Role REVersal: understanding how RRE RNA binds its peptide ligand

Dilara Grate and Chuck Wilson

The structure of a complex between the HIV Rev-responsive element (RRE) RNA and a fragment of the Rev protein has recently been determined by NMR. Together with previous studies of the Tat–TAR complex, these results show how RNA elements with considerable tertiary structure are able to play a more active role in directing binding to elements of protein secondary structure.

Address: Department of Biology and Center for the Molecular Biology of RNA, University of California at Santa Cruz, 1156 High Street, Santa Cruz, CA 95064, USA.

E-mail: grate@biology.ucsc.edu
wilson@biology.ucsc.edu

Electronic identifier: 0969-2126-005-00007

Structure 15 January 1997, 5:7–11

Introduction

Specific protein–nucleic acid interactions are critical for the proper functioning of countless biological processes, including DNA replication, regulated gene transcription and protein translation [1–3]. In the last ten years, a series of elegant structural studies focusing mainly on DNA-binding transcription factors has yielded a detailed description of how proteins and nucleic acids associate in a sequence specific way [4]. The picture that emerges from this work almost always depicts the nucleic acid component of a complex as a static, relatively featureless ligand, falling neatly into an elaborately evolved complementary surface formed by a protein 'receptor'. Because these studies have centered almost entirely on protein–DNA complexes, the mechanisms for specific protein–RNA recognition have remained largely unexplored. In contrast to conventional duplex DNA, biologically derived RNAs have many opportunities for forming complex tertiary structures, structures which might ultimately be used to direct specific associations with proteins [5]. NMR studies of two protein–RNA complexes from human immunodeficiency virus type 1 (HIV-1) (first Tat–TAR and now Rev–RRE [6,7]) are beginning to provide key insights into the ways in which RNAs and proteins recognize each other. Whereas previous structures have invariably featured proteins with tertiary structure recognizing simple elements of nucleic acid secondary structure, the structures of TAR and RRE demonstrate how these conventional roles may be reversed, with tertiary structured RNAs forming elaborate binding sites for minimal elements of protein secondary structure [8,9]. This review focuses primarily on the recent work from the Williamson and Patel groups on the Rev–RRE interaction. Readers interested in an overview of the Tat–TAR complex are referred to a corresponding review by Wemmer [10].

