. In the complex with GAGCC RNA, the phosphate group and ribose ring of the first spacer nucleotide were successfully positioned in the averaged map (Fig. 3*b*). Even when averaged maps were used, the electron density corresponding to the second spacer nucleotide or the base of the first was not interpretable. The absence of electron density corresponding to the second spacer nucleotide could only be explained by RNA flexibility and not by RNA degradation at this position, as RNA segments with less than five or six GAG triplets do not appear to interact with TRAP (Babitzke *et al.*, 1996).

3.2. Comparison of the G1 nucleotide in different structures

In both structures the G1 base stacks against the long aliphatic side chain of Lys37, which is disordered in uncomplexed TRAP (Antson *et al.*, 1995). Comparing these new structures with that of the TRAP–GAGAU complex, it can be seen that the side chain of Lys37 occupies slightly different positions in the three structures, but remains stacked with the G1 base (Fig. 4). Presumably, hydrophobic interactions between the two stabilize both the G1 base and the side chain of Lys37. This conclusion is supported by site-directed mutagenesis studies that replace Lys37 with other amino acids with aliphatic side chains of varying lengths (A. Wendt & P. Gollnick, unpublished results). The conformation of this nucleotide in the complex with GAGUU RNA matches that in the complex with GAGAU RNA. However, in the complex with GAGCC RNA the G1 base is shifted 0.5 Å away from the protein relative to its position in the other two structures. The ribose ring is twisted, so that the 5′ C is about 2 Å further from the protein, with the atoms of the ring all shifted by 0.5–1.5 Å depending on the particular subunit. The differences in the position of this nucleotide are complemented by shifts in the position of the side chain of Asp39, which was shown to form a weak hydrogen-bonding interaction with the exocyclic N atom of this base in the complex with GAGAU RNA (Antson *et al.*, 1999). In the complex with GAGCC RNA, this side chain is shifted about 1 Å towards the G1 base relative to its position in the complex with GAGAU RNA, with the position in the complex with GAGUU RNA intermediate between the two.

