

# From genetic to structural characterization of a new class of RNA-binding domain within the SacY/BglG family of antiterminator proteins

Xavier Manival<sup>1,2</sup>, Yinshan Yang<sup>1</sup>,  
Marie Paule Strub<sup>1</sup>, Michel Kochoyan<sup>1,2,4</sup>,  
Michel Steinmetz<sup>3,†</sup> and  
Stéphane Aymerich<sup>3,4</sup>

<sup>1</sup>CBS, Faculté de Pharmacie, CNRS-UMR9955, INSERM-U414, Université Montpellier 1, 15 Avenue Charles Flahault, 34060 Montpellier, <sup>2</sup>Groupe de Biophysique, CNRS-URA1254, Ecole Polytechnique, F-91128 Palaiseau and <sup>3</sup>Génétique Moléculaire et Cellulaire, INRA and CNRS (ERS 05670), F-78850 Thiverval-Grignon, France

<sup>4</sup>Corresponding authors  
e-mail: michel@tome.cbs.univ-montp1.fr or  
stef@cardere.grignon.inra.fr

<sup>†</sup>This paper is dedicated to the memory of Michel Steinmetz, who died on February 7, 1995

**SacY is the prototype of a family of regulatory proteins able to prevent transcription termination. It interacts with a 29 nucleotide RNA sequence able to fold into a stem-loop structure and partially overlapping with a terminator sequence located in the 5' leader mRNA region of the gene it controls. We show here that the N-terminal fragment of SacY, SacY(1–55), and the corresponding fragments of other members of the family have antiterminator activities with efficiency and specificity identical to those of the full-length proteins. *In vitro*, this activity correlates with the specific affinity of SacY(1–55) for its RNA target. UV melting experiments demonstrate that SacY(1–55) binding stabilizes the RNA target structure. The NMR solution structure of SacY(1–55) is very similar to that obtained in the crystal (van Tilbeurgh *et al.*, 1997): the peptide is folded as a symmetrical dimer without any structural homology with other RNA-binding domains yet characterized. According to a preliminary NMR analysis of the SacY(1–55)–RNA complex, the protein dimer is not disrupted upon RNA binding and several residues implicated in RNA recognition are located at the edge of the dimer interface. This suggests a new mode of protein–RNA interaction.**

**Keywords:** antitermination/NMR/protein domain/protein structure/RNA binding

## Introduction

Many biological processes involve RNA–protein interactions. Examples include pre-mRNA splicing, mRNA translation, retrovirus replication and regulation of gene expression. Unlike DNA, which generally exists as a heterodimer consisting of two antiparallel molecules, RNA is often single stranded and can fold, usually by base pairing, into secondary and tertiary structures of complexity comparable with that of proteins (Mattaj, 1993).

Genetic and biochemical analyses have identified specific RNA-binding proteins and, in some cases, the protein and RNA sites which interact. Many of these proteins can be classified into several families characterized by common sequence motifs known, or thought, to interact directly with RNA (Mattaj, 1993; Burd and Dreyfuss, 1994). The structure of several RNA-binding proteins (Golden *et al.*, 1993; Schindelin *et al.*, 1993; Hoffman *et al.*, 1994; Matthews *et al.*, 1994; Willbold *et al.*, 1994; Antson *et al.*, 1995; Kang *et al.*, 1995) or RNA-binding protein domains (Nagai *et al.*, 1990; Hoffman *et al.*, 1991; Wittekind *et al.*, 1992; Bycroft *et al.*, 1997) have been determined, but the only available structural information on RNA–protein recognition is from the crystal structures of tRNAs complexed with their cognate synthetases (Rould *et al.*, 1989; Caverelli *et al.*, 1993; Biou *et al.*, 1994), of an RNA bacteriophage coat protein–operator complex (Valegard *et al.*, 1994), and from NMR or crystallographic analysis of the complexes between the homologous RNA-binding domains of two proteins, hnRNP C and snRNP U1A, with poly(rU) and U1 snRNA stem-loop II, respectively (Görlach *et al.*, 1992; Howe *et al.*, 1994; Oubridge *et al.*, 1994; Allain *et al.*, 1996). Some features revealed by these studies have been proposed to be of general significance for RNA–protein interactions. However, detailed structural information on more RNA–protein complexes will be needed to estimate the diversity of the interaction mechanisms.

SacY and BglG are homologous proteins that interact directly with RNA (Houman *et al.*, 1990; Aymerich and Steinmetz, 1992). However, their sequences do not contain any of the motifs that characterize the known families of RNA-binding proteins (Mattaj, 1993; Burd and Dreyfuss, 1994; Nagai, 1996). They are bacterial regulators able to prevent premature transcription termination, and are therefore called antiterminator proteins (AT). SacY mediates the induction of the *Bacillus subtilis sacB* gene by sucrose (Steinmetz, 1993), and BglG mediates the induction of the *Escherichia coli bgl* operon by  $\beta$ -glucosides (Amster-Choder and Wright, 1993). They bind a 29 or 30 nucleotide sequence of the *sacB* or *bgl* mRNA, respectively, called RAT for Ribonucleic AntiTerminator, which partially overlap the terminator sequences located upstream of the *sacB* and *bgl* coding sequences (Houman *et al.*, 1990; Aymerich and Steinmetz, 1992). The ability of its RAT target to fold into a stem-loop structure is essential for efficient interaction with SacY. These ATs are thus thought to prevent termination by stabilizing the RAT structure, which in turn excludes the formation of the terminator structure (Aymerich and Steinmetz, 1992). SacY and BglG are the best characterized members of a family of seven identified ATs; the RAT targets of these ATs differ only at a few positions, and the nucleotides present at some of these positions act as specificity

determinants in the interactions with the ATs (Aymerich and Steinmetz, 1992).

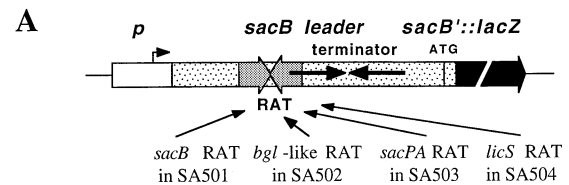
Looking for the RNA-binding domain of these ATs, we dissected the *sacY* and *bglG* genes, and identified the minimal SacY and BglG fragments sufficient for specific antitermination activity *in vivo*. Then, the relevant SacY fragment was purified and gel mobility shift experiments used to characterize its RNA-binding activity. NMR structural studies demonstrate that the peptide folds as a dimer, forming a structure not yet encountered in other nucleic acid binding domains. NMR titration of the protein–RNA complex allowed a preliminary characterization of the protein surface that contacts the RNA.

## Results

### Genetic identification of SacY and BglG minimal fragments sufficient for specific antitermination activity

To measure the specific antitermination activity of wild-type or variant proteins of the SacY/BglG family, we developed a genetic system based on isogenic *B.subtilis* strains, SA501–SA504, and the plasmid expression vector, pDG148. Strains SA501–SA504 are deleted for *sacY*, *sacT* and *licT* (which encode other ATs of the SacY/BglG family) and contain a chromosomal *lacZ* reporter gene under the control of either the wild-type *sacB* leader region encompassing the RAT target sequence of SacY, or a mutant *sacB* leader where the RAT specificity determinants have been replaced by those of the *bgl-*, *sacPA*- or *licS*-RAT, respectively (Figure 1A). The coding sequences for wild-type or mutant proteins were introduced into pDG148 in exactly the same position relative to the expression signals. Owing to their overexpression, on the one hand, and to the absence of the SacX, SacP, BglP and BglF kinases (*sacX*, *sacP* and *bglP* are deleted, *bglF* is not present in SA501–SA504) which are negative regulators of their activity, on the other hand, SacY, BglG, SacT and LicT are constitutively active in this test system.  $\beta$ -Galactosidase activity synthesized by the transformants therefore represents the antiterminator efficiency of the interaction between the protein encoded by the plasmid and the RAT target.