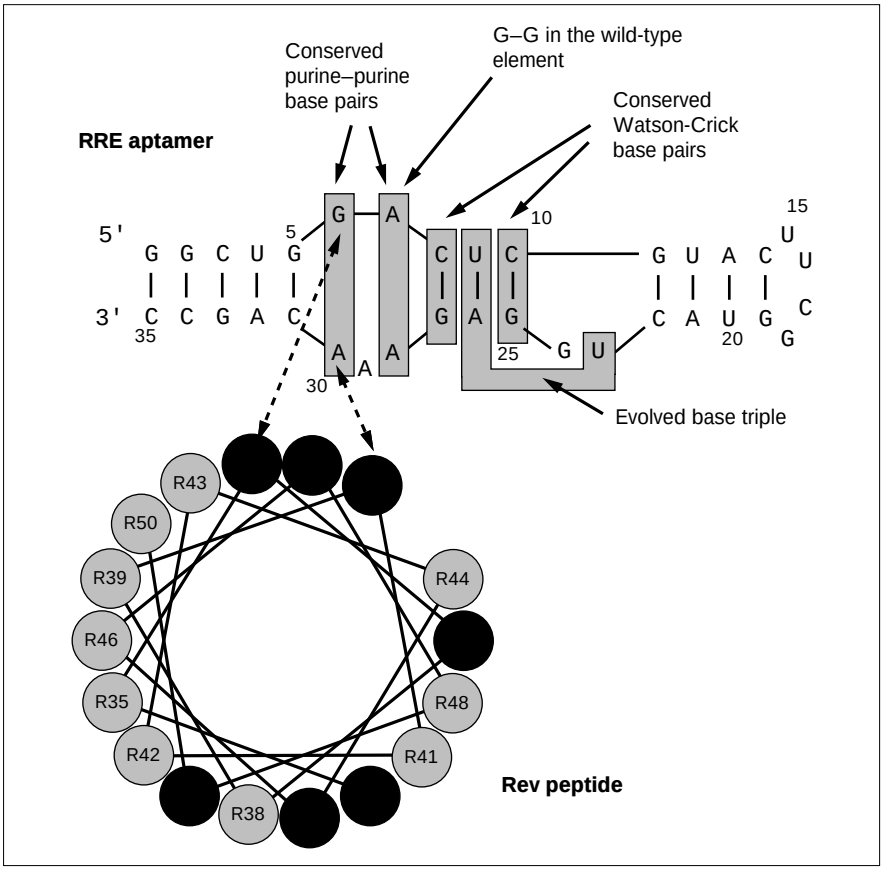
The biology of Rev–RRE

The protein Rev plays an important role in the life cycle of HIV by facilitating the cytoplasmic export of incompletely spliced mRNAs [11,12]. This activity is predictated upon Rev's binding to the *cis*-acting Rev-responsive element (RRE), carried by unspliced viral transcripts within the *env* gene [13]. Over the last decade, a wealth of both *in vivo* and *in vitro* experiments have dissected the amino acids and nucleotides required for a functional interaction between these two molecules. Deletion studies have identified an unusually arginine-rich 17 amino acid fragment of Rev (Fig. 1) that binds to RRE with an affinity comparable to that of the full length 116 amino acid protein. Other portions of Rev are thought to mediate protein oligomerization and to direct interactions with the nuclear mRNA export apparatus [14,15]. Circular dichroism showed that the 17 amino acid peptide fragment adopts an α-helical conformation which is particularly stabilized in the presence of the RRE [16]. *In vitro* selection experiments with pools of randomized RRE sequences have been useful for both defining the secondary structure of the RRE and for identifying tertiary interactions that direct the higher-order folding of this RNA [17,18]. Two purine–purine base pairs have been identified at the heart of the RRE high-affinity site. Bartel *et al.* showed that one of these, a G–G pair, could be replaced by an A–A pair without diminishing binding [19]. In addition to these studies on the individual components of the complex, a genetic screen was recently used to map out the pairwise interactions between the peptide and the RNA by searching for amino acid substitutions in Rev that could complement specific down mutations in the Rev-responsive element [20]. The essential results from these previous studies are summarized in Figure 1.

The structures

Using heteronuclear NMR techniques, two groups working independently have now determined the three-dimensional structure of the high affinity RRE site complexed with the arginine-rich Rev peptide. The relatively large number of experimental constraints used in each case, together with a significant degree of agreement between the two structures, suggests that their results provide an accurate representation of the biologically relevant structure. The major technical difficulty in solving this structure was interpre-tation of the many overlapping arginine

Figure 1



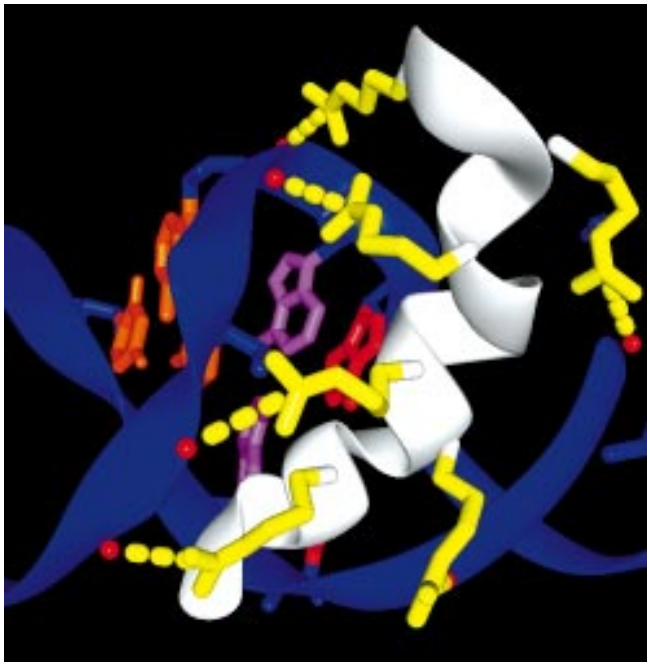
Two-dimensional representations of the RRE RNA aptamer structure (above) and the Rev peptide α helix (below). The dashed arrows indicate interactions suggested by suppressor mutations. (Both representations correspond to the constructs used in the Patel group NMR study.)