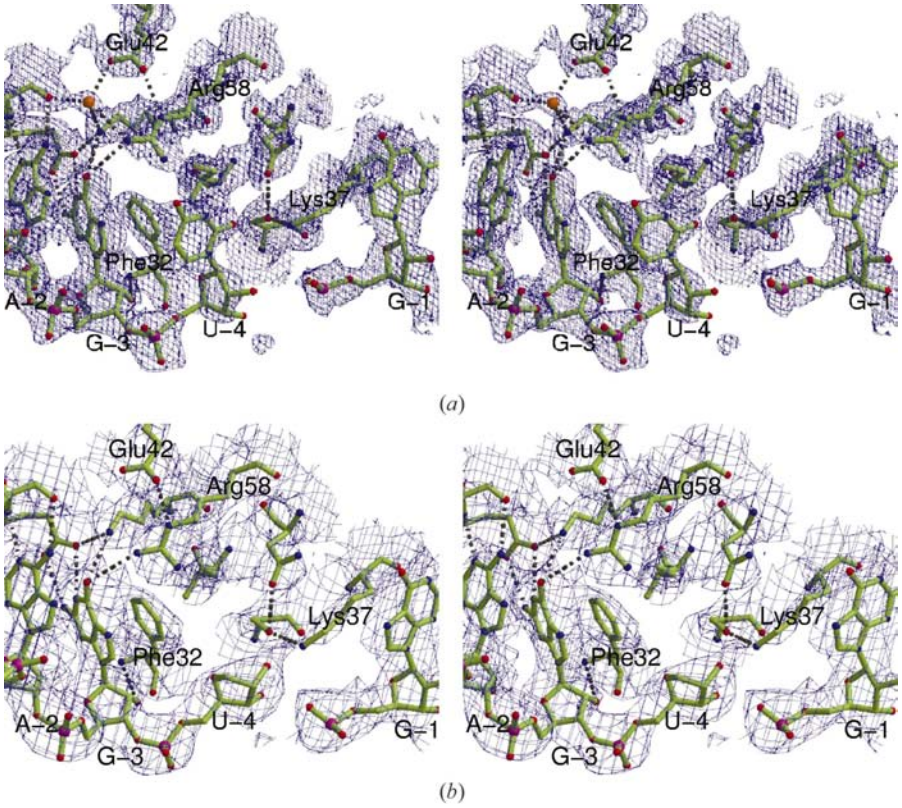

Figure 3
(*a*) Stereo diagram showing electron density calculated with likelihood-weighted $2|F_o| - |F_c|$ coefficients and contoured at the 1σ level corresponding to the TRAP–GAGUU complex. The figure shows an area around the spacer nucleotides which is typical of the 11 subunits. The U5 nucleotide shows no clear electron density and is assumed to be disordered. In five of the 11 subunits a water molecule, not present here, was placed in the model between Arg58 and the U4 base. (*b*) Stereo diagram showing 11-fold averaged likelihood-weighted $2|F_o| - |F_c|$ electron density corresponding to the spacer region of the TRAP–GAGCC RNA structure. The map is contoured at the 1σ level. The C5 nucleotide shows no clear electron density and is assumed to be disordered. The base of the C4 nucleotide is also unclear in the electron-density map and is assumed to have some rotational freedom around the sugar–phosphate backbone.

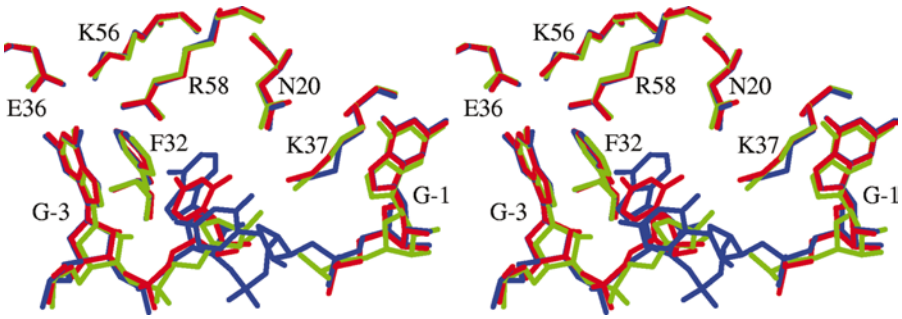

Figure 4
Stereo diagram showing overlapped structures of the spacer region of different RNA molecules bound to TRAP. The previously determined TRAP–GAGAU RNA complex is shown in blue and the TRAP–GAGUU RNA and TRAP–GAGCC RNA complexes are shown in red and green, respectively. The adjacent G3 and G1 nucleotides are also shown, as are selected aminoacyl residues, labelled with single-letter codes.

The reason why the G1 nucleotide is less closely bound to the protein in the complex with GAGCC RNA could be a consequence of greater flexibility in the neighbouring spacer nucleotides in this structure. The electron density of the spacer region is least clear in the structure with CC spacers, with the second spacer, adjacent to the G1 nucleotide, not placeable at all. The fact that the greatest effect on the G1 nucleotide is at the 5′ end, with the ribose ring twisted so that the effect on the rest of the nucleotide is minimized, supports this conclusion. Although changes in the conformations of the surrounding amino-acid side chains may compensate to some degree for the shift of the G1 base, the distance between the exocyclic N atom of this base and the side chain of Asp39 is increased from 2.7 Å in the complex with GAGAU RNA to 3.1 Å in that with CC spacers. The geometry of this arrangement is also

altered, resulting in disruption of the hydrogen bond between these two groups in the complex with GAGCC RNA. Although the fifth nucleotide is also disordered in the complex with GAGUU RNA, the position of the G1 base is not significantly affected in this structure and the geometry of the possible hydrogen bond with Asp39 is intermediate between those in the other two structures.

3.3. Comparison of the A2 and G3 nucleotides in different structures

In both new structures the A2 and G3 bases point towards the protein, with A2 stacking against G3, which in turn stacks against the aromatic ring of Phe32. The A2 base makes two hydrogen bonds with the main-chain atoms of Lys37; G3 makes three hydrogen-bond contacts with the side chains of Glu36 and Lys56. In some subunits, a fourth hydrogen bond forms between the G3 base and Arg58. This arrangement is identical to that in TRAP complexed with GAGAU RNA. This lack of variation is unsurprising considering the number of specific interactions involved. Experiments with nucleoside analogues (Elliot *et al.*, 1999) and alanine-scanning mutagenesis studies (Yang *et al.*, 1997) have shown the importance of these contacts in the overall capacity of TRAP to bind target RNA sequences. The sugar–phosphate backbone in this region also shows little variation between the three structures, adopting a right-handed coil conformation. This means that the G3 2′-hydroxyl, which has been shown to be important in RNA binding (Elliot *et al.*, 1999), makes the same hydrogen-bond contact with the main-chain N atom of Phe32 in all three complexes.

