The EMBO Journal Vol. 21 No. 8 pp. 1987–1997, 2002

Solution structure of the LicT–RNA antitermination complex: CAT clamping RAT

Yinshan Yang[1], Nathalie Declerck[1,2], Xavier Manival[1], Stéphane Aymerich[2] and Michel Kochoyan[1,3]

[1]Centre de Biochimie Structurale, CNRS-UMR 9955, INSERM-U554, Université de Montpellier I, 29 rue de Navacelles, F-34090 Montpellier and [2]Laboratoire de Génétique des Microorganismes, INRA, CNRS-URA 1925, F-78850 Thiverval-Grignon, France

[3]Corresponding author
e-mail: michel@cbs.univ-montp1.fr

LicT is a bacterial regulatory protein able to prevent the premature arrest of transcription. When activated, LicT binds to a 29 base RNA hairpin overlapping a terminator located in the 5′ mRNA leader region of the target genes. We have determined the solution structure of the LicT RNA-binding domain (CAT) in complex with its ribonucleic antiterminator (RAT) target by NMR spectroscopy (PDB 1L1C). CAT is a β-stranded homodimer that undergoes no important conformational changes upon complex formation. It interacts, through mostly hydrophobic and stacking interactions, with the distorted minor groove of the hairpin stem that is interrupted by two asymmetric internal loops. Although different in sequence, these loops share sufficient structural analogy to be recognized similarly by symmetry-related elements of the protein dimer, leading to a quasi-symmetric structure reminiscent of that observed with dimeric transcription regulators bound to palindromic DNA. Sequence analysis suggests that this RNA-binding mode, where the RAT strands are clamped by the CAT dimer, is conserved in homologous systems.

Keywords: antitermination complex/NMR/RNA binding/structure/transcriptional antiterminator

Introduction

The transcriptional antiterminator LicT from *Bacillus subtilis* belongs to a family of proteins that regulate gene expression by preventing premature transcription termination upstream of the genes they control. This family of antiterminator (AT) proteins contains >50 members, which have been shown or suggested to be involved in carbohydrate metabolism control in Gram-positive and Gram-negative bacteria. The prototypes of this AT family are *B.subtilis* SacY (Aymerich and Steinmetz, 1987) and *Escherichia coli* BglG (Mahadevan and Wright, 1987), which control the expression of the *sacB* gene and of the cryptic *bgl* operon, respectively (Figure 1). LicT, which is studied here, mediates the induction of the *B.subtilis licS* gene and the *bglPH* operon involved in the utilization of aryl-β-glucosides and β-glucanes (Kruger and Hecker, 1995; Schnetz *et al*., 1996).

Intrinsic transcription terminators contain two sequence motifs required for RNA release from the transcription machinery: a stable stem–loop hairpin immediately followed by a U-rich segment where abortive transcription complexes are disassembled (Gusarov and Nudler, 1999; Yarnell and Roberts, 1999). Besides these two features, conditional terminators targeted by the protein of the LicT/SacY family are preceded by a short RNA sequence (called RAT for ribonucleic antiterminator) overlapping by six nucleotides the terminator sequence. In the presence of their specific inducing sugar, the AT proteins are activated and bind to their cognate RAT sequence on nascent mRNAs. By inhibiting the complete formation of the large RNA stem–loop terminator structure, they enable the polymerase to proceed through a region where transcription is otherwise interrupted (Houman *et al*., 1990; Aymerich and Steinmetz, 1992; Arnaud *et al*., 1996).

The RAT target sequences of the AT proteins (Figure 1) are usually 30 nucleotides long and are proposed to adopt a hairpin structure with a variable apical loop and two asymmetric internal loops interrupting a central stem. Mutational analysis on different RATs has shown that base pairings in the stem are required for the antitermination function and that the non-conserved nucleotides within the internal loops are involved in the control of the specificity of the AT–RAT interaction. In contrast, RAT recognition by the AT appears largely independent of the length and nucleotide sequence of the apical loop (Aymerich and Steinmetz, 1992).

AT proteins from the LicT/SacY family have a modular structure. RNA recognition is embedded in the 55 amino acid N-terminal fragment. This domain, which was called CAT for co-antiterminator, presents constitutive antitermination activity *in vivo* as well as efficient and specific RNA binding activity *in vitro* (Manival *et al*., 1997; Declerck *et al*., 1999; Langbein *et al*., 1999). In the full-length protein, the activity of CAT is modulated by two homologous regulatory domains, called PRD1 and PRD2, which are phosphorylated reversibly on conserved histidines in response to the availability of carbon sources. It is believed that the phosphorylation state of the PRDs influences the tertiary and quaternary structure of the protein and thereby determines the ability of CAT to interact with its RAT target (van Tilbeurgh and Declerck, 2001).

The structure of the CAT domain from SacY and LicT has been solved previously by NMR and/or crystallography (Manival *et al*., 1997; van Tilbeurgh *et al*., 1997; Declerck *et al*., 1999). In solution as well as in crystals, CAT folds as a β-stranded symmetric dimer that shares no structural homology with other known RNA-binding motifs. Preliminary NMR footprint experiments have demonstrated that the protein surface interacting with