The results indicated that SacY interacts *in vivo* as efficiently with its cognate RNA target (*sacB*-RAT) as with the *bgl*-RAT, whereas BglG interacts efficiently only with *bgl*-RAT (Figure 1B), in accordance with our previous report (Aymerich and Steinmetz, 1992). To identify the regions of these ATs that determine their RNA-binding properties, we tested the activity of hybrid proteins containing various portions of SacY and the complementary portions of BglG. The hybrid gene consisting of the first 62 codons of *sacY* and the complementary portion of *bglG* (encoding the Y62G protein) led to exactly the same phenotype as the *sacY* wild-type gene, whereas the symmetrical hybrid gene (G65Y; residue 62 in SacY is aligned with residue 65 in BglG) led to the same phenotype as the *bglG* wild-type gene (Figure 1B). On the other hand, the G35Y chimeric protein appeared to interact with both *sacB*-RAT and *bgl*-RAT more efficiently than BglG, but less efficiently than SacY, whereas the symmetrical chimeric protein, Y32G, failed to interact with either *sacB*-RAT or *bgl*-RAT (Figure 1B). We then tested trunc-



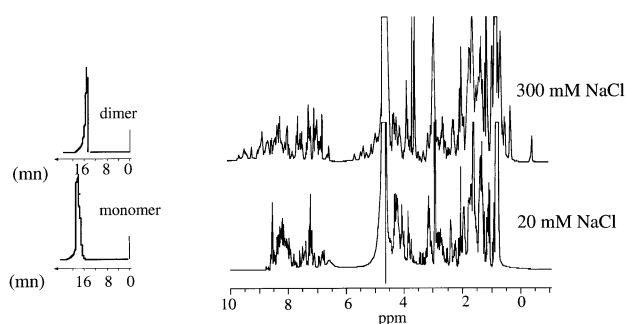
### B

| plasmid | protein    | $\beta$ -galactosidase activity |       |       |       |
|---------|------------|---------------------------------|-------|-------|-------|
|         |            | SA501                           | SA502 | SA503 | SA504 |
| pDG148  | none       | <1                              | <1    | <1    | <1    |
| pTY100  | SacY       | 94                              | 81    | 82    | 67    |
| pTG001  | BglG       | 3                               | 33    | 6     | 34    |
| pTG621  | G65Y       | 3                               | 31    | nd    | nd    |
| pTY162  | Y62G       | 121                             | 78    | nd    | nd    |
| pTG351  | G35Y       | 22                              | 66    | nd    | nd    |
| pTY132  | Y32G       | <1                              | <1    | nd    | nd    |
| pTY155  | SacY(1-55) | 102                             | 97    | 94    | 75    |
| pTY152  | SacY(1-52) | 73                              | 39    | nd    | nd    |
| pTY146  | SacY(1-46) | <1                              | <1    | nd    | nd    |
| pTG581  | BglG(1-58) | 2                               | 22    | 4     | 19    |
| pTG358  | BglG(3-58) | 3                               | 26    | 4     | 21    |
| pTS100  | SacT       | 24                              | 70    | 84    | 60    |
| pTS153  | SacT(1-53) | 33                              | 74    | 92    | 65    |
| pTL100  | LicT       | <1                              | 41    | 3     | 39    |
| pTL156  | LicT(1-56) | 1                               | 30    | 3     | 26    |

**Fig. 1.** (A) Structure of the reporter gene fusions harboured by the *B.subtilis* strains SA501–SA504. SA501 contains a tripartite fusion of a constitutive promoter (*p*), the promoter-free *sacB* leader region (dotted box) and the *lacZ* coding sequence (black box), integrated into the chromosome (thin line). SA502, SA503 and SA504 are isogenic with SA501, except for the RAT sequence (Figure 3D) which contains the *bgl*, *sacPA* and *licS* specificity determinants, respectively. (B) Antiterminator activity of entire, chimeric or truncated SacY, BglG, SacT or LicT proteins. The number that appears in the name of the chimeric proteins corresponds to the number of residues immediately preceding the junction point between the two parts of the chimera. Culture samples were assayed for  $\beta$ -galactosidase activity four times with samples from independent clones; activities are expressed in Miller's units (Sambrook *et al.*, 1989), standard deviations were all  $\leq 15\%$  of the mean for activities of 5 U or more, and  $\leq 30\%$  of the mean for lower activities. A representative result is given for each strain/plasmid combination. nd, not determined.

ated proteins and showed that the 55 amino acid N-terminal fragment of SacY [SacY(1–55)] and the 58 amino acid N-terminal fragment of BglG [BglG(1–58)] are as efficient for antitermination as the entire proteins, SacY and BglG, respectively. Moreover, they exhibit a specificity indistinguishable from that of the entire proteins (Figure 1B). Shorter SacY fragments [SacY(1–52) and SacY(1–46)] indicated that SacY(1–55) is the minimal fragment required to obtain an antitermination as efficient and specific as with the entire protein (Figure 1B).

The SacY(1–55) domain corresponds to 53–56 N-terminal amino acid regions in the other members of the SacY/BglG family: SacT, ArbG, LicT, AbgG and BglR from *B.subtilis*, *Erwinia chrysanthemi*, *B.subtilis*, *Clostridium longisporum* and *Lactococcus lactis*, respectively (Figure 7D). Because the first two amino acids of BglG have no counterpart in any of the other proteins, the SacY(1–55) domain corresponds to the region between residues 3 and 58 of BglG. By producing the appropriate protein fragments and testing them *in vivo* using reporter genes under the control of leader regions that contain either their cognate or a non-cognate RAT sequence, we



**Fig. 2.** Folded (top) and unfolded (bottom) NMR spectra of the SacY(1–55) peptide. A profile of gel filtration analysis, performed on an aliquot of the NMR sample, is shown on the left. The peptide is folded and dimeric at high salt concentration, unfolded and monomeric at low salt concentration. Once unfolded, the peptide can hardly be refolded, explaining why, even when transferred to the high-salt gel filtration buffer, the unfolded molecule remains monomeric.

confirmed that the 53 and 56 N-terminal amino acids of SacT [SacT(1–53)] and LicT [LicT(1–56)], and the segment between residues 3 and 58 of BglG [BglG(3–58)], constitute functional domains for specific antitermination (Figure 1B). We concluded that the 55 amino acid N-terminal region of SacY and the corresponding fragments of the other proteins of the family contain the RNA-binding domain including the specificity determinants for RNA target recognition.

#### **Characterization of SacY(1–55) and SacY in solution**

In order to characterize their RNA-binding properties, SacY and SacY(1–55) have been produced as fusion proteins with glutathione-*S*-transferase (GST) and purified on a glutathione affinity column. SacY and the SacY(1–55) peptide were then released by thrombin cleavage and further purified by ion-exchange chromatography. At low to intermediate salt concentration (i.e.  $\leq 100$  mM NaCl), the molecules tended to precipitate. This could be partly avoided by increasing the salt concentration to 0.4 M NaCl. However, even in these conditions, partial precipitation occurs after a few days.

Gel filtration analysis revealed that, in the micromolar range, both molecules predominantly exist in a dimeric state. A one-dimensional NMR spectrum was obtained for SacY(1–55) (Figure 2). The spreading of the amide proton resonances from 10 to 7 p.p.m. and of methyl and methylene aliphatic resonances from 5 to 0.5 p.p.m. in the spectrum of SacY(1–55) at 300 mM NaCl indicates that, at least, a fraction of the molecules is folded. In contrast, the spectrum obtained in 20 mM NaCl exhibits a much smaller chemical shift dispersion, characteristic of a random coil structure (Wüthrich, 1986). NMR titrations, performed at various protein and salt concentrations, showed that the ratio of folded to unfolded molecules correlates with the ratio of dimer to monomer concentrations as determined by gel filtration analysis performed on an aliquot of each NMR sample.

#### **SacY(1–55) specifically binds the *sacB*-RAT RNA sequence and stabilizes it**

The ability of SacY and SacY(1–55) to bind the *sacB*-RAT RNA sequence was investigated by gel mobility shift

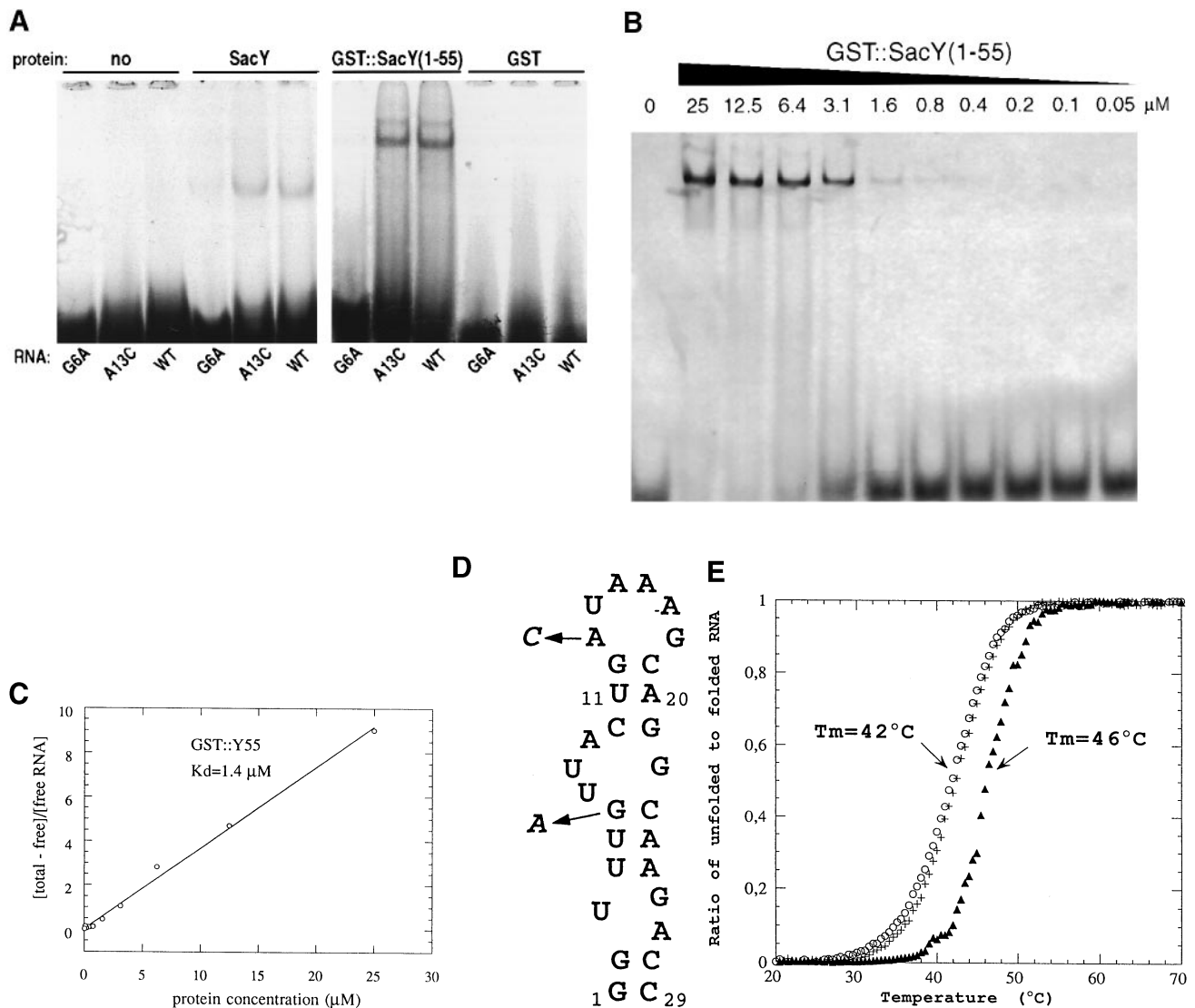
experiments. For every molecule, one major shifted band is observed when the protein is incubated with the wild-type RNA sequence or with the A13C RNA mutant (Figure 3A), two RNA sequences (Figure 3D) that have been shown to be equally efficient for antitermination by SacY (Aymerich and Steinmetz, 1992). A much weaker or no shifted band is observed for the G6A RNA mutant, inefficient for antitermination (Aymerich and Steinmetz, 1992). Under these test conditions, the affinity of SacY(1–55) for the wild-type or the A13C mutant RNA target appears to be slightly stronger than that of SacY.