3.4. Conformation of the spacer nucleotides

The previously determined structure of TRAP complexed with GAGAU RNA showed the A4 base packing against the protein surface, only 3.1 Å from the side chain of Arg58 and 3.5 Å from the aromatic ring of Phe32, though perpendicular to it (Antson *et al.*, 1999). The U5 and G1 nucleotides exhibited the conformation of a left-handed coil with the bases pointing outwards (Pauling & Corey, 1953), which has been shown to be adopted by DNA when overstretched (Allemand *et al.*, 1998). This countered the twisting effect of the right-handed coil conformation of the other nucleotides and allowed the base of the U5 nucleotide to stack with that of A4 at the interface between these two conformations. This arrangement seemed independent of any direct influence from the protein, with neither base making hydrogen-bond or stacking interactions with any aminoacyl residues. Based on only this structure, it might be predicted that any two spacer nucleotides could stack in a similar way, accounting for the apparent relative unimportance of the sequence in these spacer positions. However, our results show that UU and CC spacers do not adopt a stacked conformation.

No clear electron density is observed corresponding to the base of the first spacer nucleotide in the complex with GAGCC RNA, suggesting that it makes no stabilizing interactions with the protein and instead has significant rotational freedom around the sugar–phosphate backbone. The ribose ring is clearly visible in the averaged electron-density map and shows a conformation different from that found in the GAGAU complex. All of the atoms of the ring are shifted 2.4–3.4 Å away from the protein relative to the previous structure. These shifts in position are smallest at the C3′ and C4′ positions and greatest at the C1′ atom, resulting in a twisting of the ring and suggesting that the base is exposed to the surrounding solvent.

When UU spacer nucleotides are present, the first spacer shows a conformation similar to that in the previous structure. The U4 base faces the protein, although it is positioned further from the side chains of Arg58 and Phe32 than A4 in the GAGAU structure. The ribose ring occupies a position midway between those found in the AU and CC structures. This allows the base to move up slightly to fill the space occupied by the larger A4 base in the complex with GAGAU RNA, but also for the 3′-hydroxyl to point away from the protein into the solvent where the second U is presumed to be.

Neither of the new structures showed clear electron density corresponding to the second spacer nucleotide. This nucleotide has therefore been left out of the final models and is assumed to be very flexible. This suggests that in both cases the base of this nucleotide makes no stabilizing interactions with the protein and is instead able to adopt different positions, sometimes facing the surrounding solvent. This is in contrast to the complex with GAGAU RNA, in which the sugar–phosphate backbone faces the solvent, with the bases of the spacer region stacked against the protein surface.

3.5. Specificity of TRAP–RNA interaction

We now have crystal structures of three different TRAP–RNA complexes, allowing analysis to an extent that is not possible for other systems of the ways in which TRAP binds single-stranded RNA. The importance of the A2 and G3 bases in TRAP binding is well established from a number of studies (Yang *et al.*, 1997; Antson *et al.*, 1999; Elliot *et al.*, 1999) and confirmed here by our finding that the positions of these bases is invariant in the three structures. The reason why G is preferred in the first position remains unclear. Only a G here is predicted to make a hydrogen bond to Asp39 (Antson *et al.*, 1999), but mutagenesis experiments suggest that the most important interaction is stacking of the base with the side chain of Lys37 (Yang *et al.*, 1997). The structures presented here show that even when a G is present, the hydrogen bond can be disrupted depending on the nature of the spacer region. The evidence suggesting that G1 is not as tightly held by the protein as A2 and G3 is supported by the variation in its position across the three structures. In the complex with CC spacers, where the spacer region is least ordered, the G1 base is ∼0.5 Å further from the protein than in the other two structures. In both the complex with GAGUU and that with GAGCC, the second of the two spacer nucleotides is less ordered than the first. The tight binding of the G3 nucleotide may tether the neighbouring spacer in the fourth position into a relatively ordered state, whereas the fifth nucleotide is more

free to move because neither of the adjacent nucleotides are tightly bound to the protein. The AU spacers are significantly more ordered than either CC or UU, stacking with each other, with the A4 base in van der Waals contact with the protein. This may be because of the increased size of the base in the fourth position relative to in the other two structures. According to this theory, larger purine bases in the fourth position, with their movement restricted by the tightly bound neighbouring G3 nucleotide and by their own size, would pack stably close to the protein surface. This would then affect the other spacer nucleotide which, although more free to move, would find stacking onto the preceeding purine most energetically favourable. However, further structures are needed to test whether a purine in the fourth position is sufficient to cause stacking of the spacer bases. The structures presented here though do show for the first time that the spacers can adopt different conformations with different degrees of flexibility.