© European Molecular Biology Organization

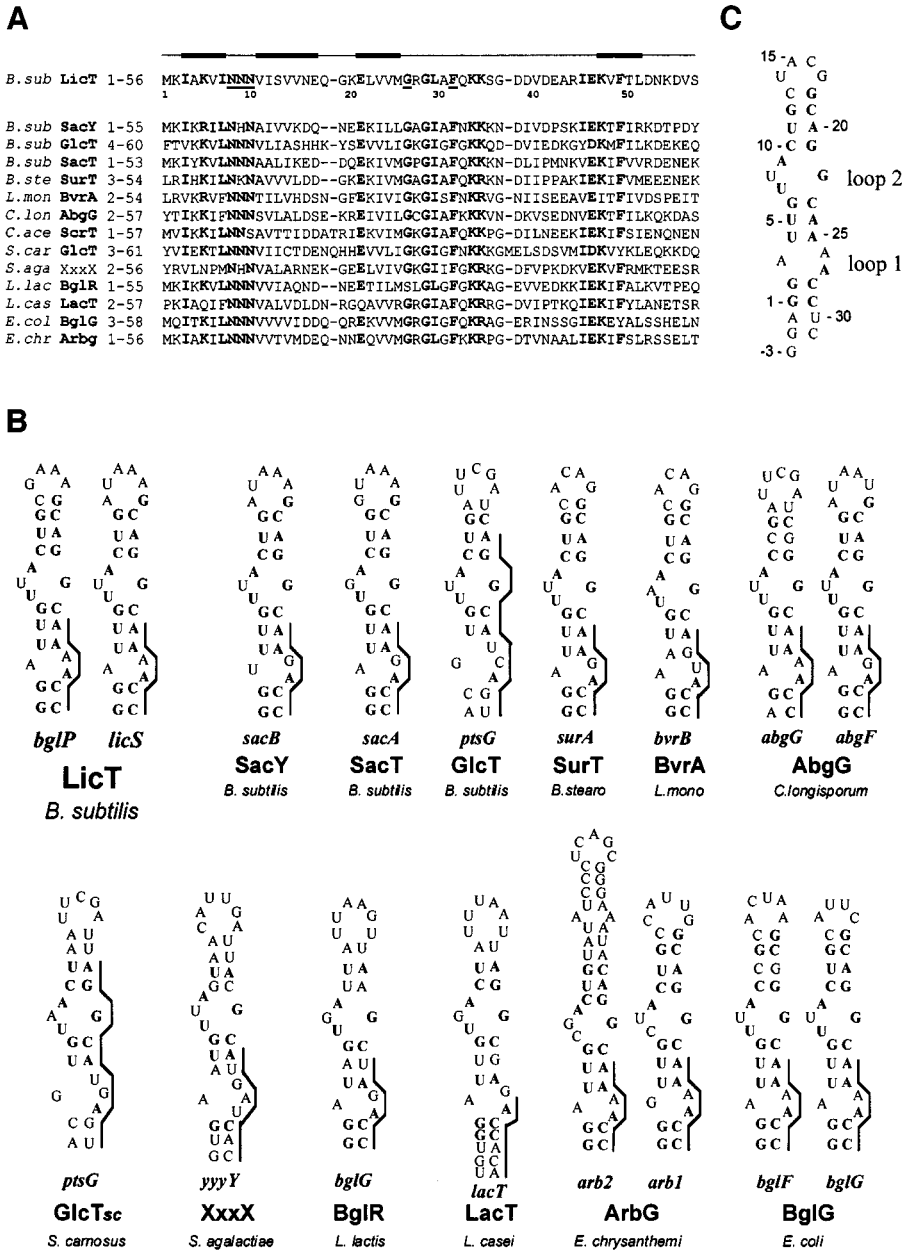

Fig. 1. (**A**) Amino acid sequence and comparison of the RNA-binding domain (CAT) of LicT and 13 representative members of the LicT/SacY family of transcriptional ATs from *Bacillus subtilis*, *Bacillus stearothermophilus*, *Listeria monocytogenes*, *Clostridium longisporum*, *Clostridium acetobutylicum*, *Staphylococcus carnosus*, *Streptococcus agalactiae*, *Lactococcus lactis*, *Lactobacillus casei*, *Escherichia coli* and *Erwinia chrysanthemi* (DDBJ/EMBL/GenBank; P.Glaser, personal communication). The residues forming the four β-strands are indicated by thick lines on top of the alignment. Residues underlined are involved in base-specific contacts with the RAT targets; residues in bold are conserved >90% in the 52 examined CAT sequences. (**B**) Secondary structure of the RAT targets that have been shown or suggested to interact with the CAT domain of the above ATs. Highly conserved nucleotides in the 45 examined RAT sequences are shown in bold. The line drawn on the right side of each hairpin indicates the nucleotides shared with the terminator sequence. The fold proposed for the *ptsG*-RAT target of the *B.subtilis* and *S.carnosus* GlcT ATs is discussed in the text. (**C**) The sequence and numbering scheme of the RNA oligonucleotide used in the present study.

RNA is located on one side of the dimer and that both monomers are involved in RNA recognition (Manival *et al.*, 1997). A puzzling question is thus how a symmetric dimer can recognize an obviously non-symmetric RNA motif. Will both monomers participate to the same extent in RNA recognition and, if so, how do identical residues in the dimer interact with distinct sites on the RNA? What is the structural basis of specificity of CAT–RAT recognition and how can the binding of CAT to the RAT target promote antitermination?

In order to address these questions, we have solved the NMR solution structure of the LicT RNA-binding domain complexed with its target RNA. Both the protein and nucleic partners remained practically unchanged upon complex formation. Strikingly, the complex conserves an apparent symmetry, each monomer recognizing in a very similar way a different asymmetric internal loop of the RNA. This is made possible by the very close conformation adopted by the loops despite their different sequence. Comparison of an extensive set of CAT–RAT sequences

suggests that this novel RNA-binding mode is very well conserved within the LicT/SacY family of ATs.

Results

Structure of the free and complexed protein

The affinity and specificity of several CAT–RAT complexes have been determined by gel shift assays and surface plasmon resonance measurements (Declerck *et al*., 1999; our unpublished results). Within the set of proteins and RNAs tested, the CAT domain from LicT presents the highest affinity for its RNA target (K_d near 10 nM compared with 3 μM for SacY) and forms a complex of longer lifetime. Indeed, whereas the NMR spectra of the SacY-CAT–RAT complex exhibit broad resonances, indicative of the formation of a short-lived complex on the NMR time scale (i.e. in the millisecond range), the spectra obtained for the LicT-CAT–RAT complex exhibit sharp lines, consistent with the formation of a long-lived (i.e. >10 ms), stable complex. A high-resolution structural NMR study was therefore undertaken on the more stable LicT-CAT–RAT complex.

The solution structure of the 56 amino acid fragment corresponding to the LicT RNA-binding domain was first solved using classical NMR techniques. The structure is very similar to that determined previously for LicT-CAT by crystallography (Declerck *et al*., 1999) and for SacY-CAT by NMR (Manival *et al*., 1997). CAT folds as a symmetric homodimer, each monomer being composed of a four-stranded antiparallel β-sheet. In the dimer, the β-sheets of the two monomers face each other to form an eight-stranded β-barrel covered on both sides by a long loop joining strand 3 to strand 4 of each monomer. The dimer is maintained by hydrophobic packing of the residues at the interface, including the well-conserved Phe48, by the antiparallel interaction between the last β-strand of each monomer, and by a possible salt bridge between the well-conserved Glu21 in one monomer and Lys46 in the other monomer, as previously observed in the LicT-CAT crystal structure.

The solution structure of the liganded protein was solved next using two- and three-dimensional homo- and heteronuclear NMR spectroscopy. As previously observed for SacY, the global architecture of the CAT dimer is maintained upon interaction with RNA. In the absence of RNA, due to the symmetry of the dimer, only 56 amino acid spin systems are observed in the NMR spectra of LicT-CAT. In contrast, in the presence of RNA, when the complex forms, the symmetry of the dimer is broken and most of the amino acid spin systems split into two components. The amino acids of each monomer are no longer equivalent and therefore give rise to two different sets of resonances (Figure 2). Paradoxically, this increased complexity of the spectra facilitated the NMR structure determination of the protein, since it was then possible to discriminate inter- from intramonomer nuclear Overhauser effects (NOEs).

Very few differences are observed between the three-dimensional structures of the ligand-free and complexed monomers, except for the tips of the two long loops joining β3 and β4, which move slightly towards the protein core in the complex. There is also a slight modification in the relative orientation of the two monomers, the β1–β2 turns at the edge of the dimer interface being 2 Å closer to each other in the complex than in the free protein.

Structure of the RNA

The non-exchangeable and exchangeable proton NOESY spectra of the free RNA confirmed the secondary structure of the molecule proposed after sequence and genetic analyses (Aymerich and Steinmetz, 1992). For the present structural study, the nucleotides corresponding to the apical loop of the *bglP*-RAT hairpin (G14–A17), which are not involved in the interaction (Aymerich and Steinmetz, 1992), have been substituted for nucleotides forming a hyperstable UACG tetraloop. The low resolution structure of the free RNA, obtained at low temperature (i.e. 7°C), confirmed the formation of this UACG apical tetraloop and of a regular stem with canonical base pairings interrupted by asymmetric loop 1 (A3, A26 and A27) and loop 2 (U7, U8, A9 and G22) (Figure 1C). Stacking of all three adenines in loop 1 of the free RNA was inferred from the couple of NOEs observed between the H2 aromatic protons of A26 and A3, A27 and A3, and A26 and A27. In asymmetric loop 2, formation of an U7–A9–G22 triplet with a GU wobble pair and a sheared AG pair was inferred from the presence of a couple of strong NOEs between the G22-imino and U7-imino protons and between G22-amino and A9-H8 protons.

In the presence of the protein, the signature of most of these secondary structure elements remains present in the spectra. The resonances of the imino protons involved in base pairing in the vicinity of the protein contact region are much sharper than in the free RNA, especially at high temperature (Figure 2). This indicates a slower exchange rate of these protons with water protons and might be due to either a higher stability of the base pairs or to a decreased accessibility of the water to the imino protons in the complex. However, since the major groove of the RNA is fully accessible to solvent, even when the complex is formed, we believe that the increased stability of the base pairs is mainly responsible for the reduced imino proton exchange observed. This would be in agreement with UV melting experiments performed on the SacY–RAT system, indicating that complex formation stabilizes the RNA hairpin (Manival *et al*., 1997).

The high-resolution solution structure of the complex (Figure 3) was obtained in a new stage of restrained molecular dynamic calculations (see Materials and methods). The nature of the constraints used during the final stage of modelling and the statistics concerning the structures used for structural analysis are summarized in Table I.

The most striking differences between the bound and free RNA concern loop 1, which is now formed of an A3–A27 sheared pair neighbouring the A26 nucleotide expelled from the helix core. In contrast, the conformation of loop 2, with a U7–G22–A9 base triplet and the bulged-out nucleotide U8, is not modified but just stabilized in the complex.

