A docking and modelling strategy for peptide–RNA complexes: applications to BIV Tat–TAR and HIV Rev–RBE

Jayashree Srinivasan[1,2], Fabrice Leclerc[1], Wei Xu[3], Andrew D Ellington[3] and Robert Cedergren[1]

Background: In spite of the great interest in the interaction between RNAs and proteins, no general protocol for modelling these complexes is presently available. This methodological vacuum is particularly acute because the structure of few such complexes is known.

Results: A general strategy for docking and modelling RNA–protein complexes has been developed. The docking procedure involves minimizing electrostatic and van der Waals' interaction energies of conformationally rigid structures during docking. After docking, libraries of amino acid sidechain conformations are searched to obtain the best interactions between the peptide and the RNA. Using this method, we have reproduced the structure of a bovine immunodeficiency virus (BIV) Tat peptide bound to BIV TAR RNA and have developed a model for the structure of the arginine-rich HIV-1 Rev peptide ($Rev_{34–50}$) interacting with the Rev-binding element (RBE).

Conclusions: The resulting model of the $Rev_{34–50}$–RBE complex predicts that although no single arginine sidechain is responsible for complex formation, residues Arg2, Arg5 and Arg11 are more important for binding than the other arginine residues in the peptide. One model is supported by binding measurements performed on wild-type and mutant RBE molecules with the peptide.

Addresses: [1]Département de Biochimie, Université de Montréal CP 6128, succursale Centre-ville Montréal, Québec H3C 3J7, Canada. [2]Present address: Department of Molecular Biology, The Scripps Research Institute, La Jolla, California 92037, USA. [3]Department of Chemistry, Indiana University, Bloomington, Indiana 47405, USA.

Correspondence: Robert Cedergren
e-mail: ceder@bch.umontreal.ca

Key words: docking, electrostatics, hydrogen bonding, RBE RNA, Rev peptide

Received: 30 Sep 1996
Revisions requested: 17 Oct 1996
Revisions received: 04 Nov 1996
Accepted: 11 Nov 1996

Published: 25 Nov 1996
Electronic identifier: 1359-0278-001-00463

Folding & Design 25 Nov 1996, 1:463–472

Introduction

The three-dimensional modelling of a macromolecule, such as RNA, is not yet a straightforward exercise, and the modelling of complexes involving proteins and RNA is subject to a number of additional complications. The docking problem for proteins has been widely recognized and innovative methods have been developed to generate realistic models of interactions between polypeptides and ligands [1,2]. Interactions between nucleic acids and proteins, however, have so far proven more difficult to model. In fact, attempts to model nucleic acid complexes have previously been largely restricted to DNA–drug complexes [3,4]. The number of nucleic acid–protein complexes whose structures have been solved to atomic resolution has been limited and, until recently, there were few examples to guide model-building exercises. In addition, less is known about the mutual adjustments that are required to maximize interactions between inherently deformable RNA and protein molecules.

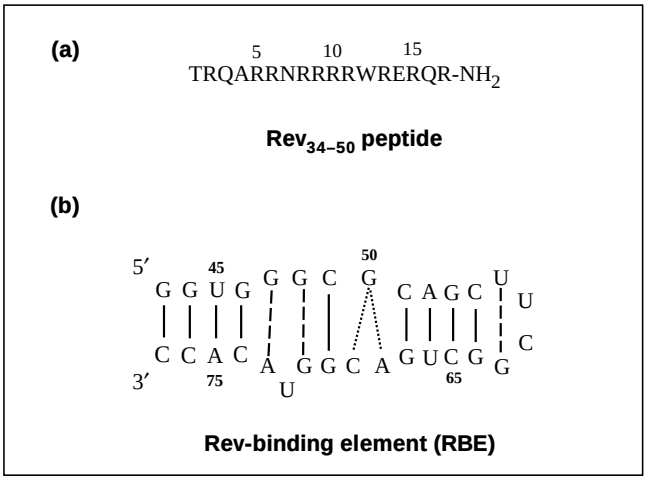
Recent work on the structure of peptide–RNA complexes has provided critical information for directing model building (reviewed in [5]). Peptides derived from the bovine immunodeficiency virus (BIV) Tat protein can bind specifically to the BIV TAR element. The structures of two BIV Tat peptide–TAR complexes have been solved by multidimensional NMR analysis [6,7] and provide a convenient control to assess methods for docking and modelling. Similarly, a 17 amino acid peptide derived from residues 34–50 of the human immunodeficiency virus Rev protein ($Rev_{34–50}$; Fig. 1a) can specifically bind to a 30 nucleotide RNA, the Rev-binding element (RBE [8–10]; Fig. 1b). We have previously modelled the structure of the RBE [11], and NMR analyses confirmed key aspects of our prediction [12,13]. A unique pattern of canonical and noncanonical base pairs [14–16] is thought to substantially widen the major groove of the RBE and permit access of amino acids in the Rev peptide to normally unapproachable nucleotide functional groups in the RNA [5,11]. However, at the time that our study was performed, the precise structure of the Rev–RBE complex was not known, and this system therefore provided a means to assess whether our methods could accurately predict interactions between peptides and nucleic acids. We present in detail a strategy for the docking and modelling of the complex between the Rev peptide and the RBE.