In order better to characterize SacY(1–55) binding to RNA, the gel shift experiments were repeated with various amounts of GST::SacY(1–55) (Figure 3B). The intensity of the shifted band increased with increasing protein concentration. The banding pattern was quantified using a PhosphorImager and the plot of the ratio of the total-minus-free RNA to free RNA intensities as a function of the protein concentration (Figure 3C) was linear. Gel filtration analysis indicated that, for concentrations  $> 1$   $\mu$ M, i.e. those giving the most significant points of the plot, the protein was mostly dimeric before addition of the RNA. The linearity of the plot could therefore be ascribed to binding of one dimer to one RNA molecule. The slope of the line provided an estimate of the apparent dissociation constant of the complex: 1.4  $\mu$ M. Since this calculation assumes that 100% of the protein is dimeric and active, this value should be considered as an upper limit.

The observation of a concentration-independent UV melting transition ( $\sim 42^\circ\text{C}$  with 100 mM NaCl, 10 mM phosphate buffer pH 6.7) indicates that the RAT oligoribonucleotide is folded into an intramolecular structure (by opposition to a double helical, intermolecular dimer). Whether this structure corresponds to the model proposed (Figure 3D) on the basis of genetic analysis and sequence comparison (Aymerich and Steinmetz, 1992) remains to be demonstrated. However, UV melting experiments (Figure 3E) indicated that the melting temperature of the RAT motif is increased by  $\sim 4^\circ\text{C}$  in the presence of saturating amounts of SacY(1–55). This shift is unlikely to be due to unspecific electrostatic effects since equivalent amounts of spermidine, which has a positive charge comparable with that of SacY(1–55) (i.e. +3 to +4 as calculated from the amino acid composition of the peptide), did not modify the melting temperature. These results support the idea that the antitermination activity of SacY(1–55) results from its capacity to stabilize the RAT motif, which in turn prevents the formation of the terminator structure.

#### **Solution structure of SacY(1–55)**

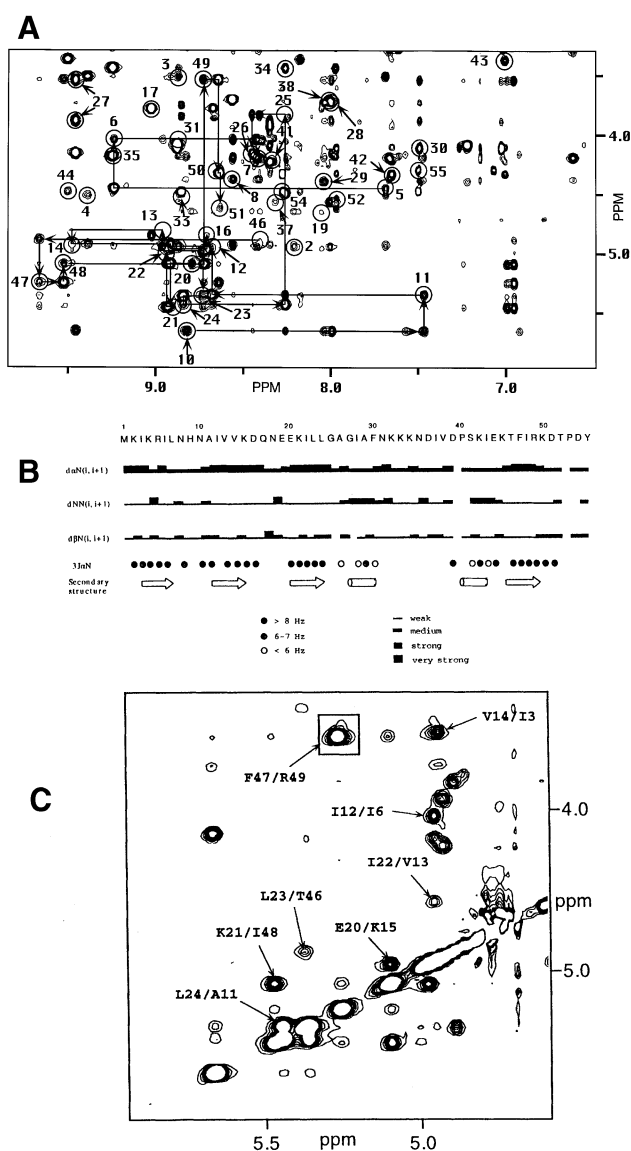
The structure of SacY(1–55) was investigated by NMR structural spectroscopy. As demonstrated above, the folded peptide is dimeric. Only 55 spin systems are observed, indicating that the dimer is symmetrical (Figure 4A). The dimer is only stable at high salt concentration and tends to precipitate irreversibly after a few days in the NMR tube, compromising the possibilities of forming heterolabelled N15, N14 dimers. In a first step, sequential assignment of the residues was obtained classically (Wüthrich, 1986). This analysis led us to a rather precise description of the secondary structure of the monomer: the sequential Nuclear Overhauser Effect (NOE) connectivities together



**Fig. 3.** (A) Gel mobility shift assays of SacY (5  $\mu$ M) and GST::SacY(1-55) (5  $\mu$ M) [and GST (10  $\mu$ M) as a control] with *sacB*-RAT wild-type or mutant RNAs. (B) Characteristics of specific complex formation between the GST::SacY(1-55) protein and SacB-RAT oligoribonucleotide. Gel mobility shift assays after incubation of the oligoribonucleotide with no protein, or various amounts of GST::SacY(1-55). (C) Plot of the ratio of total-minus-free RNA to free RNA as a function of the protein concentration for GST::SacY(1-55) [from the gel shown in (B)]. (D) Secondary structure model of the *sacB*-RAT sequence (Aymerich and Steinmetz, 1992). The G6A and A13C mutations are indicated. *bgl*-RAT, *sacPA*-RAT and *licS*-RAT present in strains SA502, SA503 and SA504 correspond to mutations 3A/13C/26A, 3A/8G/13G and 3A/26A, respectively. (E) UV melting of the RAT oligoribonucleotide (2  $\mu$ M) alone ( $\circ$ ), in the presence of 10  $\mu$ M of SacY(1-55) ( $\blacktriangle$ ) or 10  $\mu$ M spermidine (+). Experimental conditions: 300 mM NaCl, 0.5 mM EDTA, 20 mM phosphate buffer (pH 6.5).

with the values of the J3 $\alpha$ N coupling (Figure 4A and B) allowed a clear characterization of four  $\beta$ -strand regions lying from residues 3-8, 11-14, 19-24 and 45-48. Most of the strong H $\alpha$  to H $\alpha$  NOE connectivities (Figure 4C) observed within the strands are in agreement with a secondary structure consisting of a four-stranded anti-parallel  $\beta$ -sheet structure with sharp turns connecting strand 1 to strand 2 and strand 2 to strand 3, and with a long loop connecting strand 3 and strand 4. Based on this secondary structure, the cross-peaks observed between the H $\alpha$  protons of residues 47 and 49, together with those observed between the amide proton of residue 50 and the H $\alpha$  proton of residue 47, and between the amide proton of Ile48 and the H $\alpha$  proton of Arg49, could only be assigned as intermolecular. These connectivities point towards the formation of an intermolecular  $\beta$  sheet associating the C-terminal  $\beta$  strands of each monomer (Figure