CAT–RAT recognition

LicT-CAT interacts with the minor groove of the double-stranded portion of the RNA containing the two asymmetric internal loops and the stem in between the two

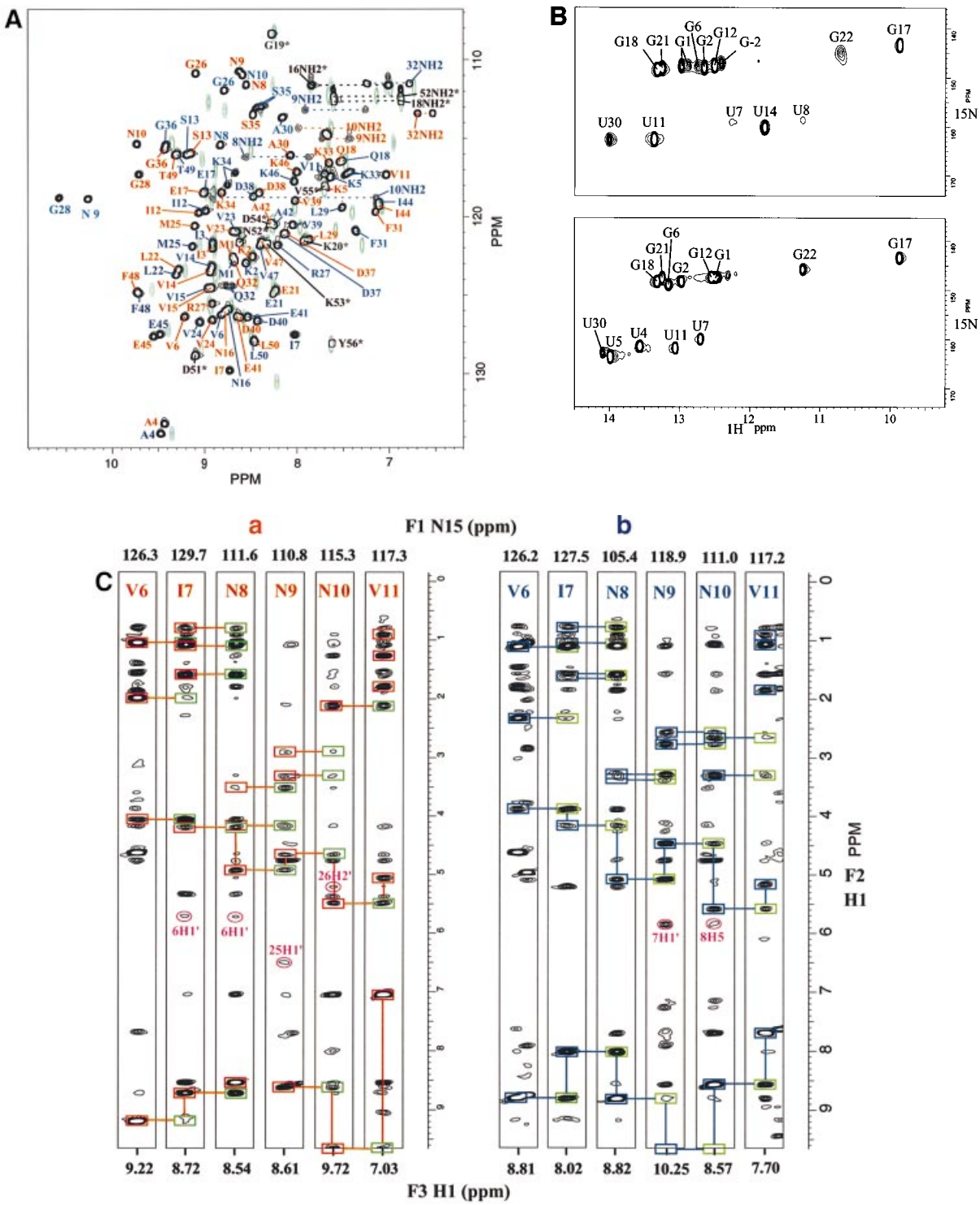

Fig. 2. (**A**) Superimposition of the ^{15}N-HSQC spectra of the amide region of the free (light green cross-peaks) and complexed (black cross-peaks) protein recorded at 28°C. In the presence of RNA (unlabelled in this experiment), the symmetry of the protein dimer is broken and most of the amide cross-peaks are split into two components. The names of the corresponding amino acids are in red and blue, depending on the monomer to which they belong. Cross-peaks not affected by the loss of symmetry (unsplit) are labelled in black. (**B**) ^{15}N-HMQC spectra of the imino proton region of the RNA alone (top) and in the presence of stoichiometric amounts of unlabelled protein (bottom). The spectra are recorded at different temperatures (7°C for free RNA and 28°C for the complex). Exchange broadening of imino protons cross-peaks is, however, weaker in the complex than in the free RNA (the cross-peaks of U7, G22, U4, U5 and G6 are much broader or even absent from the free RNA spectrum but present in the spectrum of the complex recorded at 21°C above). This is indicative of a stabilization of the base pairs in the complex (see text). (**C**) Sequential attributions of the protein peaks within each monomer (coloured in blue and red, as in A), using the 3D-NOESY-HSQC experiment recorded with a ^{15}N-labelled protein and an unlabelled RNA. Sequential protein cross-peaks (amide to HA and HB protons) are boxed in red and blue (intraresidue NOEs), and in green (sequential NOEs); intramolecular (protein–RNA) cross-peaks are circled in pink.

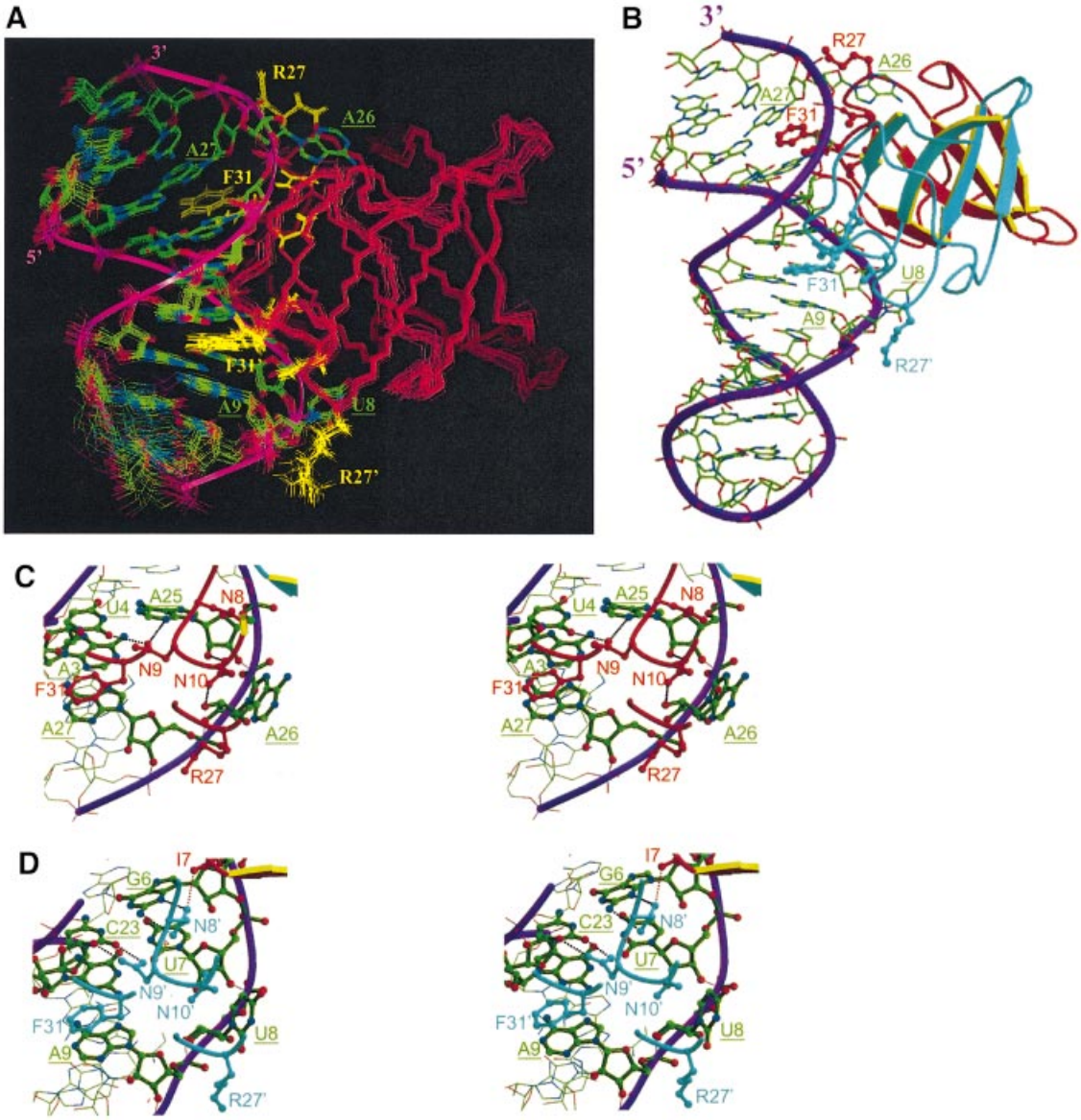

Fig. 3. (**A**) Ensemble of NMR structures of the LicT-CAT–RAT complex showing the protein backbone (in red) with some of the interacting amino acid side chains (in yellow), and the RNA helix (phosphodiester backbone in purple and nucleotides in standard atom colours). (**B**) MOLSCRIPT (Kraulis, 1991) representation of the LicT-CAT dimer interacting with its RAT hairpin target. The two CAT monomers, each composed of a four-stranded antiparallel β-sheet, are coloured in red and blue. Some important side chains interacting with the RNA are shown in ball-and-stick representation. The RNA phosphodiester backbone is shown in purple and the nucleotides are in standard atom colours. (**C** and **D**) Stereo views showing the pseudo-symmetric recognition of the RNA asymmetric internal loop 1 and loop 2, respectively, by each CAT monomer. The nucleotides forming loop 1 (the A3–A27 sheared pair and the bulged-out A26) and loop 2 (the U7–A9–G22 triplet and the bulged-out U8) are shown in ball-and-sticks as well as the neighbouring canonical base pairs (U4–A25 in loop 1, G6–C23 in loop 2). Relevant hydrogen bonds between protein and RNA residues are indicated as dotted lines.