Results

The modelling process

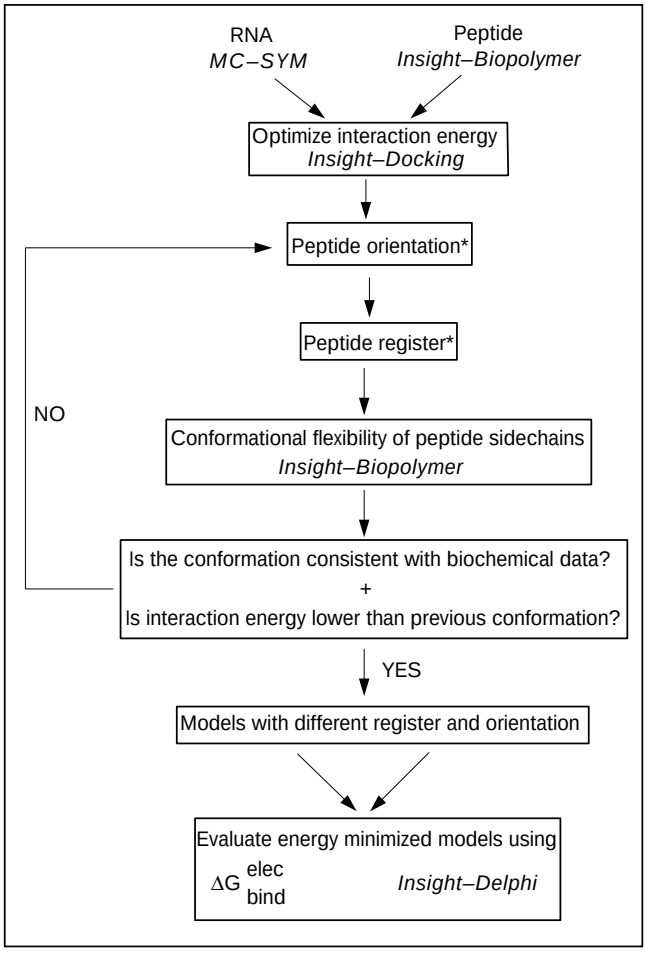
Computational methods used to construct molecular complexes involve the search for optimal intermolecular geometry and interaction energy. A rigorous search in geometric

Figure 1



The sequences of the peptide and RNA involved in the Rev–RBE complex. (a) Sequence of the $Rev_{34–50}$ peptide (in the one-letter amino acid code). The C-terminal modification to form a neutral amide group was based on the fact that this group increased the helicity of the sequence compared to the unmodified sequence (AD Ellington, unpublished data). (b) Primary sequence and secondary structure of the Rev-binding element (RBE) used in modelling the three-dimensional structure by the constraint satisfaction algorithm MC-SYM [11,35,36]. The solid lines indicate Watson–Crick base pairs, dashed lines indicate noncanonical base pairs and the dotted lines indicate alternative base pairing schemes. The nucleotide numbering convention adopted here is the same as in Leclerc *et al.* [11].