4D). Starting from there, the NOE connectivities were sorted into three classes: those which are unambiguously intramolecular (related to the secondary structure elements characterized previously), the intermolecular NOEs and the ambiguous ones. A preliminary step of modelling was performed to obtain an initial three-dimensional structure of the monomeric unit using only the NOEs of the first class together with dihedral restraints obtained from coupling data (580 NOEs and 32 dihedrals). In this first model, the surface of the four-stranded  $\beta$  sheet appears to be covered on one side by the loop joining strand 3 to strand 4 (encompassing residues 25-44), whereas the C-terminal  $\beta$  strand, involved in dimerization, is pointing toward the other side of the surface. Based on this first model, NOE connectivities such as those observed between the side chain protons of Phe47, of Glu20 and Val13, or of Leu24 and Arg5, previously considered as ambiguous,



**Fig. 4.** (A) A 200 ms mixing time NOESY experiment showing the fingerprint region of the folded peptide. Sequential NH to H $\alpha$  connectivities are indicated. Strong NH to H $\alpha$  connectivities characteristic of  $\beta$  strands are indicated by arrows. They allow the identification of four  $\beta$ -strand regions from residues 3–8, 11–14, 19–24 and 45–48. (B) Summary of the sequential connectivities observed for the peptide. The secondary structure elements are defined according to the observed pattern of short-range connectivities. Helix-like turns between residues 26–30 and 40–43 are mentioned since the arrays of connectivities and coupling constants within these segments are in agreement with the formation of a nascent helix. The second segment appears to be present in both the crystal and the NMR structure. The first segment differs in both structures, but residue Asp31 is involved in a packing interaction within the crystal. (C) NOESY region showing the H $\alpha$ , H $\alpha$  connectivities characteristic of the  $\beta$ -sheet structures. The connectivity observed between residues Phe47 and Arg49 is clearly incompatible with the structure of the monomer and was classified as intermolecular. (D) Secondary structure of the dimer resulting from the NMR data reported above. Observed intramonomer NOESY are indicated by dashed lines, intermonomer ones (between Phe47 and Arg49) by plain lines.

could not be intramolecular and were reassigned as intermolecular. It was clear at this stage that dimerization occurred via a packing interaction between the exposed  $\beta$ -sheet surface of each monomer. Residues of the long loop, located on the other side of this surface, were then assumed to give only intramolecular NOEs which were added to the constraint list (600 intramolecular NOEs) and used to generate a new ensemble of monomeric conformers. Analysis of this ensemble again allowed the reassignment of ambiguous NOE cross-peaks incompatible with intramolecular connectivities as intermolecular.

A set of dimeric conformers was then generated by duplication of the monomeric conformers (both monomers completely overlapping). The protocol used to obtain the dimer structures was described by Nilges (1993) and is summarized in Table I. A non-crystallographic symmetry (NCS) potential was used to maintain the global similarity of the monomeric units. The symmetry of the dimers was imposed by a harmonic potential maintaining equal within  $\pm 0.2$  Å the 55 couples of intermolecular distances between each pair of equivalent atoms of the monomers. Only 600 unambiguous intramolecular and 15 intermolecular NOE-

derived distance restraints, and 32 dihedral restraints were used for preliminary calculations. In an iterative process, analysis of converged conformers (at least 10) allowed addition of a few extra unambiguous NOEs to the constraint list, whereas 23 NOEs remained ambiguous. Final calculations were performed with 673 intramolecular and 21 intermolecular unambiguous distance restraints, 23 ambiguous distance restraints and 32 dihedral restraints. Forty structures were computed, out of which 28, having a restraint energy violation within 30% of the conformer of lower restraint energy violation, were analysed. They are shown in Figure 5. The average pairwise r.m.s.d. between the conformers of the ensemble is 0.65 Å for the backbone atoms (residues 1–50) and 1.3 Å for all heavy atoms. There is no NOE violation over 0.5 Å and the average violation is 0.02 Å. Average deviations from ideality of bonds, angles and impropers are  $4 \times 10^{-5}$  Å,  $0.4^\circ$  and  $0.4^\circ$ , respectively. The average van der Waals energy of the structures is  $-160 \pm 30$  kcal. It should be stressed that this NMR structure has been obtained without the help of the crystal structure. The similarity of the results obtained (see van Tilbeurgh *et al.*, 1997) strongly supports the approach used in the present work.

Each monomer is folded into a four-stranded antiparallel  $\beta$  sheet. Tight  $\beta$  turns join strand 1 to strand 2 and strand

2 to strand 3, whereas a long loop, including two helix-like turns (residues 26–30 and 40–43), joins strand 3 and strand 4, and caps the twisted rectangular surface defined by the  $\beta$  sheet. The surfaces of the  $\beta$  sheets of each monomer are facing each other, in a roughly perpendicular orientation (Figure 7A). The dimer is maintained by hydrophobic interactions between residues located at the interface, by the formation of an intermolecular antiparallel  $\beta$  sheet between the C-terminal part of the fourth  $\beta$  strand (Phe47–Arg49) of each unit, and by two symmetrical hydrogen bonds between the side chain of Asn8 and the backbone carbonyl of Leu7. The loss of antitermination activity of SacY(1–46) (Figure 1B) and of a SacY(1–55) mutant protein in which Asn8 or Phe47 is replaced by an alanine (S.Aymerich, unpublished data) confirmed the crucial role of these two residues.

**Table I.** Simulated annealing protocol used for modelling the dimer

| Stage                       | Conformational search | Cooling | Minimization |
|-----------------------------|-----------------------|---------|--------------|
| Temperature (K)             | 800                   | 800–250 |              |
| Masses (a.m.u.)             | 100                   | 100     |              |
| Energy constants (kcal/mol) |                       |         |              |
| $K_{\text{bond}}$           | 1000                  | 1000    | 1000         |
| $K_{\text{angles}}$         | 125                   | 125–500 | 500          |
| $K_{\text{planar}}$         | 50                    | 50–500  | 500          |
| $K_{\text{repel}}$          | 0.003                 | 0.003–4 | 4            |
| $K_{\text{NOE safe}}$       | 50                    | 50      | 50           |
| $K_{\text{NOE ambiguous}}$  | 2                     | 2–25    | 25           |
| Asymptote ambiguous         | 2                     | 2       | 2            |
| $K_{\text{NCS}}$            | 0.01                  | 0.01–10 | 10           |
| $K_{\text{Symm}}$           | 1                     | 1       | 1            |
| Simulation time             | 10 ps                 | 20 ps   | 250 steps    |

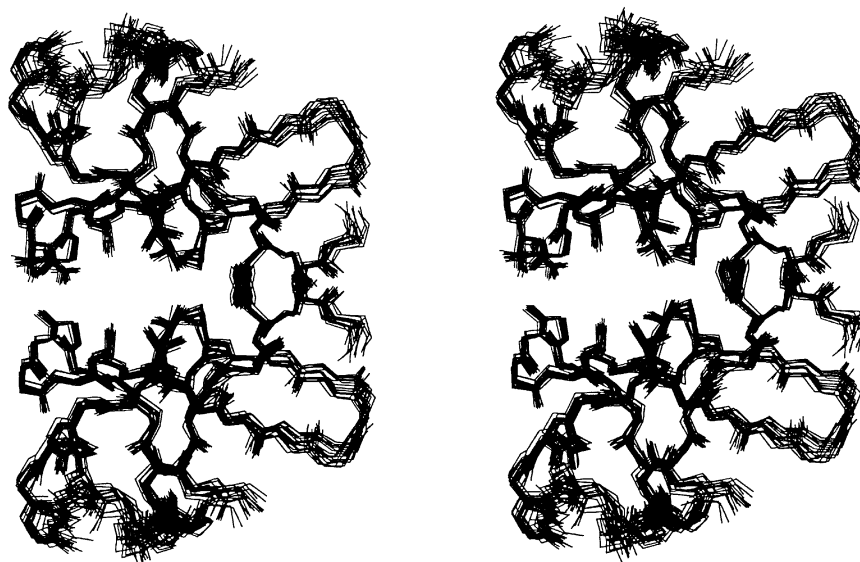
The starting structures were overlapping monomers obtained from a preliminary DG-SA step (see text).