loops (Figures 3 and 4). As expected, the highly variable apical loop is not recognized by CAT. The amino acids involved in the interaction lie in strand β1 (Lys5, Val6 and Ile7), in the short β1–β2 turn (Asn8–10) and in the beginning of the long loop joining β3 to β4 (Gly26, Arg27 and Phe31). These residues are all located on one side of the dimer, opposite the C-terminal end of the CAT peptidic fragment. Both monomers are required for the interaction.

Each RNA strand interacts with amino acid residues 5, 6 or 7, and 9 in one monomer and residues 8, 10, 26, 27 and 31 in the other monomer (Figures 3 and 4). In the dimer, most of these residues are used in a symmetrical manner to recognize similar structural features of the RNA internal loops and of the adjacent nucleotides. This is the case, for instance, for the strictly conserved aromatic side chain of Phe31, which makes an identical stacking interaction with either A9 in loop 1 or A27, the equivalent residue in loop 2 (Figure 3). In both cases, the aromatic ring is co-planar with the Watson–Crick base pairs, U4–A25 and G6–C23, neighbouring the loops. Both of these base pairs make hydrogen bond interactions with one or the other Asn9 functional group (Figures 3 and 4). The bulged-out bases

Table I. Characterization of the 20 NMR structures of the LicT–RNA complex retained for structural analysis

Distance and dihedral angle constraints	Number	Maximum violation (No.)
NOE-derived upper distance limits	1867	0.21 Å (3 ± 1 ⩾0.1 Å)
RNA–RNA (nucleotides 1–11, 20–29)	262	
Protein (residues 1–51)	1485	
Intramonomer A	709	
Intramonomer B	707	
Intermonomer	69	
Protein–RNA	120	
Hydrogen bonds initial (final modelling)		
RNA (nucleotides 1–11, 20–29)	19 (21 final)	
Protein–RNA	0 (1 final)	
Dihedral	45	2.3° (1 ± 1 ⩾2°)
RNA	32	
Protein	13	
R.m.s.d. from ideal geometry		
Bonds	0.006 ± 0.001 Å	
Angles	0.8 ± 0.1°	
Impropers	0.4 ± 0.05°	
R.m.s.d. between pairs of structures		
RNA (nucleotides 1–11, 20–29)	0.6 Å	
Protein	1.1 Å	

of the internal loops (U8 or A26) are docked in symmetry-related cavities on each side of the dimer interface (Figure 5).

The symmetry is not complete, however: in loop 1, the 2′ hydroxyl groups of A25 and A26 form hydrogen bonds with, respectively, ND2 and OD1 of Asn10 from one protein unit (Figure 3). This interaction, by pinching the phosphodiester backbone of the RNA, might stabilize the bulged-out conformation of the purine A26 that, contrary to equivalent nucleotide U8, is stacked inside the helix in the free RNA. In contrast, the functional group of Asn10 in the other protein unit is not involved in any particular interaction with loop 2. The hydrogen bonds formed by the side chain of Asn8 of monomer B with both the N3 and amino atoms of G6 and with the backbone carbonyl of Ile7 of monomer A are a good illustration of the interdependence between dimer interface geometry and RNA recognition. Hydrogen bonding from the protein backbone is observed for the carboxyl oxygen of Val6 in one monomer and Ile7 in the other monomer, which are interacting with the 2′ hydroxyl groups of G6 and A24, respectively.

Finally, electrostatic and/or hydrogen bond interactions between the phosphodiester backbone of the RNA and the positively charged side chain of some amino acid residues might complete the recognition mode of the complex. Due to the lack of easily observable protons close to the phosphate oxygen, these bonds cannot be inferred confidently from the observed NOEs. Nevertheless, it can be assumed from the refined model that contacts between Lys5-NZ and the phosphate group of U8 or A26, and between the guanidino group of Arg27 and the phosphate group of C28 or C10, might stabilize the complex further.

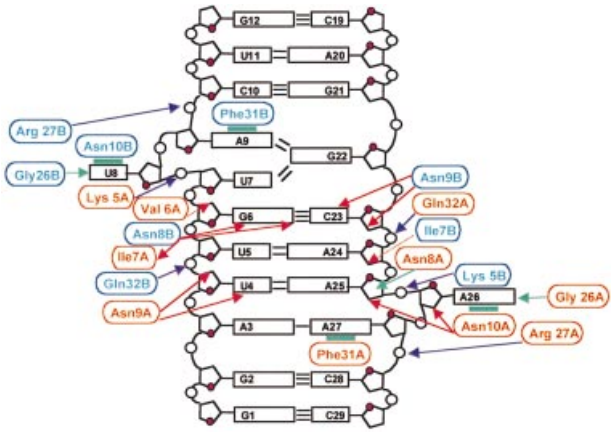

Fig. 4. Scheme of protein–RNA interactions. The colour codes for the amino acids (blue and red) correspond to that used for the monomers in Figures 2 and 3. Plain red lines indicate hydrogen bonds involving an amino acid side chain atom, broken red lines indicate hydrogen bonds involving a protein backbone atom, green lines indicate van der Waals interactions and purple lines indicate possible interactions (see text) with the RNA phosphodiester backbone.

Discussion

Implications for RNA recognition by the ATs of the LicT/SacY family

The RNA targets as well as the CAT domains of the transcriptional ATs from the LicT/SacY family share a high level of sequence similarity (Figure 1). Most nucleotides interacting with the protein are conserved among the already identified RAT sequences. Conversely, most amino acids directly involved in RNA recognition are strictly or quasi-strictly conserved (Figure 5C). These observations strongly suggest that a very high level of structural homology might exist between the different complexes formed by the RNA-binding domain of the proteins of the LicT/SacY family and their specific RAT targets.

The amino acid residues that are involved in base-specific contacts (Asn8, Asn9, Asn10, Gly26 and Phe31) are all highly conserved within the family. Those residues that interact with the nucleic acid phosphodiester backbone (Lys5, Val6, Ile7 and Arg27) are less conserved. An extensive mutational analysis performed on SacY has confirmed the crucial role of the conserved residues at positions 5, 8, 9, 10, 26 and 31 for efficient antitermination activity *in vivo* and RNA binding *in vitro* (N.Declerck, Y.Yang, M.Kochoyan and S.Aymerich, unpublished results). There are only a few conserved residues on the protein surface that are not directly involved in RNA recognition (Figure 5C). This is the case for the highly conserved Glu21 and Lys33, and for the less conserved Lys34, Glu45 and Lys46. Mutagenesis studies on SacY-CAT have shown that the charged side chain of these residues is indeed not essential for RNA binding. These residues might therefore be involved in other intra- or intermolecular interactions which are necessary for the antitermination function of the full-length protein under physiological conditions.