resonances; the two groups arrive at similar results by employing different strategies. Battiste *et al.* took advantage of a doubly-labeled (^{13}C,^{15}N) peptide and ^{13}C-labeled RNA to facilitate triple-resonance experiments, while Ye *et al.* have used ^{13}C-,^{15}N-labeled RNA and a series of mutant peptides with single amino acid substitutions to make their assignments. In addition to these spectroscopic differences, the two groups have solved the structures of slightly different complexes. The Rev peptide used by the Williamson group is longer at both the N and C termini while the RRE RNA used by the Patel group has been 'optimized' by *in vitro* selection and contains a number of tertiary interactions not observed in the original HIV-1 sequence. Generally speaking, the Patel group structure is based on more NOE constraints, allowing the construction of a more detailed model of the structure. While the particulars differ, the bulk of the interactions observed in the final structures are generally duplicated and provide a clear view as to how the high affinity and specificity of binding is achieved.

A snapshot impression of the structures

A cursory examination of the Rev–RRE structure immediately reveals several striking features which define the nature of the complex. Firstly, two purine–purine base pairs separated by a non-conserved bulged residue cause the RRE backbone to adopt an unusual S-shaped fold. Local undertwisting of the base pairs in this region helps to effectively double the width of the major groove, making it broad enough to readily accommodate a helical peptide. Secondly, arginines (as expected) play a critical role in binding and do so by making three different types of interactions. Six arginines are directed up and away from the major groove and are involved in direct neutralization of the phosphate backbone. These residues form two stripes along either side of the helical peptide (Fig. 2). One can readily imagine that these 'arginine zippers' act to bridge the two phosphate ridges that define the major groove. In addition, three other arginines (Arg35, Arg39 and Arg44) participate in direct hydrogen bonds to specific bases (all involving either the O6 or N7 atoms of the RRE guanosines). A third type of interaction is observed in the Patel group structure where two arginine residues are found to make close stacking interactions with a uridine base involved in a base triple (described below). Thirdly, there are virtually no sidechain interactions within the Rev peptide; contacts with the RRE dominate to define its conformation. The main exception

Figure 2



Highlighted view of the RRE–Rev interaction (based on coordinates provided by D Patel and A Gorin). Coloring is as follows: arginines in the Rev peptide that interact with RRE phosphates (Arg38, Arg42, Arg46, and Arg50 on one face; Arg41 and Arg48 on the other) are in yellow; the conserved G–A base pair which interacts with Asn40 is in red; the structurally important A–A base pair (G–G in the original HIV-1 RRE) is in purple; the U–A–U base triple introduced by *in vitro* selection is in orange.