Figure 2



Flowchart for the docking procedure employed for modelling peptide–RNA complexes. The programs used at each step are indicated in italics, and steps involving manual docking are designated by an asterisk.

space involves a large number of trajectories, each of which is a rugged energy surface where small changes in intermolecular distance and orientation can result in large changes in interaction energy. To streamline this process, fully automated strategies to model protein–protein complexes have been developed based on either a geometric method, i.e. a search for complementarity between the ligand and substrate binding surfaces, or an energy-driven method, i.e. minimization of their interaction energy [1,2]. Unfortunately, these methods have yet to address nucleic acid–protein interactions. Thus, we have established an obvious but robust hierarchy for energetic analysis and have used this hierarchy to develop a generally useful docking algorithm. First, the structures of the individual molecules are energetically minimized independently and then treated as rigid bodies during the docking procedure. Next, the molecules are brought into contact along a small set of trajectories dictated by experimental data. Since nucleic acids are predominantly negatively charged while the proteins interacting with them are frequently positively charged, their interaction energies are largely electrostatic. Thus, the trajectory that gives the best initial minimization of electrostatic and van der Waals' interactions is retained. Finally, the juxtapositions of the partners and the positions of sidechains along the most favourable trajectory are extensively analyzed. The steps involved in modelling a test complex between BIV Tat peptide and TAR RNA and a peptide from the HIV-1 Rev protein, $Rev_{34–50}$, and the RBE are shown in Figure 2.

The BIV Tat–TAR complex

In order to validate the docking algorithm, we first applied it to a known structure, the BIV Tat–TAR complex [6]. Specifically, we have tested the ability of the method to identify the correct orientation (5′,3′ relative to N and C termini) and register (juxtaposition) of the Tat peptide within the RNA recognition site and evaluated its sensitivity to small changes in peptide register within the recognition site. In generating these models for the complex, conformations of both partners were fixed. Starting from the published NMR structure of the complex (designated R), three other models were constructed. The inverted model I was obtained by the rotation of the original peptide axis 180° about its midpoint while the position

of the RNA was frozen. A second model, R-C2, representing the case of a complex with the proper orientation but an incorrect register was constructed by translating the peptide from its original position one base pair step along the recognition site towards the 5′,3′ end of the TAR RNA. A third model, R-C1, was obtained by a 0.5 Å translation from its original position in the direction of the loop of the RNA. Calculation of the free energy of binding for different models confirmed that the free energy of the complex structure determined by NMR was indeed the lowest (Table 1). Judging from the free energy of binding calculated for the model R-C1, the method would be rather insensitive to small perturbations in peptide register. However, these calculations do demonstrate the validity of using the free energy of binding as a metric for determining the global register and orientation of peptides docked with RNA.

Table 1

The free energy of binding determined from the electrostatic energies of the different BIV Tat–TAR complexes and the corresponding RNA and peptide molecules.

Model	$\Delta G_{binding}$ (kcal mol^{-1})
R	–308.5
R-C1	–302.7
R-C2	3.1
I	15.0

The NMR structure of the complex is referred to as R. Model R-C1 has the peptide shifted about 0.5 Å towards the loop. Model R-C2 has the same peptide orientation as R but has been shifted by a two base pair step along the RNA axis towards the 5′,3′ end of the TAR RNA. Model I (for inverted) features the peptide in the orientation obtained by a 180° rotation about the perpendicular of the helical axis.

Docking the Rev peptide and the RBE

Having validated our method, we wished to test it on a previously unknown structure, a complex between the HIV-1 Rev peptide ($Rev_{34–50}$) and its cognate RNA, the RBE. One of the benefits of our procedure is that the precise geometry of each of the molecules within the complex need not be known in advance. Since the original models for the isolated molecules were generated under the conditions of a complex formed *in vivo*, the degree to which they can be brought together to form a tight complex is an independent measure of their accuracy. The $Rev_{34–50}$ peptide was initially modelled as a canonical α-helix, consistent with results from recent binding and circular dichroism studies [17–19]. Starting structures for the RBE RNA were taken from our previous modelling study [11] and differed only in the orientation of residue 71 — structure ‘A’ had G71 in the *anti* conformation, whereas structure ‘S’ had G71 in the *syn* conformation. They differed from each other by an all-atom root mean square deviation (RMSD) of 2.6 Å, but were indistinguishable on the basis of their electrostatic potential surfaces.

Docking was guided by the electrostatic potential surfaces (Fig. 3a,b) and published experimental data [10,14–18,20–22]. Amino acid substitution experiments had identified arginine residues in the $Rev_{34–50}$ peptide that were critical for interaction with the RBE [10,17,21]. These arginines, Arg2, Arg5, Arg6, Arg9 and Arg13, all lay along one side of an α-helical wheel (Fig. 3c). *In vitro* selection experiments [14,16] and modification interference analysis [10,15,22] had both shown that the major groove of the RBE internal loop was the site of peptide binding (Fig. 3b). Initial modelling therefore focused on fitting one face of the peptide to the major groove of the RNA.