Based on sequence comparison and the present structural data, it can be concluded that all the RAT targets will fold into a hairpin structure with a regular double-stranded stem interrupted by two asymmetric internal loops. The

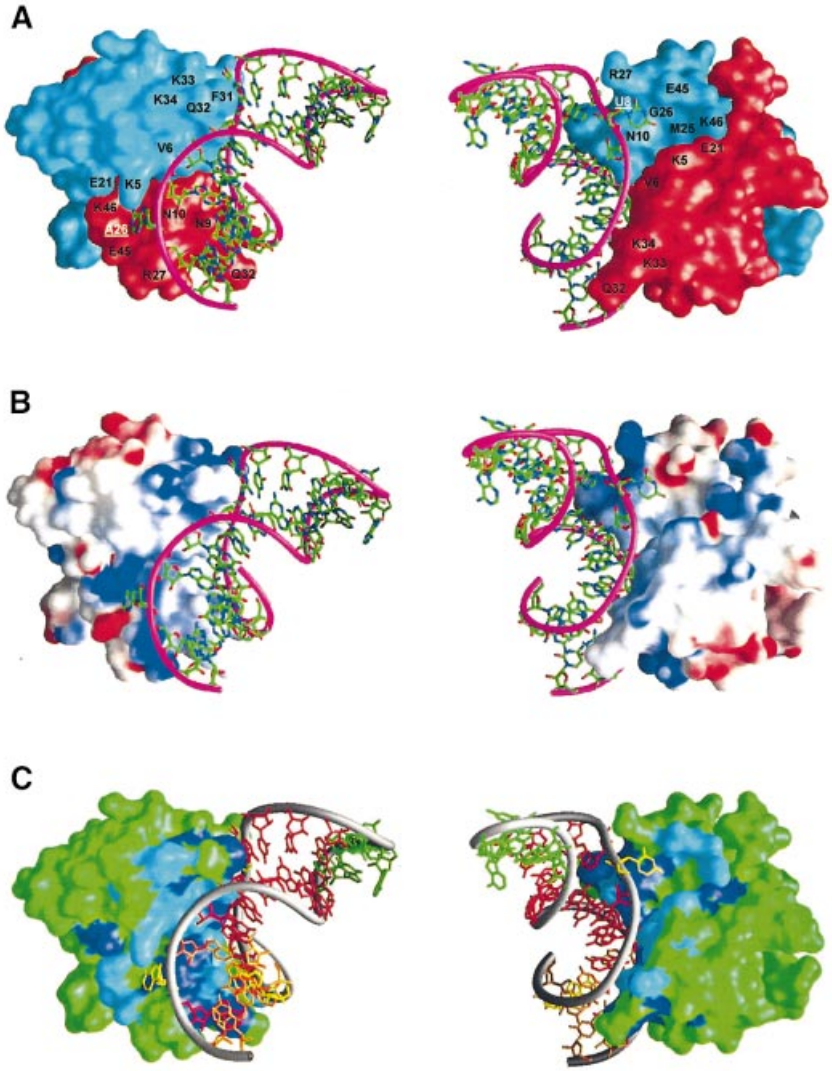

Fig. 5. GRASP (Nicholls *et al*., 1991) representations of the protein–RNA complex showing the symmetric role of the CAT monomers and the cavity on each side of the dimer receiving the bulged-out base in the RNA internal loop 1 (left side views) and loop 2 (right side views). In each case, the left and right side views showing the protein surface and the RNA backbone are rotated by ~180° with respect to each other. (**A**) The protein monomers are coloured in red and blue as in Figure 3. Amino acid residues are labelled in black. The bulged-out bases are labelled in white. (**B**) The electrostatic surface potential as calculated for the free CAT dimer using GRASP. The amino acids lying in the minor groove of the RNA helix are essentially neutral. They are surrounded by two spines of basic residues, interacting with the phosphodiester backbone. (**C**) Conserved amino acids and nucleotides coloured as a function of their level of conservation among the LicT/SacY family. Strictly conserved amino acids within the AT family are coloured in dark blue, conserved residues in blue and others in green. Similarly, the nucleotides are coloured in red, orange, yellow and green as their level of conservation within the RAT sequences decreases.

highly variable apical loop is not expected to be recognized by any CAT domain. The RNA stem region between the apical loop and asymmetric loop 2 contains highly conserved base pairs (C10–G21, U11–A20 and G12–C19) that are not involved in direct interaction with any amino acid residues but might contribute to the formation and stability of the RAT hairpin (Figure 1). Similarly, the basal portion of the RNA helix contains at least two conserved base pairs (G1–C29 and G2–C28) that are not interacting with CAT. The only base-specific interactions involving canonical base pairs are in the 3 bp stem joining the asymmetric internal loops (Figure 4). Nevertheless, except for G6, conservation of the pyrimidine and purine positions (independently of the nature of the base) should be sufficient to allow proper recognition by the ATs. The high conservation of these base pairs, as well as of the GC pairs in the basal stem, might be due to the fact that any mutation at these positions requires two compensatory mutations (one in the complementary strand of the RAT stem and another in the downstream sequence of the terminator) to preserve the functionality of the system. Sequence requirements within and at the 3′ end of the terminator sequence (in order to avoid the formation of alternative structures between the terminator and the RAT or to favour interactions of the terminator hairpin with the transcription machinery) might also impose the conservation of some nucleotides of RAT.

The two internal loops of the RNA hairpin constitute the major structural elements involved in CAT–RAT recognition. The architecture of these loops in the complex

(Figure 3) might be conserved throughout the entire family. In loop 2, the conserved pyrimidine at position 7 (U7 in the LicT RAT targets) allows formation of a Y7–G22–A9 base triplet, with a sheared AG pair and a wobble or Watson–Crick UG or CG pair (Leontis and Westhof, 1998), which is expected to be common to all RATs. The displaced nucleotide A9 acts as a stacking platform for the conserved aromatic side chain of Phe31 from one protein monomer. Loop 1 is characterized by an A3–A27 (or an isomorphous GA) sheared pair that allows nucleotide A27 to be displaced toward the minor groove, providing a stacking platform for Phe31 from the other monomer. One of the few exceptions to the formation of this sheared pair in homologous systems is in the *sacB*-RAT targeted by SacY in which A3 is replaced by a uridine. Formation of a U3–A27 Watson–Crick base pair is expected to be detrimental to the positioning of A27 towards the shallow groove and therefore to the stacking of Phe31 (Phe30 in SacY). This might partly explain why the SacY-CAT–*sacB*-RAT antitermination complex exhibits poor stability (Declerck *et al*., 1999). Another noticeable exception concerns the *ptsG*-RAT target of the GlcT AT. This RNA can be folded like the other members of the family with an apical loop, a canonical loop 2, containing a U–G–A base triple and a bulged-out U, but not with a canonical loop 1 (Figure 1). Strikingly, however, the basal portion of the stem can form a second U–G–A triplet with a bulged-out C, i.e. a structure identical to that of loop 2. *ptsG*-RAT could thus adopt a symmetrical structure at the protein recognition site. In contrast to what is observed for the other RATs, the stem between the two recognition loops would contain only two base pairs instead of three. Preliminary modelling indicates that a slight increase in the rise per base pair along the stem, a kink of the RNA helix or a slight rearrangement of the monomers within the dimer probably could compensate for the shortening of the RNA stem. The complex formed by GlcT-CAT–*ptsG*-RAT would then be completely symmetrical.

In both internal loops of the RNA, the base preceding the displaced adenine is looped out in the LicT-CAT–RAT complex. These bases might adopt the same extruded conformation in all the complexes. Since they are the main variable nucleotides of the two recognition regions, they are likely to be responsible for most of the recognition specificity of the ATs for their cognate RNA. Indeed, a genetic analysis of the RAT motifs has demonstrated that the nucleotides at positions 8 and 26 are specificity determinants of the RAT–AT interaction (Aymerich and Steinmetz, 1992; N.Declerck, unpublished results). However, since recognition of these bases by the LicT RNA-binding domain is provided mainly by the stacking interactions of two highly conserved amino acid residues (Asn10 and Gly26), no clear understanding of the structural elements involved in specificity has yet been reached. The cavity on each side of the CAT dimer (Figure 5), where the bulged-out bases are docked, may vary in shape and surface properties depending on the nature of the amino acid found in the surroundings, in particular at the less conserved positions 5, 27 and 45. On the RNA side, the nature of the bulged-out bases may influence the fine structure of the internal loops and/or of the phosphodiester backbone. The stacking and electrostatic interactions that maintain the complex (Figures 4 and 5) may thereby vary in strength and modulate the affinity of the proteins for their RNA target. Finally, the free energy cost of the in/out switch undergone by the nucleotide at position 26 could vary depending on the loop sequence and could also play a role in the specificity of recognition. Mutational and structural studies are currently under way in order to gain further insights into how specific recognition is achieved within the CAT–RAT family of antitermination complexes.