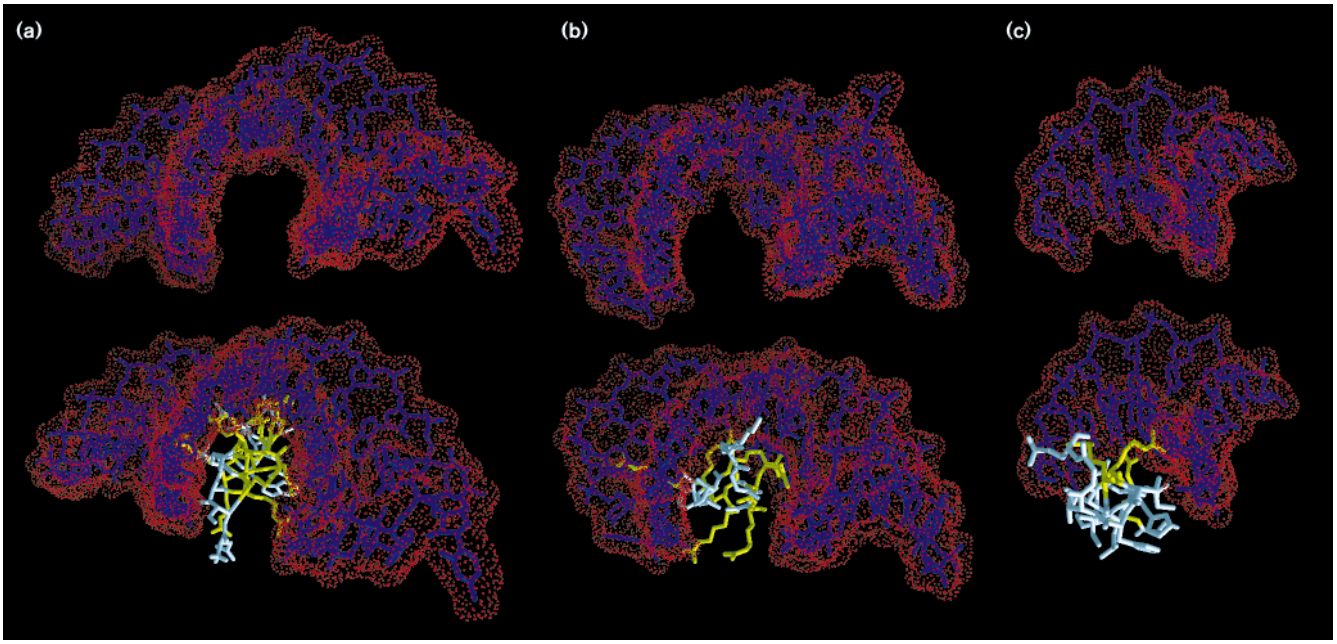
to this involves Glu47, the sole negatively charged amino acid in the Rev peptide. In both the Patel and Williamson structures, the glutamate sidechain carboxylate forms hydrogen bonds or salt bridges to multiple neighboring arginines, helping to define their conformation and thus direct their interactions with the RRE. Fourthly, the Patel structure contains a U–A–U base triple, introduced into the RRE in the course of *in vitro* optimization by a process termed the selective evolution of ligands by exponential enrichment (SELEX). The pairing of a bulged uridine with an A–U basepair (Fig. 1) accomplishes two things. By wedging itself into the major groove, the uridine base creates two well defined pockets on either face of the groove for Arg35 and Arg39; stacking between the nucleotide and these sidechains presumably contributes to the binding energy. In addition, the non-Watson-Crick uridine base in the triple acts as a convenient tether to help anchor that section of the RRE backbone that contacts the Rev peptide. Finally, previous genetic studies have strongly suggested that Asn40 plays a role in making base-specific contacts [20]. In particular, mutation of the conserved G–A pair at the center of the RRE may be compensated for by mutation of Asn40. In nice agreement with this result, the NMR structures place Asn40 in close proximity to the adenosine of the G–A base pair, positioned to hydrogen bond via its exocyclic amine.

The uniqueness of the Rev–RRE complex may be best appreciated by comparison to other protein–RNA and protein–DNA structures. The complexes formed by the $tRNA^{Gln}$ synthetase and the U1A protein with RNA are typical of the few previously determined protein–RNA structures [21,22]. In both of these cases, complex formation is driven in part by distinct binding pockets on the surface of each protein which are optimized for interaction with single nucleotides on the corresponding RNA partners. As a consequence of binding, the recognized nucleotides lose their previous structure and make interactions almost exclusively with the protein. The interaction observed between Rev and the RRE is fundamentally different in two respects: the structure of the RNA allows it to wrap around the protein instead of vice versa, and the protein component has virtually no internal structure; amino acid sidechains predominantly interact with the RNA and not with each other.