The electrostatic potential surfaces of the RNA and the peptide were determined. The two molecules were brought together under the criterion of minimizing both van der Waals’ and coulombic energies of the nascent complex. Two different orientations for the peptide were considered during the docking procedure: one in which the N-terminal end of the peptide pointed towards the open end of the RBE (‘NO’) and the other in which it pointed towards the tetraloop region of RBE (‘NL’). In conjunction with the two forms of the RNA, ‘A’ and ‘S’, four different initial models were generated: A–NL, A–NO, S–NL and S–NO.

Identifying the register of the peptide

The same general techniques that were initially developed to assess the BIV Tat–TAR structure served to optimize the HIV-1 Rev–RBE models. The peptide was translated and rotated along its helical axis relative to the major groove of the RBE. At each local energy minimum, conformations for the amino acid sidechains on the side facing the RNA were additionally searched in the case of the RBE complex in order to promote interactions between the RNA and the peptide. If the resulting model was consistent with data from alanine mutation studies of the peptide [17,21] and chemical protection studies of the RBE available in the literature [10,15,20], the electrostatic and van der Waals’ interaction energy of the complex was then evaluated. Selected models underwent energy minimization (see Materials and methods). The binding energies for different models of the $Rev_{34–50}$–RBE complexes are shown in Table 2.

The lowest energy complex is formed when the N terminus of the $Rev_{34–50}$ peptide points towards the UUCG tetraloop region of the RBE (NL orientation) for both the ‘A’ and ‘S’ starting structures of the RBE. For the NL orientation of the peptide, two different registers were consistent with the biochemical data. One juxtaposition is represented by the model A–NL in which the N-terminal residues Thr1 and Arg2 of the peptide are aligned with

