RNA recognition and base flipping by the toxin sarcin

C. C. Correll,* X. Yang, T. Gerczei, J. Beneken and M. J. Plantinga

Department of Biochemistry and Molecular Biology, The University of Chicago, 920 E58th Street, Chicago, IL 60637, USA. E-mail: ccorrell@uchicago.edu

Sarcin is a member of a fungal toxin family that enters cells and specifically cleaves one of the thousands of RNA phosphodiester bonds in the ribosome. As a result, elongation factor binding is disrupted, translation is inhibited and apoptosis is triggered. The toxin targets a universal RNA structure in the ribosome called the sarcin/ricin loop (SRL). A 1.11 Å resolution structure of a minimal SRL RNA substrate (~30-mer) shows that the loop portion of the substrate folds into two common building blocks of RNA structure: a bulged-G motif (recognition site) and a GAGA tetraloop (cleavage site). To elucidate the structural basis of toxin action, we determined two cocrystal structures of the sarcin homologue restrictocin bound to different analogs of a minimal SRL RNA substrate. Our studies argue that site selection by the toxin depends on direct base and shape recognition of the SRL RNA, and that cleavage by the toxin depends on a base flipping mechanism that positions the nucleophile for in-line attack on the scissile bond.

Keywords: **RNA motifs; RNA-protein recognition; induced fit.**

1. Introduction

GTP hydrolysis by four protein factors with related GTPase domains help drive initiation, elongation and termination of ribosome-directed protein synthesis. The emerging view is that, when bound to GTP, these factors interact with the large subunit of the ribosome in a similar manner (Moore & Steitz 2002). After the factor has accomplished its task, the ribosome activates GTP hydrolysis, triggering factor release. The ribosomal RNA and/or protein sites responsible for activating GTP hydrolysis and for interacting with the switch regions of these GTPase protein factors are unknown. One intriguing candidate that has been shown to lie near the switch regions of elongation factors (EFs) by cryoelectron microscopy is the universal sarcin/ricin loop (SRL) of 23 rRNA (Valle *et al.* 2002). As shown in Fig. 1, the SRL RNA lies on the surface of the ribosome. The site (2653 - 2667; *E. coli* 23S numbering is used throughout) derives its name from two separate families of site-specific toxins—sarcin and ricin—that target it for covalent modification. Sarcin family members and the ricin analog pokeweed antiviral protein (PAP) target ribosomes from the three kingdoms of life; whereas ricin targets only eukaryotic ribosomes (Endo & Tsurugi 1987; Marchant & Hartley 1995). Endonucleolytic cleavage by sarcin of the P-O5' bond in A2662 or depurination by PAP of A2660 disrupts the binding of EFs. As a consequence of toxin action, protein synthesis is inhibited and apoptosis is triggered by an unknown mechanism (Olmo *et al*. 2001). The SRL RNA thus forms a critical component of the binding site for EFs, and possibly for other GTPase protein factors, such as initiation factor IF2 (La Teana *et al.* 2001; Cameron *et al.* 2002).