LicT is a protein clamp stabilizing an RNA hairpin

The LicT–RNA antitermination complex confirms the original mechanism of RNA recognition by the ATs of the BglG/SacY family. The CAT domain is not structurally related to any other known RNA-binding motif, and its interaction mode with RNA is radically different from that of other protein domains, including those interacting with RNA hairpins or double-stranded RNAs. The detailed RNA-binding features of a reasonable number of RNA-binding proteins are known and it appears that most of them recognize RNA as a monomer (Nagai and Mattaj, 1994; Cusack, 1999; Draper, 1999; Perez-Canadillas and Varani, 2001). The few oligomeric RNA- binding motifs characterized to date recognize single-stranded repeated sequences, each monomer interacting with a single repeat, as observed for the *trp* RNA-binding attenuation protein (TRAP) (Antson *et al*., 1999) or the transcription attenuation factor Rho (Bogden *et al*., 1999). The MS2 coat protein is the only exception. As for LicT, the RNA-binding domain is a homodimer and the amino acids that participate in RNA recognition are essentially the same for both monomers. However, contrary to what we observe here, there is little structural similarity in the manner in which both monomers recognize two distinct RNA sites (Valegard *et al*., 1994; Peabody and Chakerian, 1999).

Dimeric binding motifs are, by contrast, very common in DNA-binding proteins. Both monomers recognize identical or nearly identical sequences on the double-stranded DNA molecule that are either in the same orientation or in inverted orientation, such as, for example palindromic sequences. As expected, each monomer recognizes its DNA target in a very similar manner. In this sense, LicT provides the first example of a dimeric RNA-binding motif that recognizes a double-stranded RNA in a mode that is highly reminiscent of the recognition of palindromic DNA sequences by prokaryotic repressors or transcription factors. However, whereas DNA-binding proteins recognize Watson–Crick base pairs in double-stranded inverted repeats that are either symmetric or quasi-symmetric, LicT, like most RNA-binding proteins (Draper, 1999; Hermann and Patel, 2000; Westhof and Fritsch, 2000), recognizes the non-canonical structures formed by asymmetric internal loops. In RAT, the two internal loops are different in sequence but structurally related. The conformational similarities of these loops solve the problem of understanding how a symmetric dimer can recognize an asymmetric RNA hairpin. The loops constitute equivalent targets for the CAT dimer and can therefore interact in a similar way with the two monomers. As seen in the LicT-CAT–RAT complex structure, most of the interactions between the nucleotides of one loop and the amino acid residues of one monomer have their symmetrical counterparts in the other

loop and the other monomer. Even the non-conserved bases U8 and A26 are in fact recognized by the same conserved residues of each monomer (Figures 4 and 5).

For most RNA groove-binding proteins characterized to date (Conn *et al*., 1999; Stoldt *et al*., 1999; Wimberly *et al*., 1999; Agalarov *et al*., 2000; Batey *et al*., 2000; Nikulin *et al*., 2000; Ramos *et al*., 2000; Worbs *et al*., 2001), there are no major conformational rearrangements of the interacting partners upon complex formation. Recognition often accompanied by stabilization of a pre-existing RNA structure by an already folded protein domain is governed mostly by the structural complementarity of the free molecules. The LicT–RNA complex appears to follow this rule: the structure of the LicT dimer and most of the RNA hairpin remains practically unchanged in the bound form. A26 in loop 1, which is bulged out in the complex, is stacked inside the loop in the free RNA. A similar flip-out mechanism of adenine bases is observed when the IF1 protein binds to the 30S ribosomal subunit (Carter *et al*., 2001). However, as indicated by the broad line width of the exchangeable protons (Figure 2), most base pairs of the free RNA are highly unstable. A minor fraction of the population of the free RNA may thus adopt the conformation observed in the bound state, with both U8 and A26 already bulged out. In this situation, RNA–protein recognition may proceed through a conformational capture mechanism (Bouvet *et al*., 2001; Leulliot and Varani, 2001) in which a minor populated RAT conformer is selected by the CAT dimer.

Stabilization of the antitermination complex is achieved primarily through stacking and hydrophobic interactions. The shielding from the solvent of a total of four bases (U8, A9, A26 and A27), which are expelled from the core of the RNA helix in the bound state, and of two highly hydrophobic side chains (Phe31 of each monomer), which are exposed to the solvent in the free protein, might be the energetic factors contributing the most to the binding process. As shown in Figure 5D, most of the protein residues inserted within the minor groove and in contact with the bases are neutral. They are surrounded by two spines of basic residues probably involved in electrostatic interactions with the phosphodiester backbone and which may also contribute significantly to the stability of the complex. Due to the pseudo-symmetry of the complex formed between the CAT dimer and its RAT target, both monomers make an equivalent number of interactions with the two strands of the RNA stem. The protein dimer can therefore be viewed as an RNA clamp that prevents dissociation of the RNA stem and establishes a coupling between the RNA strand–strand interactions and the monomer–-monomer interactions in the CAT domain. Besides its possible relevance regarding the control of the AT activities within the entire protein, this architecture, by trapping at least the six last 3′ nucleotides of RAT, prevents the formation of the six basal base pairs of the terminator and thus allows the transcription to proceed further through the downstream coding sequences. A different way to stabilize a hairpin structure using a two-domain protein clamp has been described recently for the nucleolin–RNA complex (Allain *et al*., 2000).

Materials and methods

Production and purification of LicT-CAT

A DNA fragment encoding the (1–55) N-terminal peptidic fragment of LicT was inserted into the plasmid pGEX-2T (Pharmacia). Bacterial growth, protein purification and thrombin cleavage were performed as described previously (Manival *et al*., 1997). As in the case of SacY(1–55), a high salt concentration had to be used at all purification steps in order to maintain CAT dimeric structure (Manival *et al*., 1997). ^{15}N-labelled and doubly labelled (^{15}N, ^{13}C) samples were obtained by growing the bacteria in M9 minimal medium (Sambrook *et al*., 1989) with [^{15}N]ammonium chloride (Eurisotop) as the unique source of nitrogen or an Martek 9-CN medium for doubly labelled samples.

RNA synthesis

RNAs were obtained either by large-scale *in vitro* transcription (Milligan and Uhlenbeck, 1989) or by chemical synthesis on an Applied synthesizer, using Amersham PAC amidites, deprotected according to the manufacturer's protocol and then purified by ion exchange chromatography on a Q-HR column (Pharmacia). Doubly labelled RNA was obtained by *in vitro* transcription using NTP produced as described previously (Batey *et al*., 1992; Nikonowicz *et al*., 1992). The sequence of enzymatic RNA, GGA**GGAUUGUUACUGCUACGGCAGGCAAAACC**UC, includes the wild-type RAT-*bglP* sequence (bold nucleotides) found in the leader region of the *bglPH* operon controlled by LicT. The hairpin structure was stabilized by the insertion of an UNCG tetraloop (underlined residues) and by addition of two base pairs in the basal stem, to reduce end fraying effects in the vicinity of the protein-binding site. The sequence was chosen in order to avoid formation of alternative pairing and to optimize *in vitro* transcription yield. The synthetic RNA, C**GGAUUGUUACUGCUACGGCAGGCAAAACC**G, was designed to be as short as possible and contains a single base pair added to the basal stem. Prior to complex formation, the diluted RNA (0.1 mM) was heated at 80°C for 2 min, then quickly cooled in an ice–water bath to favour the formation of unimolecular folded-back stem–loop structures.

Complex formation and stabilization

Complex formation was monitored in the NMR spectrometer. When stoichiometry was reached, most of the protein's cross-peaks were split into two components of equal intensity, indicating the loss of symmetry of the monomeric units. The samples were usually stable for only a few days in the spectrometer. Two phenomena contribute to their poor stability: (i) the action of the nucleases which were never completely removed (even when additional steps of purification using ion exchange or heparin affinity chromatography, complicated by the necessity to maintain the protein at high salt concentration, were performed); and (ii) the progressive increase of all NMR resonance line widths (due to the RNA conversion from a single-stranded hairpin loop to a duplex structure, as suggested by the disappearance of the peaks characteristic of the UNCG tetraloop). These hypotheses were confirmed by the results of the gel filtration chromatography performed after the NMR experiments where large molecular weight species (duplex RNA probably complexed with two protein dimers) and low molecular weight species (degraded RNA) are observed in addition to the native stoichiometric complex. Transition of the RNA structure from a hairpin to a duplex is a consequence of the high salt and high oligonucleotide concentrations used in the experiments. Unfortunately, none of these parameters can be modified: the RNA concentration range is dictated by the NMR sensitivity; the salt concentration stabilizes the protein dimer, which otherwise dissociates and precipitates irreversibly in a few hours.