3.6. Effect of spacers on RNA binding

The binding of TRAP to RNA has been shown to be driven by an increase in entropy which overcomes the associated unfavourable enthalpic change of +66.5 kJ mol^{-1} (Baumann *et al.*, 1996). This increase in entropy is thought to primarily arise from the liberation of water molecules associated with TRAP in the unbound state. With the crystal structures of three TRAP–RNA complexes, we can now address the question of why this favourable change in entropy is larger for some RNAs than for others, explaining the differences in TRAP's affinity for different sequences. Because TRAP only has a short time in which to halt transcription, the target sequence found in the *trp* leader should represent something close to the ideal TRAP-binding site. However, the *trp* leader also has to form the terminator and antiterminator hairpins under the appropriate conditions and the TRAP-binding site must remain single stranded, at least in part, as TRAP cannot bind RNA that forms a stable duplex (Babitzke *et al.*, 1996). These factors must influence the sequences of TRAP-binding sites. In total, the *trp* leaders from several bacilli contain 30 dinucleotide, ten trinucleotide and only two single-nucleotide spacers between adjacent (G/U)AG repeats (Fig. 1). These statistics agree well with the results of binding studies, which indicate a preference for dinucleotide repeats (Babitzke *et al.*, 1996). Among the 30 dinucleotide spacers, there is a bias towards pyrimidine–pyrimidine pairs, consistent with the results of SELEX experiments (Baumann *et al.*, 1997). Based on those results, it was proposed that TRAP favours pyrimidine spacers because they unstack more easily than purines. The two structures presented here show that pyrimidine spacers do unstack as the RNA binds to TRAP, whereas the structure of TRAP bound to GAGAU RNA shows that a single purine spacer can, in at least one case, cause stacking of the spacers in the TRAP–RNA complex (Antson *et al.*, 1999). Therefore, TRAP can accommodate both stacked and unstacked spacers, but the question remains as to how these two conditions affect the strength and kinetics of binding. The new structures show parts of the spacer region to be free to move when pyrimidine–pyrimidine spacers are present. The fact that they are disordered in these complexes means that the spacer nucleotides themselves should contribute to the increase in entropy upon binding, as the more ordered arrangement of stacked bases in unbound RNA is disrupted. This is not true when AU spacers are present, as these nucleotides are stacked in the complex. The freedom of pyrimidine spacers to move would also cause a greater increase in the entropy of the surrounding water molecules compared with the binding of GAGAU RNA. Our results therefore support a model in which pyrimidine spacers are conducive to binding because of their greater ability to unstack and the greater increase in entropy upon binding because of this unstacking. However, the natural sequence of the *B. subtilis trp* leader RNA shows a bias towards U or A in the spacer positions rather than U or C and filter-binding assays support the suggestion that these nucleotides are conducive to binding (Babitzke *et al.*, 1996). In these experiments, GAGUU RNA was found to form the most stable complex with TRAP, but RNAs with AU or AA spacers were also found to form strong complexes. The fact that GAGCU RNA does not bind easily to TRAP in such filter-binding experiments can be explained by the formation of stable secondary structure (predicted $\Delta G = -56.1$ kJ mol^{-1}) and results are not available for GAGCC or GAGUC RNA molecules. The reasons behind the strong binding of GAGAA RNA to TRAP remain unclear. Similarly, the high affinity of TRAP for GAGAU RNA but not for GAGGU is intriguing. Further structural studies of TRAP in complex with these RNA sequences are required to answer these questions.

4. Summary

The structures described here show that the conformation of the sugar–phosphate backbone in TRAP-bound RNA molecules can change to accommodate different spacer nucleotides. The stacking of spacer bases at the interface between right-handed and left-handed coiled conformations is not stable in all cases, since in both new structures the backbone of the second spacer nucleotide appears to be flexible and the bases are unstacked. The second spacer nucleotide is assumed to be able to adopt several different conformations in the solvent, rather than one conformation stabilized by interactions with the protein. The conformations of the neighbouring nucleotides support this model. The RNA is least ordered when CC spacers are present; in this case, the base of the first spacer is also free to adopt different conformations. When UU spacers are present, the base of the first spacer adopts a relatively stable conformation, though it is not as closely packed to the protein as the corresponding base in the AU structure. The lack of order in the new structures relative to that with AU spacers suggests that pyrimidine spacers are conducive to binding because of a greater increase in entropy upon binding.