### Preliminary characterization of the SacY(1–55)–RNA complex

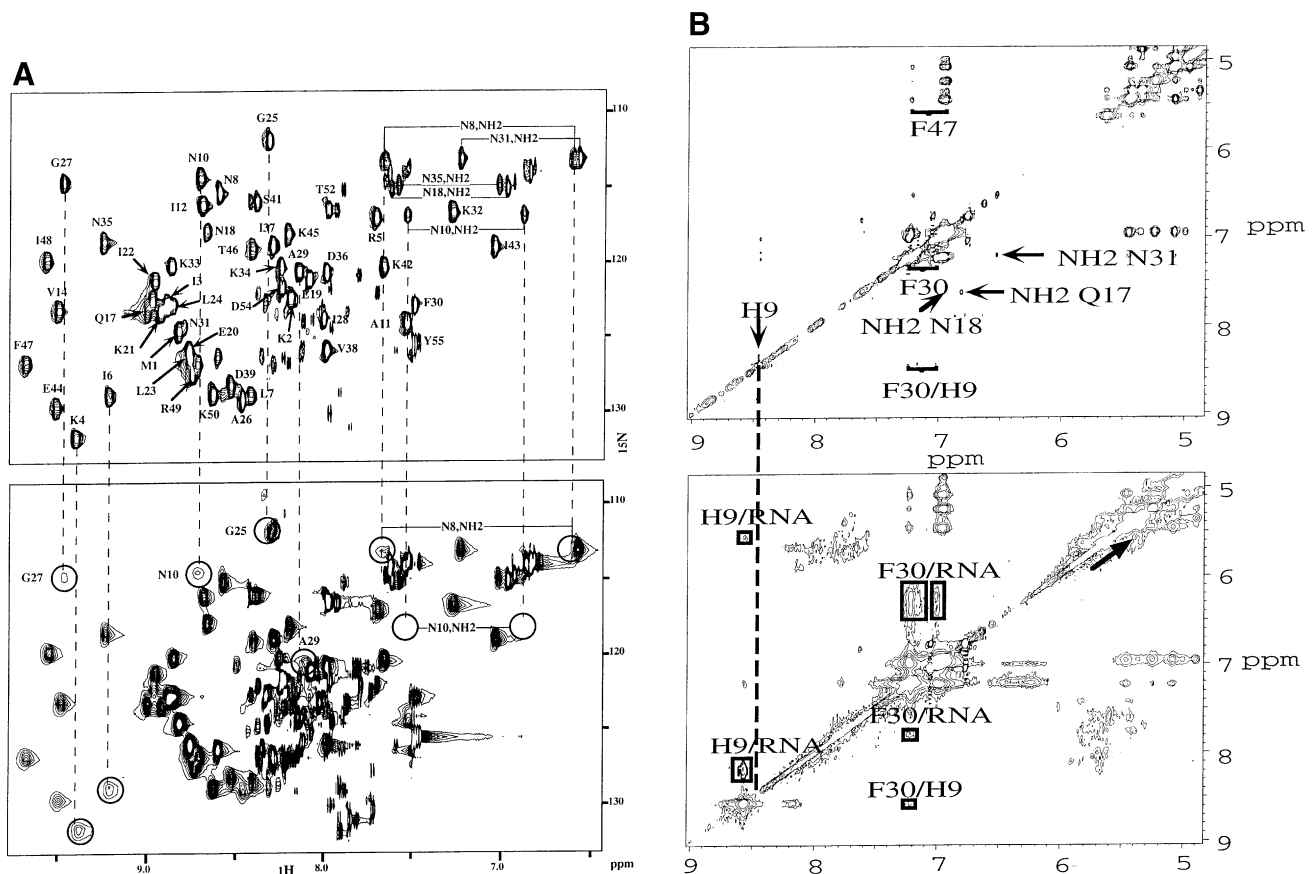
Upon addition of SacB–RAT oligoribonucleotide, most of the NMR peaks arising from the protein remain unchanged and intramolecular NOE cross-peaks are conserved, indicating that the global architecture of the dimer is maintained; the RAT motif thus binds *in vitro* specifically to the dimeric form of SacY(1–55) and does not disrupt it upon binding. However, the backbone (Figure 6A) and/or side chain resonances (data not shown) of a few residues broaden and/or are shifted, indicating the formation of a complex in fast to intermediate exchange at the NMR time scale (lifetime in the millisecond range in our experimental conditions). NOEs are observed between the aromatic protons of Phe30 and His9 and the aromatic and H1' protons of the RNA (Figure 6B). The residues affected by the presence of the nucleic acid, namely Lys4–Ile6, Asn8–Asn10, Gly25, Gly27, Phe30 and Asn31, are clustered at the surface of one side of the three-dimensional structure of the free protein structure (Figure 7). Involvement of some of these residues (Arg5, His9, Gly25, Gly27, Phe30) in the interaction with the RNA was confirmed by the weak or null antitermination activity of the corresponding single mutant molecules (S.Aymerich, unpublished data).

### Discussion

To identify the RNA-binding domain of the SacY and BglG proteins, we constructed various chimeric or truncated variants of the corresponding genes that allowed us to test *in vivo* various fragments of these proteins for antitermination activity. The SacY fragment consisting of the 55 N-terminal amino acids, SacY(1–55), and the homologous fragment of BglG, BglG(3–58), were thus shown to be functional for antitermination. Moreover, they were as efficient as the entire SacY and BglG proteins when tested



**Fig. 5.** An ensemble of 20 conformers obtained from the NMR-derived restraints. Side chains of the well-ordered residues of the hydrophobic core of the dimer (Asn8, Ala11, Ile22, Leu24, Val13 and Phe47) are shown.



**Fig. 6.** Preliminary NMR characterization of the complex. Experimental conditions: protein concentration  $\sim 2$  mM, 500 mM NaCl, 10 mM MgCl<sub>2</sub>, temperature 37°C in (A) and 28°C in (B). These stringent conditions were used in order to avoid the formation of unspecific complexes observed at lower temperature and salt concentration and resulting in a general broadening of all resonances in the NMR spectra. (A) HMQC experiment of <sup>15</sup>N-labelled SacY(1–55) in the absence (top) and in the presence (bottom) of 0.5 M equivalent of RAT oligoribonucleotide. The resonances affected by the presence of the RNA are labelled. (B) 100 ms mixing time NOESY spectra in D<sub>2</sub>O of the aromatic region of unlabelled SacY(1–55), in the absence (top) and in the presence of 0.5 M equivalent of RAT oligoribonucleotide (bottom). NOE cross-peaks are observed between the aromatic protons of F30 and H9 (shifted downfield), and the aromatic and H1' sugar protons of the RNA. The side chain cross-peaks (data not shown) modified by the presence of RNA are the same as those whose amide resonances are affected.

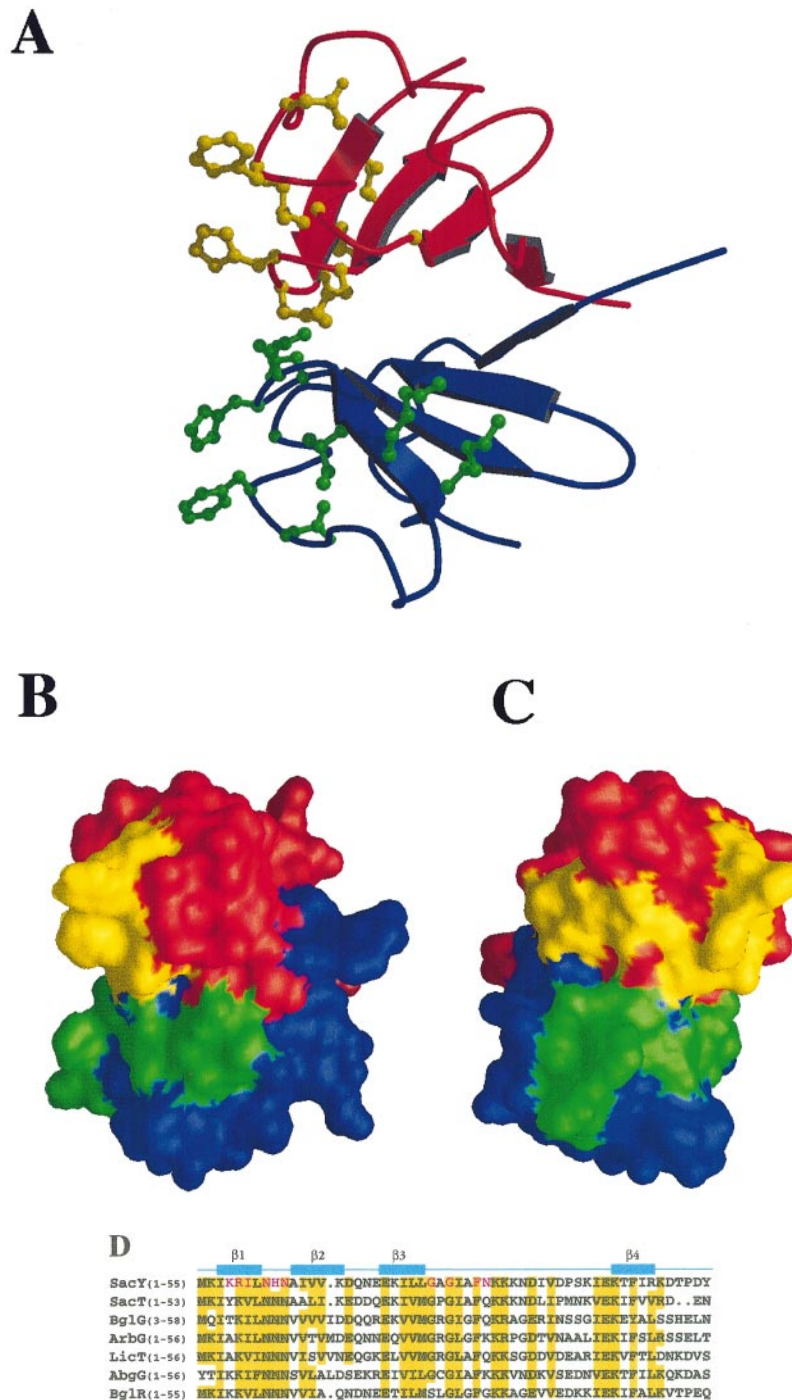
in our genetic system using an identical expression vector and the same reporter gene fusion. They exhibited the same recognition specificity of the RNA targets tested (cognate or non-cognate) as the corresponding full-length proteins. Similarly, we demonstrated that the relevant N-terminal fragments of SacT and LicT, two other proteins of the SacY/BglG family, also had antitermination activities with efficiencies and specificities identical to those of the entire proteins, thus confirming that the homologous N-terminal regions of the proteins of this family contain their RNA-binding domains.