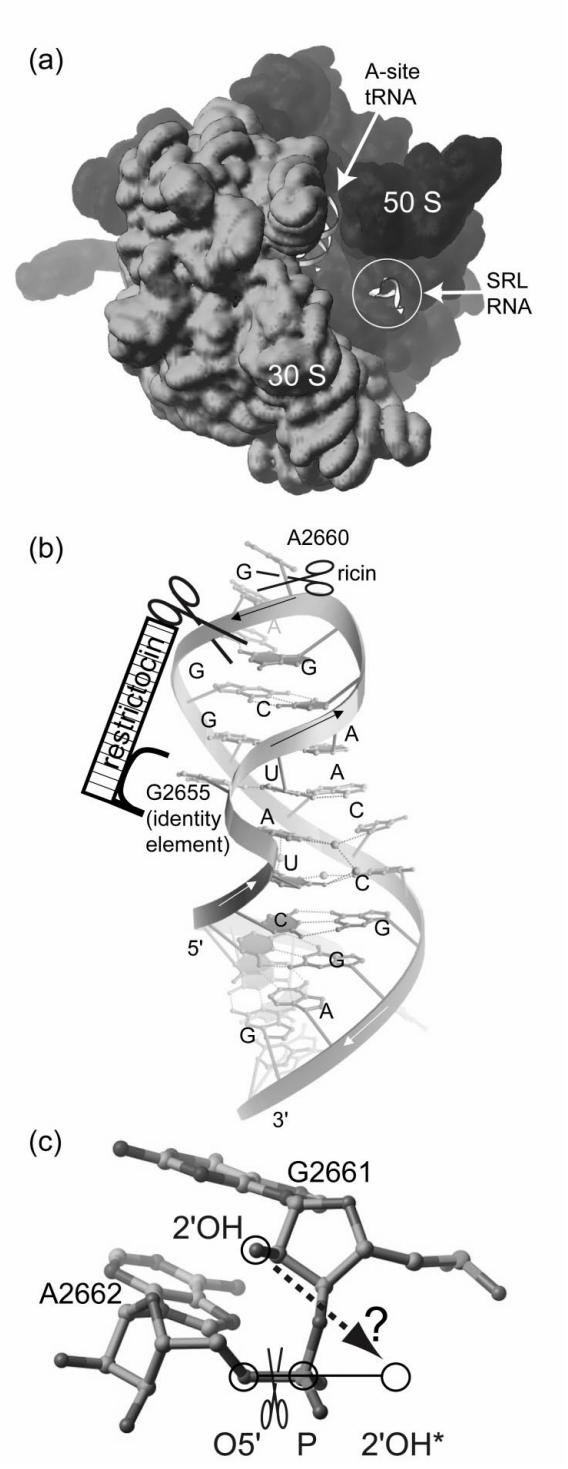

Figure 1

The SRL RNA substrate. (a) The contact surface location of the SRL RNA on the ribosome from PDB entries 1GIX and 1GIY (Yusupov *et al.* 2001). (b) Ribbon drawing of the *E. coli* SRL RNA schematically showing ricin depurination and restrictocin recognition and cleavage. (c) The ground state arrangement of the nucleophilic 2'-hydroxyl group is inconsistent with an in-line attack. Panels (b) and (c) were drawn from PDB entry 483D (Correll *et al*. 1999).

2. Substrate SRL RNA structure

To better understand the structural basis of sarcin recognition, we have carried out structural studies using minimal rRNA substrates. The RNAs are synthetic oligonucleotides (~30-mers) with the SRL nucleotide sequence that mimic the form and function of the SRL in the ribosome (Ban *et al.* 2000). SRL mimics have served as a minimal substrate for EF-G binding (Munishkin & Wool 1997), for sarcin and ricin activity (Wool 1997), and for other structural studies (Szewczak *et al.* 1993; Szewczak & Moore 1995; Correll *et al.* 1998; Seggerson & Moore 1998; Correll *et al.* 1999; Rife *et al.* 1999). As shown in Fig. 1b, the stem of the SRL RNA connects to the loop portion that folds into a GAGA tetraloop and a bulged G motif. In the GAGA tetraloop, ricin depurinates the first A and sarcin cleaves the scissile P-O5' bond of the second A (Fig. 1b). Ricin recognition depends solely on the GAGA tetraloop (Gluck *et al.* 1992); sarcin and EF recognition depends primarily on the bulged-G (2655) of the bulged-G motif located ~12 Å away from the scissile bond (Correll *et al.* 1999). G2655 is the only nucleotide known to be critical for recognition of the SRL RNA by sarcin and is, therefore, referred to as the identity element (Wool 1997). The termini generated by sarcin (a 5'-hydroxyl group and 2'-3'-cyclic phosphate) indicate that the nucleophile attacks the scissile bond via an in-line attack mechanism with a trigonal bipyramidal transition state. The ground state SRL RNA structure shows that the nucleophilic 2'-hydroxyl group lies adjacent to and not in-line with the scissile bond (Fig. 1c). Apparently, conformational change is required to position the nucleophile for in-line attack on the scissile bond. To better understand the specificity of sarcin and the nature of this conformational change we determined two co-crystal structures using the toxin restrictocin from *Aspergillus restrictus*, which shares 86% sequence identity with sarcin (Yang *et al.* 2001).

3. Insights into specificity

To mimic a Michaelis intermediate for cocrystallization trials, we designed SRL RNA variants that minimize cleavage and closely mimic the substrate by substituting certain nucleophilic 2'-hydroxyl groups with poor nucleophiles of comparable size. One cocrystal structure, determined to 2 Å resolution, illustrates site-selection with specific contacts to the identity element (G2655); therefore, we refer to it as the "bound" structure (Yang *et al.* 2001) (Fig. 2). Contacts between two loops of the toxin and the identity element guide site selection and may mimic critical contacts in a Michaelis intermediate. Interestingly, a hydrated potassium ion that bridges the enzyme-substrate interface may also guide site selection.