NMR

NMR samples contained 1–1.5 mM complex (protein dimer and RNA hairpin), in 250 mM NaCl, 10 mM Na phosphate buffer pH 6.8. Spectra were acquired on the 600 MHz Bruker AMX spectrometer of the Centre de Biochimie Structurale and on the 800 MHz spectrometers of the National Facilities in Grenoble (Varian Unity spectrometer) and Gif/Yvette (Bruker DMX spectrometer). Spectra were processed using the Gifa software (Pons *et al*., 1996). Complete assignments of the free protein and RNA were obtained using a set of two- and three-dimensional homo- and heteronuclear NMR experiments (Y.Yang and M.Kochoyan, unpublished results) performed on the unlabelled, ^{15}N- and ^{15}N/^{13}C-labelled molecules. Due to the high level of sequence homology between the LicT(1–56) fragment and the SacY(1–55) fragment previously studied (Manival *et al*., 1997), assignment of the free protein peaks was rather straightforward. Several RAT fragments were synthesized corres-

ponding to the RAT target of LicT (sequence above) or SacY (double mutant A3U, A26G, Figure 1), and the comparative analysis of the spectra obtained facilitates the assignments of the free RNA peaks (Y.Yang and M.Kochoyan, unpublished results). Assignments of all the protein resonances in the complex were obtained by analysis of a ^{15}N-NOESY-HSQC in 90% H_2O/10% D_2O and a 2D-^{13}C HCCH-TOCSY in D_2O performed, respectively, on a ^{15}N- and ^{15}N/^{13}C-labelled protein sample with unlabelled RNA. Assignments of the non-exchangeable RNA resonances (Pardi and Nikonowicz, 1992; Pardi, 1995) were obtained from ^{15}N- and ^{13}C-NOESY-HMQC and 2D-HCCH COSY in D_2O performed on a complex with doubly labelled RNA and unlabelled protein. Exchangeable protons of the RNA were assigned with a combination of ^{15}N-HMQC, CPMG-NOESY (Mueller *et al*., 1995) and 2D-NOESY (with JR solvent suppression; Plateau and Guéron, 1982) in 90% H_2O/10% D_2O. All the complexed RNA resonances, with the exception of about half of the H4′, H5′ and H5″ protons, were assigned.

Distance restraints were obtained from 80 ms mixing time 2D-NOESY experiments performed with unlabelled samples in D_2O and H_2O at 303 and 308 K. Distance restraints were classified as strong ($\leqslant$2.7 Å), medium ($\leqslant$3.3 Å), weak ($\leqslant$3.8 Å) and very weak ($\leqslant$4.5 Å). Most ambiguities were solved by comparison of the NOESY recorded at 303 and 308 K. A few non-resolved or ambiguous cross-peaks on these spectra were assigned from the heteronuclear 3D-NOESY-HSQC spectrum recorded with the ^{13}C-labelled RNA and the unlabelled protein. In this case, distance constraints were considered as very weak ($\leqslant$4.5 Å). Hydrogen bonds were constrained for canonical Watson–Crick base pairs exhibiting normal chemical shifts for the imino and amino proton resonances as well as for GU wobble pair of the AGU triplet, for which the G and U imino protons are strongly NOE connected (19 constraints). No hydrogen bonding constraints were imposed initially either between the G and A nucleotide of this base triplet or between A3 and A26 nucleotides and, of course, between the protein and the RNA. The UACG tetraloop and the stems that are far away from the protein-binding site (U11–A20) were modelled using fake constraints derived from the structure obtained by Varani and co-workers (Allain and Varani, 1995) or accepted for a standard A-form geometry (Saenger, 1984), and are not included in the list presented in Table I. The C3′-endo conformation of the ribose moiety was constrained in the absence of a H1′–H2′ TOCSY cross-peak (18 dihedrals). In the same way, the anti-conformation of nucleotides without strong H6/8–H1′ NOEs was also constrained (14 dihedrals).

Modelling

All modelling was performed using the X-PLOR 3.8 package (Brünger, 1992). In a first step, the free protein was modelled according to the protocol described (Manival *et al*., 1997; Nilges *et al*., 1997). Twenty-six ambiguous NOEs (for which distinction between inter- and intramonomers constraints was impossible) remained in the final stage of modelling. When the complex forms, 22 out of these 26 NOEs could be assigned unambiguously due to the spitting of the amino acid spin systems accompanying the loss of symmetry of the dimer. Only non-ambiguous constraints were then used for modelling the protein within the complex. Modelling of the complex consisted of a robust protocol of high temperature simulated annealing starting from one randomly chosen structure of the free protein generated in the previous stage of modelling and from a fully extended oligonucleotide chain. This protocol was adopted since the set of experimental constraints characterizing the structure of the free and complexed protein was nearly identical, except for the values of a few distance restraints (which were modified accordingly in the constraint list). The modelling protocol of the complex consisted of an 8 ps high temperature (2000 K) molecular dynamic (md) with reduced van der Waals radii, followed by a 16 ps md with a slow increase of the atom radii, followed by a 3.5 ps md with progressive cooling to 300 K. All these steps were performed in the absence of electrostatic, dihedral or hydrogen bonding potential and with a repulsive van der Waals potential. The structures were then submitted to 500 steps of conjugate gradient minimization with attractive van der Waals potential. Twenty or more structures were generated, and only those with a constraint energy within the limit of 30% of the lowest constraint energy obtained for the best structure (usually more than half and at least 10) were retained for structural analysis. New hydrogen bonds were added to the constraint list as short distance restraints (2.2 $\pm$ 0.3 Å) between a hydrogen and an acceptor atom after the following conditions were satisfied: (i) the acceptor and donor heavy atom were closer than 4 Å in all the analysed structures; (ii) the bonded protons were in slow exchange with the solvent (and in slow rotational exchange for the RNA amino groups); and (iii) the structure generated in the presence of the new constraint did not result in higher violation of the other experimental constraints than those generated in its absence. As a result, three hydrogen bonds were added to the constraints list. One, from the A27 amino protons to the A3-N3 nitrogen, defines the A3–A27 base pair (supported by the strong A27-amino to A3-H2 NOE). Another, from the G22 amino group to the A9-N7 nitrogen, defines the A9–G22 pair (supported by the strong G-amino to A-H8 NOE). The last hydrogen bond, from the carboxyl group of Asn10 to the 2′ hydroxyl of A26, is supported by a strong NOE from the NH_2 group of N10 to the slow exchanging H2′ sugar proton of A26.

Acknowledgements

We are grateful to C.Gaillardin at INA-PG for his constant support and interest in this work, and to E.Westhof for helpful discussion and advice. We thank P.Glaser at Institut Pasteur in Paris for giving information on sequences before release. J.P.Simore and E.Guittet are acknowledged for their help with the 800 spectrometers in Grenoble and Gif/Yvette. This work was supported by a PCV grant from the CNRS and by a EEC-TMR grant CT97-0154.

References

Agalarov,S.C., Sridhar Prasad,G., Funke,P.M., Stout,C.D. and Williamson,J.R. (2000) Structure of the S15,S6,S18–rRNA complex: assembly of the 30S ribosome central domain. *Science*, **288**, 107–113.

Allain,F.H. and Varani,G. (1995) Structure of the P1 helix from group I self-splicing introns. *J. Mol. Biol.*, **250**, 333–353.

Allain,F.H., Bouvet,P., Dieckmann,T. and Feigon,J. (2000) Molecular basis of sequence-specific recognition of pre-ribosomal RNA by nucleolin. *EMBO J.*, **19**, 6870–6881.

Antson,A.A., Dodson,E.J., Dodson,G., Greaves,R.B., Chen,X. and Gollnick,P. (1999) Structure of the trp RNA-binding attenuation protein, TRAP, bound to RNA. *Nature*, **401**, 235–242.

Arnaud,M., Debarbouille,M., Rapoport,G., Saier,M.H.,Jr and Reizer,J. (1996) *In vitro* reconstitution of transcriptional antitermination by the SacT and SacY proteins of *Bacillus subtilis*. *J. Biol. Chem.*, **271**, 18966–71892.

Aymerich,S. and Steinmetz,M. (1987) Cloning and preliminary characterization of the *sacS* locus from *Bacillus subtilis* which controls the regulation of the exoenzyme levansucrase. *Mol. Gen. Genet.*, **208**, 114–120.

Aymerich,S. and Steinmetz,M. (1992) Specificity determinants and structural features in the RNA target of the bacterial antiterminator proteins of the BglG/SacY family. *Proc. Natl Acad. Sci. USA*, **89**, 10410–10414.

Batey,R.T., Inada,M., Kujawinski,E., Puglisi,J.D. and Williamson,J.R. (1992) Preparation of isotopically labeled ribonucleotides for multidimensional NMR spectroscopy of RNA. *Nucleic Acids Res.*, **20**, 4515–4523.

Batey,R.T., Rambo,R.P., Lucast,L., Rha,B. and Doudna,J.A. (2000) Crystal structure of the ribonucleoprotein core of the signal recognition particle. *Science*, **287**, 1232–1239.