Gel-shift assays demonstrated that SacY and its truncated form, SacY(1–55), have similar affinities for their RNA target, the RAT oligoribonucleotide, but do not recognize a mutant RAT sequence unable to promote antitermination *in vivo*. The RAT oligoribonucleotide was shown to adopt a folded back stem-loop structure that is rather unstable, as indicated by its low, concentration-independent, melting temperature ( $\sim 42^\circ\text{C}$ ) and the qualitative analysis of its NMR imino proton spectrum (M.Kochoyan, unpublished data). In the presence of SacY(1–55), the melting temperature of the motif is increased by  $\sim 4^\circ\text{C}$ . These results are in agreement with the following model for antitermination: SacY inhibits the termination of transcription by stabilizing the RAT

structure, which itself excludes the formation of the terminator, because of the six nucleotide overlap between the RAT and the terminator sequences.

The singularity of the RNA-binding domain of the ATs of the SacY/BglG family, predictable by the absence of primary sequence homology with other RNA-binding proteins, is confirmed at the structural level by the NMR studies. The SacY(1–55) peptide is folded as a symmetrical dimer. The four-stranded antiparallel  $\beta$  sheets of each monomer are facing each other in a roughly perpendicular orientation to form an eight-stranded  $\beta$  barrel covered on both sides by the loops joining strand 3 and strand 4 of each monomer. The sequence conservation of the N-terminal RNA-binding moiety within the SacY/BglG family (Figure 7D), especially regarding the residues involved in the  $\beta$  sheet and in the hydrophobic core, is a strong indication that this fold is preserved within the family (van Tilbeurgh *et al.*, 1997). We propose to call this new type of RNA-binding domain CAT, for ‘Co-AntiTerminator’.

An antiparallel  $\beta$  sheet is frequently found in RNA-binding domains (Schindelin *et al.*, 1993; Valegard *et al.*, 1994; Allain *et al.*, 1996; Nagai, 1996; Bycroft *et al.*, 1997). The residues splayed on the solvent-exposed face of the  $\beta$  sheet have been often shown (Valegard *et al.*,



**Fig. 7.** (A) MOLSCRIPT (Kraulis, 1991) representation of a converged NMR structure of low residual constraint energy. Residues affected by RNA binding (Figure 6) are shown in ball and stick representation in yellow or green on each monomer. (B) Surface representation of the dimer (same orientation and same colour code as in (A) using the GRASP program (Nicholls, 1992). (C) Same as in Figure 4B after a 90° rotation of the molecule. (D) The SacY(1-55) domain corresponds to 53-56 N-terminal amino acid regions in the other members of the SacY/BglG family: SacT (Débarbouillé *et al.*, 1990), ArbG (el Hassouni *et al.*, 1992), LicT (Schnetz *et al.*, 1996), AbgG (GeneBank) and BglR (Bardowski *et al.*, 1994) from *B.subtilis*, *Erwinia chrysanthemi*, *B.subtilis*, *Clostridium longisporum* and *Lactococcus lactis*, respectively. Because the first two amino acids of BglG have no counterpart in any of the other proteins, the SacY(1-55) domain corresponds to the region between residues 3 and 58 of BglG. Residues involved in the interaction with the RNA are in red.

1994; Allain *et al.*, 1996; Nagai 1996), or proposed, to participate in nucleic acid binding. A  $\beta$  sheet also constitutes the central structural element of the SacY(1-55) monomer. However, one face of the sheet is involved in dimerization, while the other is capped by the long loop joining strand 3 and strand 4. Therefore, the  $\beta$  sheet is

almost totally inaccessible from the solvent. In order to probe the SacY(1-55) protein for RNA contacts, we titrated a  $^{15}\text{N}$ -labelled protein with a RAT oligoribonucleotide. This preliminary footprinting of the RNA recognition surface revealed: (i) that the global architecture of the dimer is maintained upon RNA binding; (ii) that the amino



acids in contact with the nucleic acid are located on the rim of the sheet, at the edge of the dimer interface (Figure 7), and along the long loop, designing a roughly L-shaped surface (Figure 7C) on each monomer. This interaction surface clearly distinguishes SacY(1–55) from other known RNA-binding domains. This difference might be related to the differences in their RNA targets. U1A snRNP protein and MS-2 virus coat protein recognize clustered nucleotides, within and next to one flexible loop (Valegard *et al.*, 1994; Allain *et al.*, 1996). The bulged nucleotides, which are present in the RAT target at positions 3, 8, 13 and 26, and which determine the specificity of the interaction with the ATs (Aymerich and Steinmetz, 1992), are scattered all along the stem of the RAT motif (Figure 3D). They span at least 5–6 base pairs, according to the putative secondary structure of the RNA, i.e. more than half of a helical turn over a distance probably >15 Å. The protein–RNA contact region should, therefore, be rather extended, which is in agreement with the footprint observed on the protein. At present, we cannot distinguish whether the RNA contacts a single monomeric unit (the yellow or the green surface in Figure 7C) or whether it binds to both units (the yellow and the green surface). In all cases, RNA binding disrupts the symmetry of the dimer. Binding of a second RNA could restore this symmetry, but would probably lead to severe steric and electrostatic clashes between the bound nucleic acid molecules. Moreover, it is difficult to imagine that a single SacY molecule could bind at the same time two *sacB* mRNAs emerging from two distinct transcription complexes that necessarily progress one after the other along the 5' leader region of the single *sacB* gene present in the cell. Determination of the structure of the SacY(1–55)–RAT complex will show whether SacY(1–55) could bind one or two RAT molecules *in vitro*.

SacY(1–55) is present as a dimer in solution and in various crystal forms (van Tilbeurgh *et al.*, 1997) and this dimeric form is not disrupted upon RAT binding; this suggests that, *in vivo*, the full-length active form of the SacY protein is also a dimer. This is in agreement with the results of Amster-Choder and Wright (1992) showing that BglG is a dimer in its active form. The proteins of the SacY/BglG family would thus dimerize via their RNA-binding domain. The regulation of the antitermination activity of SacY in response to the external level of the inducer involves Enzyme I and HPr, the general energy coupling proteins of the phosphoenolpyruvate:sugar-phosphotransferase system (PTS), and SacX, a membrane protein homologous to SacP, the *B. subtilis* sucrose-specific PTS-permease, acting as negative regulators. Recently, it was demonstrated that SacY is phosphorylatable by HPr-(His-P) on three histidyl residues present in both of the two consecutive homologous segments of ~100 amino acids that follow the RNA-binding domain in the SacY molecule. These two segments together thus constitute the sensor domain. One of these phosphorylatable histidyl residues has been proven to be directly involved in the regulation of the antitermination activity of SacY (Tortosa *et al.*, 1997). Inhibition of SacY implies the transmission of a signal from the sensor module to the effector module, i.e. the RNA-binding domain. In the absence of sucrose, SacY would interact with phosphorylated SacX and would be phosphorylated by HPr-(His-P). SacY could be main-

tained in this inactive phosphorylated form because it would remain trapped by SacX (stoichiometric mechanism) or because the phosphorylation would induce a conformational switch within the protein masking the RNA-binding surface (catalytic mechanism). Within the latter framework and following Amster-Choder and Wright (1992), who proposed that the inactive form of BglG is phosphorylated and monomeric, phosphorylation could trigger a rearrangement of the N-terminal moiety of the protein. This domain would thus become no longer available for dimerization and RNA binding. *In vitro*, the N-terminal fragment of SacY appears folded only as a dimer and unfolded when monomeric, probably due to solvent exposure of the hydrophobic residues of the dimerization surface. Stabilization of this domain within a monomeric SacY entire protein would then require partial protection of these residues, possibly via an interaction with the sensor domain. The switch between this intramolecular interaction and the intermolecular dimerization, leading to a functional RNA-binding domain, could be triggered by SacY dephosphorylation.

To conclude, we showed that the 55 amino acid N-terminal fragment of SacY constitutes the RNA-binding domain of this protein. This domain is folded as a symmetrical dimer in solution. Its structure has been determined both by NMR spectroscopy and by X-ray crystallography (van Tilbeurgh *et al.*, 1997). NMR titration of complex formation between SacY(1–55) and its RAT target indicates that the amino acids involved in RNA recognition are located close to the dimer interface. These results, together with that of Aymerich and Steinmetz (1992), imply that the stem of the RAT element has to be located in the proximity of this interface, either parallel (in the case of binding to a single unit) or perpendicular (in the case of binding across the dimer) to it. In both cases, the dimer can be viewed as a molecular clamp, stabilizing the RAT structure in order to prevent the formation of the terminator structure. To the extent of our knowledge, a similar RNA-binding mode has not been described before. Characterization of the SacY(1–55)–RAT complex, by NMR and crystallography, is now in progress in our laboratory.

## Materials and methods

### Culture media and genetic procedures

Luria broth (LB) was used for bacterial growth except where otherwise indicated. CHG medium is minimal C medium (Aymerich *et al.*, 1986) containing 0.25% (w/v) casein hydrolysate and 1% (w/v) glucose. *Escherichia coli* strains were transformed by the calcium-shock procedure (Sambrook *et al.*, 1989). *Bacillus subtilis* strains were transformed by using their natural competence (Anagnostopoulos and Spizizen, 1961).