4. Sarcin uses a base-flipping mechanism

Another cocrystal structure, determined to 2.2 Å resolution, illustrates how the toxin enables cleavage by docking A2662 in the active site with its 2' nucleophilic oxygen atom nearly inline for attack on the scissile P-O5' bond. The complex is misdocked by one nucleotide so we refer to it as the "misdocked" structure (Yang *et al.* 2001).

Restrictocin-like toxins may use base flipping to enable cleavage at the correct site. The base-flipped geometry of A2662 in the misdocked structure is similar to that observed for the guanosine in the active site of a T1-analog complex structure (Arni *et al.* 1999) (Fig. 3). Moreover, it is possible to create a model structure with reasonable stereochemistry for the correctly docked complex that combines features of the T1-analog structure (Arni *et al.* 1999) and of the bound structure. In the model structure, a base-flipped G2661 docks in the active site, as observed in the T1-analog structure (Fig. 3a), and L2 and L4 of restrictocin contact the identity element (G2655) and the surrounding S-turn, as observed in the bound structure (Fig. 2b). The misdocked structure provides clues to the structural basis of discrimination between cleavage and miscleavage because it fails to make direct contacts to the identity element and those to the surrounding S-turn moiety are weakened. Steric clash is observed between N6 of A4326 and nearby backbone amides, which make optimal contacts to the O6 of the target guanine base. The clash prevents snug docking of A4326 in the active site and thereby favors correct docking over misdocking.

The toxin may play an active role in the base flipping and restacking associated with unfolding the tetraloop. Superposition of the bulged-G motif structure from the SRL RNA in the *H. marismortui* 50S subunit (Klein *et al.* 2001) and that from the bound complex (Yang *et al.* 2001) shows (1) that the SRL RNA is available for toxin recognition (Fig. 3b) and (2) that the minor steric clash between the toxin and the folded tetraloop in the ribosomal target is relieved by base flipping (Fig. 3c). The bulged-G motif structures were superimposed because they remain unchanged upon complex formation with the toxin. The superposition demonstrates that contacts to the S-turn are incompatible with a folded tetraloop. The energy required to unfold a GNRA is estimated at ~1 kcal/mol (Antao *et al.* 1991). Steric hindrance may therefore provide sufficient driving force to unfold the tetraloop when the toxin docks with the S-turn of the SRL RNA.

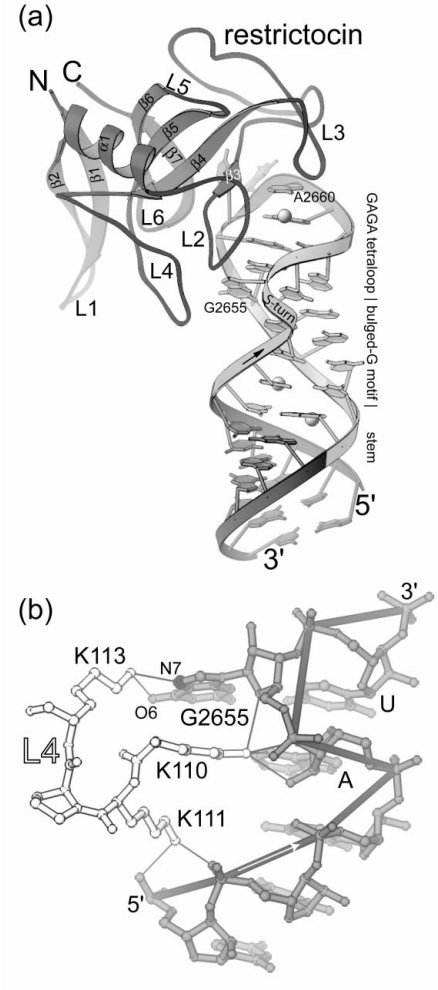

Figure 2

The bound restrictocin-inhibitor cocrystal structure from PDB entry 1JBS (Yang *et al.* 2001). (a) Ribbon drawing showing how loops L2 and L4 contact the bulged-G identity element. (b) Sequence specific contacts between lysines in L4 and the identity element and surrounding S-turn.