Bogden,C.E., Fass,D., Bergman,N., Nichols,M.D. and Berger,J.M. (1999) The structural basis for terminator recognition by the Rho transcription termination factor. *Mol. Cell*, **3**, 487–493.

Bouvet,P., Allain,F.H., Finger,L.D., Dieckmann,T. and Feigon,J. (2001) Recognition of pre-formed and flexible elements of an RNA stem–loop by nucleolin. *J. Mol. Biol.*, **309**, 763–775.

Brünger,A. (1992) *X-PLOR (Version 3.1): A System for X-ray Crystallography and NMR*. Yale University Press, New Haven, CT.

Carter,A.P., Clemons,W.M.,Jr, Brodersen.D.E., Morgan-Warren,R.J., Hartsch,T., Wimberly,B.T. and Ramakrishnan,V. (2001) Crystal structure of an initiation factor bound to the 30S ribosomal subunit. *Science*, **291**, 498–501.

Conn,G.L., Draper,D.E., Lattman,E.E. and Gittis,A.G. (1999) Crystal structure of a conserved ribosomal protein–RNA complex. *Science*, **284**, 1171–1174.

Cusack,S. (1999) RNA–protein complexes. *Curr. Opin. Struct. Biol.*, **9**, 66–73.

Declerck,N., Vincent,F., Hoh,F., Aymerich,S. and van Tilbeurgh,H. (1999) RNA recognition by transcriptional antiterminators of the BglG/SacY family: functional and structural comparison of the CAT domain from SacY and LicT. *J. Mol. Biol.*, **294**, 389–402.

Draper,D.E. (1999) Themes in RNA–protein recognition. *J. Mol. Biol.*, **293**, 255–270.

Gusarov,I. and Nudler,E. (1999) The mechanism of intrinsic transcription termination. *Mol. Cell*, **3**, 495–504.

Hermann,T. and Patel,D.J. (2000) RNA bulges as architectural and recognition motifs. *Structure Fold. Des.*, **8**, R47–R54.

Houman,F., Diaz-Torres,M.R. and Wright,A. (1990) Transcriptional antitermination in the *bgl* operon of *E.coli* is modulated by a specific RNA binding protein. *Cell*, **62**, 1153–1163.

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

Kruger,S. and Hecker,M. (1995) Regulation of the putative *bglPH* operon for aryl-β-glucoside utilization in *Bacillus subtilis*. *J. Bacteriol.*, **177**, 5590–5597.

Langbein,I., Bachem,S. and Stulke,J. (1999) Specific interaction of the RNA-binding domain of the *Bacillus subtilis* transcriptional antiterminator GlcT with its RNA target, RAT. *J. Mol. Biol.*, **293**, 795–805.

Leontis,N.B. and Westhof,E. (1998) Conserved geometrical base-pairing patterns in RNA. *Q. Rev. Biophys.*, **31**, 399–455.

Leulliot,N. and Varani,G. (2001) Current topics in RNA–protein recognition: control of specificity and biological function through induced fit and conformational capture. *Biochemistry*, **40**, 7947–7956.

Mahadevan,S. and Wright,A. (1987) A bacterial gene involved in transcription antitermination: regulation at a rho-independent terminator in the *bgl* operon of *E.coli*. *Cell*, **50**, 485–494.

Manival,X., Yang,Y., Strub,M.P., Kochoyan,M., Steinmetz,M. and Aymerich,S. (1997) From genetic to structural characterization of a new class of RNA-binding domain within the SacY/BglG family of antiterminator proteins. *EMBO J.*, **16**, 5019–5029.

Milligan,J.F. and Uhlenbeck,O.C. (1989) Synthesis of small RNAs using T7 RNA polymerase. *Methods Enzymol.*, **180**, 51–62.

Mueller,L., Legault,P. and Pardi,A. (1995) Improved RNA structure determination by detection of NOE contacts to exchange broadened amino-protons. *J. Am. Chem. Soc.*, **117**, 11043–11048.

Nagai,K. and Mattaj,I.W. (1994) *RNA–Protein Interactions*. Oxford University Press, Oxford, UK.

Nicholls,A., Sharp,K.A. and Honig,B. (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins*, **11**, 282–296.

Nikonowicz,E.P., Sirr,A., Legault,P., Jucker,F.M., Baer,L.M. and Pardi,A. (1992) Preparation of ^{13}C and ^{15}N labelled RNAs for heteronuclear multi-dimensional NMR studies. *Nucleic Acids Res.*, **20**, 4507–4513.

Nikulin,A. *et al*. (2000) Crystal structure of the S15–rRNA complex. *Nature Struct. Biol.*, **7**, 273–277.

Nilges,M., Macias,M.J., O'Donoghue,S.I. and Oschkinat,H. (1997) Automated NOESY interpretation with ambiguous distance restraints: the refined NMR solution structure of the pleckstrin homology domain from β-spectrin. *J. Mol. Biol.*, **269**, 408–422.

Pardi,A. (1995) Multidimensional heteronuclear NMR experiments for structure determination of isotopically labeled RNA. *Methods Enzymol.*, **261**, 350–380.

Pardi,A. and Nikonowicz,E.P. (1992) Simple procedure for resonance assignment of the sugar protons in ^{13}C-labeled RNAs. *J. Am. Chem. Soc.*, **114**, 9202–9203.

Peabody,D.S. and Chakerian,A. (1999) Asymmetric contributions to RNA binding by the Thr(45) residues of the MS2 coat protein dimer. *J. Biol. Chem.*, **274**, 25403–25410.

Perez-Canadillas,J.M. and Varani,G. (2001) Recent advances in RNA–protein recognition. *Curr. Opin. Struct. Biol.*, **11**, 53–58.

Plateau,P. and Guéron,M. (1982) Exchangeable proton NMR without base-line distortion, using new strong-pulse sequence. *J. Am. Chem. Soc.*, **104**, 7310–7311.

Pons,J.L., Malliavin,T. and Delsuc,M. (1996) Gifa V4: a complete package for NMR data-set processing. *J. Biomol. NMR*, **8**, 445–452.

Ramos,A., Grunert,S., Adams,J., Micklem,D.R., Proctor,M.R., Freund,S., Bycroft,M., St Johnston,D. and Varani,G. (2000) RNA recognition by a Staufen double-stranded RNA-binding domain. *EMBO J.*, **19**, 997–1009.

Saenger,W. (1984) *Principles of Nucleic Acid Structure*. Springer, New York, NY.

Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Schnetz,K., Stulke,J., Gertz,S., Kruger,S., Krieg,M., Hecker,M. and Rak,B. (1996) LicT, a *Bacillus subtilis* transcriptional antiterminator protein of the BglG family. *J. Bacteriol.*, **178**, 1971–1979.

Stoldt,M., Wohnert,J., Ohlenschlager,O., Gorlach,M. and Brown,L.R. (1999) The NMR structure of the 5S rRNA E-domain–protein L25 complex shows preformed and induced recognition. *EMBO J.*, **18**, 6508–6521.

Valegard,K., Murray,J.B., Stockley,P.G., Stonehouse,N.J. and Liljas,L. (1994) Crystal structure of an RNA bacteriophage coat protein–operator complex. *Nature*, **371**, 623–626.

van Tilbeurgh,H. and Declerck,N. (2001) Structural insights into the regulation of bacterial signalling proteins containing PRDs. *Curr. Opin. Struct. Biol.*, **11**, 685–693.

van Tilbeurgh,H., Manival,X., Aymerich,S., Lhoste,J.M., Dumas,C. and Kochoyan,M. (1997) Crystal structure of a new RNA-binding domain from the antiterminator protein SacY of *Bacillus subtilis*. *EMBO J.*, **16**, 5030–5036.

Westhof,E. and Fritsch,V. (2000) RNA folding: beyond Watson–Crick pairs. *Structure Fold. Des.*, **8**, R55–R65.

Wimberly,B.T., Guymon,R., McCutcheon,J.P., White,S.W. and Ramakrishnan,V. (1999) A detailed view of a ribosomal active site: the structure of the L11–RNA complex. *Cell*, **97**, 491–502.

Worbs,M., Bourenkov,G.P., Bartunik,H.D., Huber,R. and Wahl,M.C. (2001) An extended RNA binding surface through arrayed S1 and KH domains in transcription factor NusA. *Mol. Cell*, **7**, 1177–1189.

Yarnell,W.S. and Roberts,J.W. (1999) Mechanism of intrinsic transcription termination and antitermination. *Science*, **284**, 611–615.

Received December 28, 2001; revised February 18, 2002; accepted February 25, 2002