Studies on the Tat–TAR complex by the Puglisi and Patel groups have shown that, in certain contexts, RNA structure may play an important role in directing protein–RNA interactions [9,10]. Given that both Tat and Rev bind via short arginine-rich peptides and both TAR and RRE consist of RNA hairpins with internal loops and bulges, one might have expected significant similarities for the corresponding complexes. The Tat–TAR and Rev–RRE structures, however, are fundamentally different in many different respects (Fig. 3). In the Tat–TAR structure, the peptide folds as a β hairpin and binding with the RNA is directed in part by hydrogen bonding between the protein mainchain and the RNA. In contrast, the Rev peptide folds as an α helix, leaving only sidechains available to mediate the interaction with the RRE. While both Tar and Rev peptides bind in widened major grooves of a distorted RNA helix, the extent of the distortion is significantly less in the Tat–TAR complex (1–1.5 Å versus 4–5 Å). The relative narrowness of the major groove in TAR is a strong constraint on the sequence of Tat , requiring small glycines at three different residues in the peptide. Despite these differences, similar types of interactions are observed at a finer resolution and the formation of both complexes is largely driven by pairings between the guanidinium groups of arginines and either guanosine bases or backbone phosphates of the RNAs.

A final comparison that may be drawn is to the transcription factors that insert an α helix into the major groove of DNA to direct sequence-specific binding (including helix-turn-helix and zinc finger proteins). As shown in Figure 3, the helix from a typical DNA-binding protein penetrates

Figure 3



Different approaches to protein–nucleic acid complex formation. (a) The Rev–RRE complex (coordinates provided by J Williamson), (b) The Tat–TAR complex (coordinates provided by J Puglisi) and (c) the zif268 zinc finger–DNA complex (the α helix is part of the N-terminal finger in this three-finger complex [23]). Solvent accessible surfaces calculated for the nucleic acid component of each complex are shown above in isolation. Both RNA–protein complexes feature much deeper penetration by the protein and numerous electrostatic interactions involving arginine sidechains (colored yellow).

its nucleic acid target much less deeply than does the Rev peptide, a limitation imposed by the relative shallowness of the DNA major groove. In general, accessory elements of the protein structure provide many of the backbone phosphate interactions that direct high affinity but low specificity binding by DNA-binding proteins. The Rev α helix, perhaps because of its deep positioning within the RRE, is able to make extensive backbone and base-specific interactions and thus binds efficiently without the help of any additional protein contacts.

Conclusions

The structure of the Rev–RRE complex has provided important insights into the mechanisms by which RNA structure is able to direct binding to a specific protein. Several important features observed in this structure are likely to be seen again in other contexts : widening of the major groove by non-canonical base pairs, electrostatic zippering together of rows of backbone phosphates and arginines, direct hydrogen bonding between guanosines and arginines, base triples, and interdigitation of nucleotide bases and amino acid sidechains. The wide range of biological processes that require the formation of protein–RNA complexes has sparked a growing interest in this field. As such, we may expect these other contexts to fall victim to structural analysis sooner rather than later.