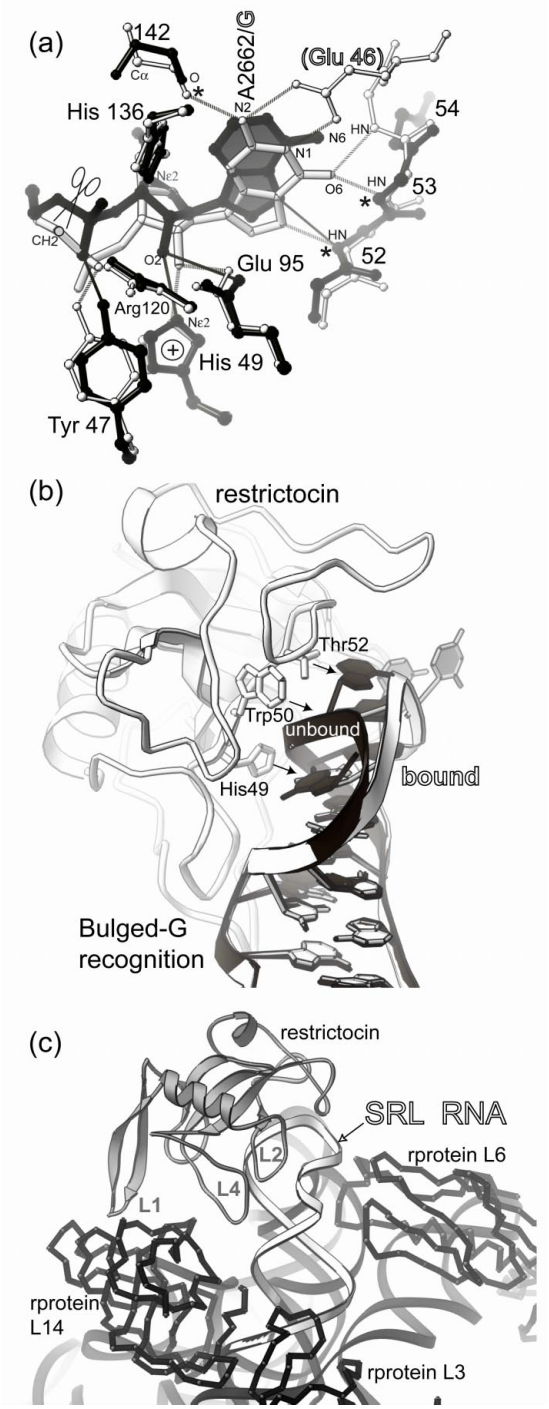

Figure 3

Base flipping. (a) The misdocked restrictocin-inhibitor cocrystal structure from PDB entry 1JBR (Yang *et al.* 2001) (dark) superimposed with the structure of RNase T1-analog complex from PDB entry 1B2M (Arni *et al.* 1999) (light). (b) Superposition of the bulged-G motif of the bound structure (white) and the equivalent motif of the 50S subunit structure from PDB entry 1JJ2 (Klein *et al.* 2001) (black). (c) Same as (b) except that the RNA from the bound structure has been omitted for clarity. Nearby ribosomal proteins (rproteins) are shown. Mutational studies show that the putative contact between L1 of restrictocin and rprotein L14 are not required for cleavage (Gluck & Wool 2002).

5. Possible base flipping mechanism of ricin

Ricin may also use a base flipping mechanism to dock its target (A2660) in the active site before depurination. In the uncomplexed structure of the SRL RNA, the glycosidic bond of A2660 is *anti* and its base stacks on the base of G4325 (Fig. 4). In the structure of ricin bound to a substrate analog (formycin monophosphate), the glycosidic bond of the analog is *syn* and its base packs snuggly in a base-binding pocket, surrounded by side chain contacts (Weston *et al*. 1994). To dock the SRL RNA in a similar manner, base flipping of A2660 would disrupt base stacking with the 3'-adjacent guanosine and could rotate the glycosidic bond of A2660 from the *anti* to the *syn* rotamer, as observed in the misdocked structure of restrictocin.