This work was supported by a BBSRC CASE studentship with Ribotargets to NH and by Wellcome Trust research career development fellowship No. 053648 to AA. This work

was also supported by NIH Grant GM62750-01 and NSF Grant No. MCB-9982652 to PG. BBSRC provided financial support to the infrastructure of the YSBL through contract No. 87/SB09829. We thank Eleanor Dodson and Garib Murshudov for useful suggestions during structure refinement. We also thank Sean McSweeney for help during data collection and the ESRF for provision of synchrotron facilities through the block allocation to York.

References

Allemand, J. F., Bensimon, D., Lavery, R. & Croquette, V. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 14152–14157.
Antson, A. A., Dodson, E. J., Dodson, G. G., Greaves, R. B., Chen, X. & Gollnick, P. (1999). *Nature (London)*, **401**, 235–242.
Antson, A. A., Otridge, J. B., Brzozowski, A. M., Dodson, E. J., Dodson, G. G., Wilson, K. S., Smith, T. M., Yang, M., Kurecki, T. & Gollnick, P. (1995). *Nature (London)*, **374**, 693–700.
Babitzke, P. (1997). *Mol. Microbiol.* **26**, 1–9.
Babitzke, P., Stults, J. T., Shire, S. J. & Yanofsky, C. (1994). *J. Biol. Chem.* **269**, 16597–16604.
Babitzke, P., Yealy, J. & Campanelli, D. (1996). *J. Bacteriol.* **178**, 5159–5163.
Baumann, C., Otridge, J. B. & Gollnick, P. (1996). *J. Biol. Chem.* **271**, 12269–12274.
Baumann, C., Xirasagar, S. & Gollnick, P. (1997). *J. Biol. Chem.* **272**, 19863–19869.
Brünger, A. T. (1992). *Nature (London)*, **355**, 472–475.
Chen, X., Antson, A. A., Yang, M., Li, P., Baumann, C., Dodson, E. J., Dodson, G. G. & Gollnick, P. (1999). *J. Mol. Biol.* **289**, 1003–1016.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Du, H. & Babitzke, P. (1998). *J. Biol. Chem.* **273**, 20494–20503.
Du, H., Tarpey, R. & Babitzke, P. (1997). *J. Bacteriol.* **179**, 2582–2586.
Elliot, M. B., Gottlieb, P. A. & Gollnick, P. (1999). *RNA*, **5**, 1277–1289.
Elliot, M. B., Gottlieb, P. A. & Gollnick, P. (2001). *RNA*, **7**, 85–93.
Gollnick, P. (1994). *Mol. Microbiol.* **11**, 991–997.
Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.
Kuroda, M. I., Henner, D. & Yanofsky, C. (1988). *J. Bacteriol.* **170**, 3080–3088.
Merino, E., Babitzke, P. & Yanofsky, C. (1995). *J. Bacteriol.* **177**, 6362–6370.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
Oldfield, T. J. (1996). *Proceedings of the CCP4 Study Weekend. Macromolecular Refinement*, edited by E. Dodson, M. Moore, A. Ralph & S. Bailey, pp. 67–74. Warrington: Daresbury Laboratory.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Pauling, L. & Corey, R. B. (1953). *Proc. Natl Acad. Sci. USA*, **39**, 84–97.
Sarsero, J. P., Merino, E. & Yanofsky, C. (2000). *J. Bacteriol.* **182**, 2329–2331.
Shimotsu, H., Kuroda, M. I., Yanofsky, C. & Henner, D. J. (1986). *J. Bacteriol.* **166**, 461–471.
Valbuzzi, A. & Yanofsky, C. (2001). *Science*, **293**, 2057–2059.
Yang, M., Chen, X., Millitello, K., Hoffman, R., Fernandez, B., Baumann, C. & Gollnick, P. (1997). *J. Mol. Biol.* **270**, 696–710.
Yang, M., de Saizieu, A., van Loon, A. P. G. M. & Gollnick, P. (1995). *J. Bacteriol.* **177**, 4272–4278.
Ziehler, W. A. & Engelke, D. R. (1996). *Biotechniques*, **20**, 622–624.