### *Bacillus subtilis* strains

Strain SA500 is a *licTΔ* derivative of GM905 (Aymerich and Steinmetz, 1992) obtained by replacing the chromosomal fragment delimited by the *Hind*III site 2 kb upstream from *bglP* and the *Sal*I site of *licS* (thus including the *licT* gene) with the kanamycin-resistance cassette, *aphA3*. SA501, SA502, SA503 and SA504 were obtained by transformation of SA500 with the integrative plasmids pIC38, pIC38/3A/13C/26A/t, pIC38/3A/8G/13G and pIC38/3A/26A/t (Aymerich and Steinmetz, 1992), respectively.

### Construction of the pTY, pTG, pTS and pTL series of plasmids

Plasmids pTY, pTG, pTS and pTL are derivatives of the replicative expression vector pDG148 (Stragier *et al.*, 1988), and contain wild-type,

chimeric or truncated *sacY*, *bglG*, *sacT* or *licT* genes under the control of the inducible *spac* promoter. These genes were obtained by PCR using standard procedures (oligonucleotide sequences are available on request) and the method described by Yon and Fried (1989). These constructions were checked by DNA sequencing.

#### **Bacillus subtilis liquid cultures and $\beta$ -galactosidase assays**

The *B.subtilis* strains transformed by one pTY, pTG, pTS or pTL plasmid were grown in CHG medium in the presence of phleomycin (0.2 mg/l); expression of the plasmid gene under the control of the *spac* promoter was induced by addition of 0.5 mM IPTG. Culture samples (Aymerich and Steinmetz, 1992) were assayed for  $\beta$ -galactosidase activity by the method of Miller (Sambrook *et al.*, 1989).

#### **Production and purification of SacY(1–55) and SacY proteins**

The coding sequences corresponding to the SacY(1–55) and SacY proteins were amplified by PCR using primers that generate a *Bam*HI site immediately upstream of the initiation codon and an *Eco*RI site downstream of the stop codon, and inserted into the plasmid pGEX-2T (Pharmacia). The resulting plasmids, pGEX-SacY(1–55) and pGEX-SacY, were checked by DNA sequencing and then used to transform *E.coli* strain BL21(DE3) (Sambrook *et al.*, 1989). Bacterial growth, protein purification and thrombin cleavage were performed according to the manufacturer's instructions (Pharmacia). SacY and SacY(1–55) cleaved fragments were further purified on a Q Sepharose fast flow or an S Sepharose fast flow column (Pharmacia), respectively. For NMR experiments, the SacY(1–55) fragment was further purified by gel filtration on a Superdex 100 HR column (Pharmacia) operated with the NMR buffer (300 mM NaCl, 10 mM phosphate buffer pH 5.5). This allowed the separation of dimeric molecules (usually >90% of the total) from monomeric ones. The purity of the proteins, checked by SDS-PAGE and Coomassie staining, was >95%. Comparison of staining intensity with three molecular weight markers provided an estimate of the protein concentration. The peptides were then concentrated using a Filtron low-molecular-weight cut-off system. SacY(1–55) purity and integrity were checked by mass spectrometry; concentration for the NMR experiments was estimated from the measurement of peak intensities performed on the isolated amide peak around 10 p.p.m. and the upfield shifted aliphatic peak at –0.5 p.p.m. (both characteristic of the folded molecule) after calibration with a reference sample. <sup>15</sup>N-labelled SacY(1–55) was obtained by growing the bacteria in M9 minimal medium (Sambrook *et al.*, 1989) with [<sup>15</sup>N]ammonium chloride (Eurisotop) as the unique source of nitrogen.

#### **RNA synthesis and labelling**

RNAs were obtained either by large-scale *in vitro* transcription (Milligan and Uhlenbeck, 1989) or by chemical synthesis, and then purified by ion-exchange chromatography on a Q-HR column (Pharmacia). The wild-type RAT oligoribonucleotide used in Figure 3B was further purified by preparative electrophoresis on a 12% denaturing polyacrylamide gel. RNA was labelled with <sup>33</sup>P using T4 polynucleotide kinase, after dephosphorylation for RNA synthesized enzymatically.

#### **Gel filtration**

Gel filtration was performed using an HPLC (Waters) system and a Protein Pack 60 (Waters) column calibrated with RNase A (13 700 Da), cytochrome C (12 500 Da), aprotinin (6540 Da) and Toxine  $\gamma$  (6800 Da), and operated at constant flow rate (0.5 ml/min) using a 20 mM phosphate buffer (pH 6.5) adjusted to 200 mM NaCl. Protein was detected by a variable-wavelength UV detector (Waters) at 220 nm.

#### **Gel mobility shift assays**

Twenty femtomoles of <sup>33</sup>P-labelled RNA were mixed with 20 pmol of non-specific competitor RNA (unrelated 55 residue stem-loop motif from a previous transcription), 10 U of RNase inhibitor (Pharmacia) and the relevant amount of protein. The solution was adjusted to 300 mM NaCl, 0.5 $\times$  TBE in a 5  $\mu$ l final volume. After equilibration for 30 min at 20°C, the loading dyes and glycerol (5% final) were added, and the sample was loaded on a 6 or 8% polyacrylamide (19/1) gel under 200 V at 5°C. Electrophoresis was continued for ~1 h at 5°C.

#### **UV melting**

UV experiments were performed on a Cary 1E spectrophotometer equipped with a temperature control unit. For melting experiments, the data were collected every minute, the temperature being increased by steps of 0.2°C. The ratio of unfolded to folded RNA,  $\alpha$ , was obtained by the equation  $A = \alpha A_{uf} + (1-\alpha)A_f$ , where  $A$  is the measured absorbance

and  $A_f$  and  $A_{uf}$  are the absorbances, at the same temperature, of the unfolded and folded species, respectively. We extrapolated these values from the low (between 20°C and 25°C) and high (between 60°C and 70°C) temperature regions assuming a linear dependence of absorbance versus temperature (Rougée *et al.*, 1992).

#### **NMR and modelling**

NMR samples contained 1–3 mM protein, in 300 mM NaCl, 10 mM phosphate buffer (pH 5.5), unless otherwise specified. Spectra were acquired on a 600 MHz Bruker spectrometer. Two-dimensional spectra (NOESY, Nuclear Overhauser Exchange Spectroscopy; DQ-COSY, Double Quantum Filtered Spectroscopy; and TOCSY, Total Correlation Spectroscopy) in 90% H<sub>2</sub>O, 10% D<sub>2</sub>O or 100% D<sub>2</sub>O were acquired at 20°C, 28°C and 37°C in order to remove ambiguities due to overlapping cross-peaks. For newly prepared samples, after gel filtration chromatography, the peptide appears to be >95% dimeric [the lines of the monomeric unfolded peptide are much sharper than those of the dimeric folded molecule and give intense and easily detectable cross-peaks on the two-dimensional, through bond coupling, COSY, TOCSY or heteronuclear multiple quantum coherence (HMQC) experiments]. The ratio of dimeric to monomeric molecules decreases after a few days. However, as indicated by the absence of exchange cross-peaks between the folded dimeric and the unfolded monomeric molecules on the NOESY spectra, interconversion between the species is slow. Complete assignments were obtained for the folded dimeric peptide and the chemical shifts have been deposited in the BioMagRes Database. Distance restraints were obtained from NOESY spectra recorded at 80 ms mixing time. They were classified as strong ( $\leq 2.7$  Å), medium ( $\leq 3.3$  Å), weak ( $\leq 3.8$  Å) and very weak ( $\leq 4.5$  Å). Complex formation was monitored via a HMQC experiment.

In a first step, the monomeric unit was modelled using the hybrid Distance Geometry-Simulated Annealing (DG-SA) protocol provided with the X-PLOR 3.1 package (Brünger, 1992). Only unambiguous intramolecular NOE derived distance restraints and dihedral data obtained from coupling constants were used at this stage. The conformers of low restraint energy violation (within 30% of the conformer of lower restraint energy violation and at least 10) were retained for analysis. Some ambiguous NOE cross-peaks were then classified as intermolecular when the intramolecular distance between the couple of protons involved was >8 Å in all conformers and when addition of this NOE to the intramolecular constraint list resulted in higher restraint energy violation for the conformers of the ensemble.

The dimer was modelled using the simulated annealing protocol described by Nilges (1993) and summarized in Table I. Symmetry of the dimer was obtained by forcing the distance between the C $\alpha$  of residue 1 in monomer A and the C $\alpha$  of residue 55 in monomer B to be equal to that between the C $\alpha$  of residue 1 in monomer B and the C $\alpha$  of residue 55 in monomer A, etc. Electrostatic, empirical dihedral and attractive (Lennard-Jones) van der Waals potentials were not used during the simulated annealing steps. Attractive van der Waals were turned on during a final stage of modelling which consisted of 250 steps of conjugate gradient minimization. The parameters used for modelling were those of the all atom parameter file parallhdg (Brünger, 1992).

## **Acknowledgements**

We thank C.Gaillardin, E.Fabre, F.Dardel, H.van Tilbeurgh and N.Declerck for critical reading of the manuscript, and O.Amster-Choder for unpublished data. This work was supported by the Institut National de la Recherche Agronomique, the Centre National de la Recherche Scientifique, the Ministère de la Recherche (Action Coordonnée Concertée Science de la Vie) and an EEC BIO-2CT930345 grant. X.M. was the recipient of a French MRT fellowship.