References

1. Pai, K.S., Bussiere, D.E., Wang, F., Hutchison, C.A., White, S.W. & Bastia, D. (1996). The structure and function of the replication terminator protein of *Bacillus subtilis*: identification of the winged helix DNA-binding domain. *EMBO J.* **15**, 3164–3173.
2. Torres, F.A. & Bonner, J.J. (1995). Genetic identification of the site of DNA contact in the yeast heat shock transcription factor. *Mol. Cell. Biol.* **15**, 5063–5070.
3. Ibba, M., Hong, K.W., Sherman, J.M., Sever, S. & Soll, D. (1996). Interactions between tRNA identity nucleotides and their recognition sites in glutaminyl-tRNA synthetase determine the cognate amino acid affinity of the enzyme. *Proc. Natl. Acad. Sci. USA* **93**, 6953–6958.
4. Nikolov, D.B., Chen, H., Halay, E.D., Hoffman, A., Roeder, R.G. & Burley, SK. (1996). Crystal structure of a human TATA box-binding protein/TATA element complex. *Proc. Natl. Acad. Sci. USA* **93**, 4862–4867.
5. Cate, J.H., *et al*., Doudna, J.A. (1996). Crystal structure of a group I ribozyme domain reveals principles of higher order RNA folding. *Science* **273**, 1678–1685.
6. Battiste, J.L., *et al*., & Williamson, J.R. (1996). Alpha helix-RNA major groove recognition in an HIV-1 rev peptide-RRE RNA complex. *Science* **273**, 1547–1551.
7. Ye, X., Gorin, A.G, Ellington, A.D. & Patel, J.D. (1996). Deep penetration of an α-helix into a widened RNA major groove in the HIV-1 rev peptide–RNA aptamer complex. *Nat. Struct. Biol.* **3**, 1026–1033.
8. Puglisi, J.D., Chen, L., Blanchard, S. & Frankel, A.D. (1995). Solution structure of a bovine immunodeficiency virus. *Science* **270**, 1200–1203.
9. Ye, X., Kumar, R.A. & Patel, D.J. (1995). Molecular recognition in the bovine immunodeficiency virus Tat peptide–TAR RNA complex. *Chem. Biol.* **2**, 827–840.
10. Wemmer, D.E. (1996). New features in RNA recognition: a Tat–TAR complex. *Chem. Biol.* **3**, 17–20.
11. Feinberg, M.B., Jarrett, R.F., Aldovini, A., Gallo, R.C. & Wong-Staal, F. (1986). HTLV-III expression and production involve complex regulation at the levels of splicing and translation of viral RNA. *Cell* **46**, 807–817.

12. Sodroski, J. (1986). A second post transcriptional transactivator gene required for HTLV-III replication. *Nature* **321**, 412–417.
13. Malim, M.H., Hauber, J., Le, S.Y., Maizel, J.V. & Cullen, B.R. (1989). The HIV-1 rev trans activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA. *Nature* **338**, 254–257.
14. Zapp, M.L., Hope, T.J., Parslow, T.G. & Green, M.R. (1991). Oligomerization and RNA binding domains of type-I human immunodeficiency virus Rev protein: a dual function for an arginine-rich binding motif. *Proc. Natl. Acad. Sci. USA* **88**, 7734–7738.
15. Kjems, J., Calnan, B.J., Frankel, A.D. & Sharp, P.A. (1992). Specific binding of a basic peptide from HIV-1 Rev. *EMBO J.* **11**, 1119–1129.
16. Tan, R. & Frankel, A.D. (1993). Costabilization of peptide and RNA structure in an HIV Rev peptide–RRE complex. *Cell* **73**, 1031–1040.
17. Werstruck, G., Zapp, M.L. & Green, M.R. (1996). A non-canonical base pair within the human immunodeficiency virus Rev-responsive element is involved in both Rev and small molecule recognition. *Chem. Biol.* **3**, 129–137.
18. Giver, L., Bartel, D.P., Zapp, M.L., Green, M.R. & Ellington, D.A. (1993). Selection and design of high-affinity RNA ligands for HIV-1 Rev *Gene* **137**, 19–24.
19. Bartel, D.P., Zapp, M.L., Green, M.R. & Szostak, J.W. (1991). HIV-1 Rev regulation involves recognition of non-Watson-Crick base pairs in viral RNA. *Cell* **67**, 529–536.
20. Jain, C. & Belasco, J.G. (1995). A structural model for the HIV-1 Rev–RRE complex deduced from altered specificity variants isolated by a rapid genetic strategy. *Cell* **87**, 115–125.
21. Rould, M.A., Perona J.J. & Steitz, T.A. (1991). Structural basis of anticodon loop recognition by glutaminyl-tRNA synthetase. *Nature* **352**, 213–218.
22. Oubridge, C., Ito, N., Evans, P.R., Teo, C.H. & 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.
23. Pavletich, N.P. & Pabo, C.O. (1991). Zinc finger-DNA recognition: crystal structure of a Zif268–DNA complex at 2.1 Å. *Science* **252**, 809–817.