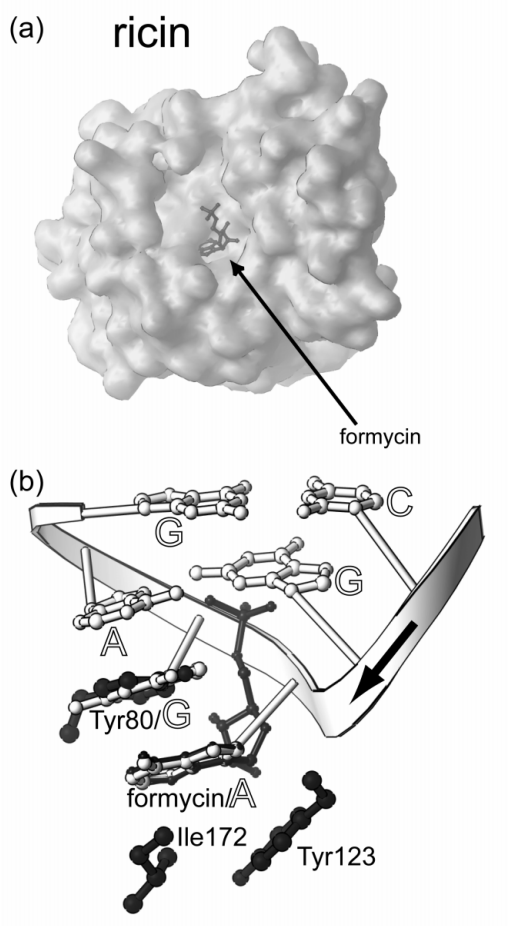

Figure 4

Ricin interacting with the noncleavable AMP analog formycin from PDB entry 1FMP (Monzingo & Robertus 1992). (a) A surface representation of ricin and (b) a superposition of the base of formycin (black) with the target A2660 of an unbound GAGA tetraloop from PDB entry 483D (Correll *et al.* 1999) shows severe steric clash between Tyr80 of ricin and the second G of the tetraloop.

References

Antao, V.P., Lai, S.Y. & Tinoco, I. (1991) *Nucleic Acids Res* **19**, 5901-5905.
Arni, R.K., Watanabe, L., Ward, R.J., Kreitman, R.J., Kumar, K. & Walz, F.G. (1999). *Biochemistry* **38**, 2452-2461.
Ban, N., Nissen, P., Hansen, J., Moore, P.B. & Steitz, T.A. (2000). *Science* **289**, 905-920.
Cameron, D.M., Thompson, J., March, P.E. & Dahlberg, A.E. (2002). *J Mol Biol* **319**, 27-35.

Correll, C.C., Munishkin, A., Chan, Y.L., Ren, Z., Wool, I.G. & Steitz T.A. (1998). *Proc Natl Acad Sci USA* 95, 13436-13441.
Correll, C.C., Wool, I.G. & Munishkin A. (1999). *J Mol Biol* **292**, 275-287.
Endo, Y. & Tsurugi, K. (1987). *J Biol Chem* **262**, 8128-8130.
Gluck, A., Endo, Y. & Wool, I.G. (1992). *J Mol Biol* **226**, 411-424.
Gluck, A. & Wool, I.G. (2002). *Biochim Biophys Acta* **1594**, 115-126.
Klein, D.J., Schmeing, T.M., Moore, P.B. & Steitz, T.A. (2001). *Embo J* **20**, 4214-4221.
La Teana, A., Gualerzi, C.O. & Dahlberg, A.E. (2001). *RNA* **7**, 1173-1179.
Marchant, A. & Hartley, M.R. (1995). *J Mol Biol* **254**, 848-855.
Monzingo, A. F. & J. D. Robertus (1992). *J Mol Biol* **227**, 1136-1145.
Moore, P.B. & Steitz, T.A. (2002). *Nature* **418**, 229-235.
Munishkin, A. & Wool, I.G. (1997). *Proc Natl Acad Sci USA* **94**, 12280-12284.
Olmo, N., Turnay, J., De Buitrago, G.G., De Silanes, I.L., Gavilanes, J.G. & Lizarbe, M.A. (2001). *Eur J Biochem* **268**, 2113-2123.
Rife, J.P., Stallings, S.C., Correll, C.C., Dallas, A., Steitz, T.A. & Moore, P.B. (1999). *Biophys J* **76**, 65-75.
Seggerson, K. & Moore, P.B. (1998). *RNA* **4**, 1203-1215.
Szewczak, A.A. & Moore, P.B. (1995). *J Mol Biol* **247**, 81-98.
Szewczak, A.A., Moore, P.B., Chan Y.L. & Wool I.G. (1993). *Proc Natl Acad Sci USA* **90**, 9581-9585.
Valle, M., Sengupta, J., Swami, N.K., Grassucci, R.A., Burkhardt, N., Nierhaus, K.H., Agrawal, R.K. & Frank, J. (2002). *Embo J* **21**, 3557-3567.
Wool, I.G. (1997). San Diego: Academic Press, Inc.
Yang, X., Gerczei, T., Glover, L. & Correll, C.C. (2001). *Nat Struct Biol* **8**, 968-973.
Yusupov, M.M., Yusupova, G.Z., Baucom, A., Lieberman, K., Earnest, T.N., Cate, J.H. & Noller, H.F. (2001). *Science* **292**, 883-896.

 Received 30 April 2003 • Accepted 15 October 2003