## **References**

- Allain,F.H.-T., Gubser,C.C., Howe,P.W., Nagai,K., Neuhaus,D. and Varani,G. (1996) Specificity of ribonucleoprotein interaction determined by RNA folding during complex formulation. *Nature*, **380**, 646–650.
- Amster-Choder,O. and Wright,A. (1992) Modulation of the dimerization of a transcriptional antiterminator protein by phosphorylation. *Science*, **257**, 1395–1398.

- Amster-Choder, O. and Wright, A. (1993) Transcriptional regulation of the *bgl* operon of *Escherichia coli* involves phosphotransferase system-mediated phosphorylation of a transcriptional antiterminator. *J. Cell. Biochem.*, **51**, 83–90.
- Anagnostopoulos, C. and Spizizen, J. (1961) Requirements for transformation in *Bacillus subtilis*. *J. Bacteriol.*, **81**, 741–746.
- Antson, A. A. et al. (1995) The structure of *trp* RNA-binding attenuation protein. *Nature*, **374**, 693–700.
- 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.
- Aymerich, S., Gonzy-Tréboul, G. and Steinmetz, M. (1986) 5′-noncoding region *sacR* is the target of all the regulation affecting the levanucrase gene in *Bacillus subtilis*. *J. Bacteriol.*, **166**, 993–998.
- Bardowski, J., Ehrlich, S. D. and Chopin, A. (1994) BglR protein, which belongs to the BglG family of transcriptional antiterminators, is involved in beta-glucoside utilization in *Lactococcus lactis*. *J. Bacteriol.*, **176**, 5681–5685.
- Biou, V., Yaremchuk, A., Tukalo, M. and Cusack, S. (1994) The 2.9 Å crystal structure of *T. thermophilus* seryl-tRNA synthetase complexed with tRNase<sup>ser</sup>. *Science*, **263**, 1404–1410.
- Brünger, A. (1992) *X-PLOR (Version 3.1): A System for X-ray Crystallography and NMR*. Yale University Press, New Haven, CT.
- Burd, C. G. and Dreyfuss, G. (1994) Conserved structures and diversity of functions of RNA-binding proteins. *Science*, **265**, 615–621.
- Bycroft, M., Hubbard, T. J. P., Proctor, M., Freund, S. M. V. and Murzin, A. G. (1997) The solution structure of the S1 RNA binding domain: a member of an ancient nucleic acid-binding fold. *Cell*, **88**, 235–242.
- Caverelli, J., Rees, B., Ruff, M., Thierry, J.-C. and Moras, D. (1993) Yeast tRNA<sup>Asp</sup> recognition by its cognate class II aminoacyl-tRNA synthetase. *Nature*, **362**, 181–184.
- Débarbouillé, M., Arnaud, M., Fouet, A., Klier, A. and Rappoport, G. (1990) The *sacT* gene regulating the *sacPA* operon in *Bacillus subtilis* shares strong homology with transcriptional antiterminators. *J. Bacteriol.*, **172**, 3966–3973.
- el Hassouni, M., Henrissat, B., Chippaux, M. and Barras, F. (1992) Nucleotide sequences of the *arb* genes, which control beta-glucoside utilization in *Erwinia chrysanthemi*: comparison with the *Escherichia coli bgl* operon and evidence for a new beta-glycohydrolase family including enzymes from eubacteria, archaeobacteria, and humans. *J. Bacteriol.*, **174**, 765–777.
- Golden, B. L., Ramakrishnan, V. and White, S. W. (1993) Ribosomal protein L6: structural evidence of gene duplication from a primitive RNA binding protein. *EMBO J.*, **12**, 4901–4908.
- Görlach, M., Wittekind, M., Beckman, R. A., Mueller, L. and Dreyfuss, G. (1992) Interaction of the RNA-binding domain of the hnRNP C proteins with RNA. *EMBO J.*, **11**, 3289–3295.
- Hoffman, D. W., Query, C. C., Golden, B. L., White, S. W. and Keene, J. D. (1991) RNA-binding domain of the ribonucleoprotein analysed by NMR spectroscopy is structurally similar to ribosomal proteins. *Proc. Natl Acad. Sci. USA*, **88**, 2495–2499.
- Hoffman, D. W., Davies, C., Gerchman, S. E., Kycia, J. H., Porter, S. J., White, S. W. and Ramakrishnan, V. (1994) Crystal structure of prokaryotic ribosomal protein L9: a bi-lobed RNA-binding protein. *EMBO J.*, **13**, 205–212.
- Houman, F., Diaz-Torres, M. and Wright, A. (1990) Transcriptional antitermination of *E. coli* is modulated by a specific RNA binding protein. *Cell*, **62**, 1153–1163.
- Howe, P. W. A., Nagai, K., Neuhaus, D. and Varani, G. (1994) NMR studies of U1 snRNA recognition by the N-terminal RNP domain of the human U1A protein. *EMBO J.*, **13**, 3873–3881.
- Kang, C., Chan, R., Berger, I., Lockshin, C., Green, L., Gold, L. and Rich, A. (1995) Crystal structure of the T4 RegA translational regulator protein at 1.9 Å resolution. *Science*, **268**, 1170–1173.
- Kraulis, P. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.*, **24**, 946–950.
- Mattaj, J. (1993) RNA recognition: a family matter? *Cell*, **73**, 837–840.
- Matthews, D. A. et al. (1994) Structure of human rhinovirus 3C protease reveals a trypsin-like polypeptide fold, RNA-binding site, and means for cleaving precursor polypeptide. *Cell*, **77**, 761–771.
- Milligan, J. F. and Uhlenbeck, O. C. (1989) Synthesis of small RNAs using T7 RNA polymerase. *Methods Enzymol.*, **180**, 51–62.
- Nagai, K. (1996) RNA-protein complexes. *Curr. Opin. Struct. Biol.*, **6**, 53–61.
- Nagai, K., Oubridge, C., Jessen, T. H., Li, J. and Evans, P. R. (1990) Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A. *Nature*, **348**, 515–520.
- Nicholls, A. (1992) *GRASP: Graphical Representation and Analysis of Surface Properties*. Columbia University, New York.
- Nilges, M. (1993) A calculation strategy for the structure determination of symmetric dimers by 1H NMR. *Proteins*, **17**, 297–309.
- Oubridge, C., Ito, N., Evans, P. R., Teo, C. H. and Nagai, K. (1994) Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature*, **372**, 432–438.
- Rougée, M., Faucon, B., Mergny, J. L., Barcelo, F., Giovannangeli, C., Garestier, T. and Hélène, C. (1992) Kinetics and thermodynamics of triple-helix formation: effects of ionic strength and mismatches. *Biochemistry*, **31**, 9269–9278.
- Rould, M. A., Perona, J. J., Soll, D. and Steitz, T. A. (1989) Structure of *E. coli* glutamyl-tRNA synthetase complexed with tRNA<sup>Gln</sup> and ATP at 2.8 Å resolution. *Science*, **246**, 1135–1142.
- Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Schindelin, H., Marahiel, M. A. and Heinemann, U. (1993) Universal nucleic acid-binding domain revealed by crystal structure of the *B. subtilis* major cold-shock protein. *Nature*, **364**, 164–168.
- Schnetz, K., Stülke, J., Gertz, S., Krüger, S., Krieg, M., Hecker, M. and Rak, B. (1996) Lic T, a *Bacillus subtilis* transcriptional antiterminator of the BglG family. *J. Bacteriol.*, **178**, 1971–1979.
- Steinmetz, M. (1993) Carbohydrate catabolism: pathways, enzymes, genetic regulation and evolution. In Sohenshein, A. L., Hoch, J. A. and Losick, R. (eds), *Bacillus subtilis, and other Gram-Positive Bacteria, Biochemistry, Physiology, and Molecular Genetics*. American Society for Microbiology, Washington, DC, pp. 157–170.
- Stragier, P., Bonamy, C. and Karmazin-Campelli, C. (1988) Processing of a sporulation sigma factor in *Bacillus subtilis*: how morphological structure could control gene expression. *Cell*, **52**, 697–704.
- Tortosa, P., Aymerich, S., Lindner, C., Saier, M. H., Jr, Reizer, J. and Le Coq, D. (1997) Multiple phosphorylation of SacY, a *Bacillus subtilis* transcriptional antiterminator negatively controlled by the phosphotransferase system. *J. Biol. Chem.*, **272**, 17230–17237.
- 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., 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.
- Willbold, D., Rosin-Arbesfeld, H., Frank, R. and Rösch, P. (1994) Structure of the equine infectious anemia virus Tat protein. *Science*, **264**, 1584–1587.
- Wittekind, M., Görlach, M., Friedrichs, M., Dreyfuss, G. and Mueller, L. (1992) 1H, 13C, and 15N NMR assignments and global folding pattern of the RNA-binding domain of the human hnRNP C protein. *Biochemistry*, **31**, 6254–6265.
- Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*. John Wiley & Sons, New York.
- Yon, J. and Fried, M. (1989) Precise gene fusion by PCR. *Nucleic Acids Res.*, **17**, 4895.

Received on March 19, 1997; revised on May 21, 1997