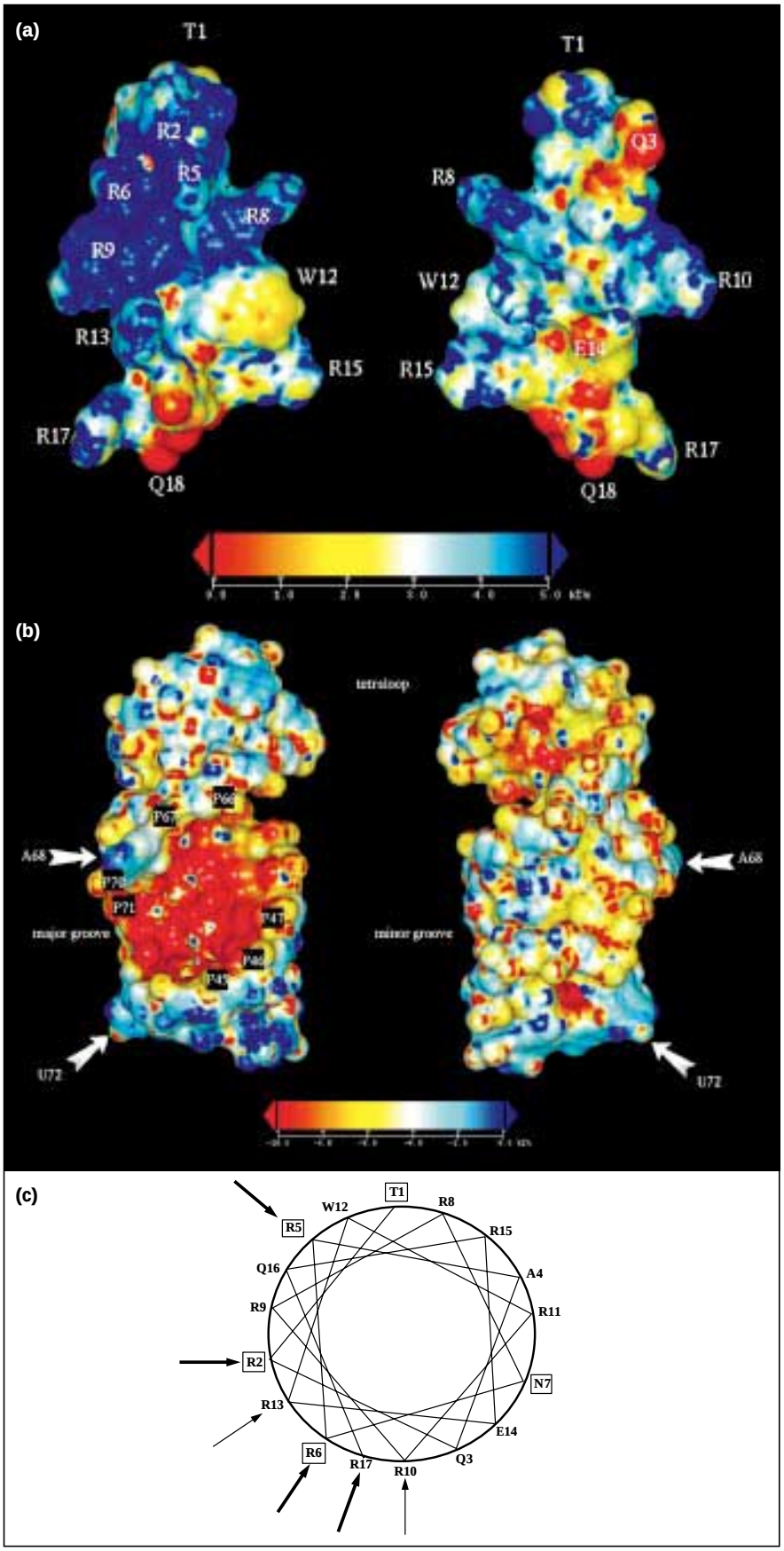


Figure 3

Electrostatic surfaces of the Rev peptide and RBE RNA. (a) Front and back views of the electrostatic surface of the $Rev_{34–50}$ peptide. Amino acid residues referred to in the text are shown (in the one-letter amino acid code). (b) Electrostatic surface of the RBE viewed into the major and minor grooves. Phosphate groups that are protected from chemical modification due to complex formation are indicated. (c) Helical wheel projection for the $Rev_{34–50}$ peptide. Arginine residues cleaved strongly (thick arrows) and weakly (thin arrows) during proteolytic cleavage studies are shown [20]. Amino acid residues whose substitution by alanine lead to peptides that do not form complexes are boxed.

Table 2

The free energy of binding determined from the electrostatic energies of the complexes and the corresponding RBE and peptide molecules using the finite difference method to solve the Poisson–Boltzmann equation.

Model	$\Delta G_{binding}$ (kcal mol^{-1})
A-NL	–71.6
A–NL_a	–21.6
S–NL	–46.0
A–NO	–9.5
S–NO	–24.0

The peptide orientation in which the N terminus points towards the tetraloop of the RBE structure is referred to as NL and that in which the N terminus points towards the 5′,3′ open end of the RBE (a 180° rotation about an axis perpendicular to the helical axis) is referred to as NO. The RBE structures 'A' and 'S' are the structures in which the nucleotide G71 is in the *anti* and *syn* conformation, respectively. The A–NL_a model has the peptide shifted two base pairs toward the tetraloop.

the base pair step G67–C51. The other juxtaposition is one in which the entire peptide is shifted two base pairs towards the tetraloop and thus Thr1 and Arg2 are aligned with the base pair step G64–C54 (represented as model A–NL_a). The calculated binding free energy suggested that the peptide register corresponding to model A–NL was preferred to that of model A–NL_a (Table 2). Further, the *anti* form of the RBE has a lower energy than the *syn* form. This finding corroborates both NMR studies and our earlier conclusion that although the conformation of the nucleotide G71 is flexible, as suggested both by NMR and theoretical studies [11–13,23], the *anti* conformation is preferred in the context of the complex [11].

Interactions between the peptide and the RBE

The best model for the complex, model A–NL, was further optimized by *in vacuo* molecular dynamics simulations with intermolecular and intramolecular hydrogen bonding constraints. The final structure was an average structure from the 300 K equilibrated trajectory (final 15 ps) of the dynamics simulation (Fig. 4a).

Based on a distance cut-off of 2.5 Å between atoms in contact, we postulate a series of interactions between the RBE and the $Rev_{34–50}$ peptide in model A–NL (Table 3). Specific interactions between the peptide sidechains and guanosine base moieties in the RBE provide a 'direct read-out' of the sequence, while nonspecific interactions between basic and polar peptide sidechains and the phosphodiester backbone of the RBE yield an 'indirect read-out' of structure. Hydrogen bonding patterns confirm that the noncanonical base pairs are preserved in the complex and are intimately involved in peptide recognition. In accord with binding studies with substituted peptides [17,20], most specific contacts to the RNA are made by arginine and asparagine, while nonspecific contacts are made by both arginine and other polar amino acid sidechains. The peptide backbone appears to be involved in only intramolecular interactions that would be important for stabilizing the helical structure.

Role of Arg2, Arg5, Arg11 and Asn7: direct read-out

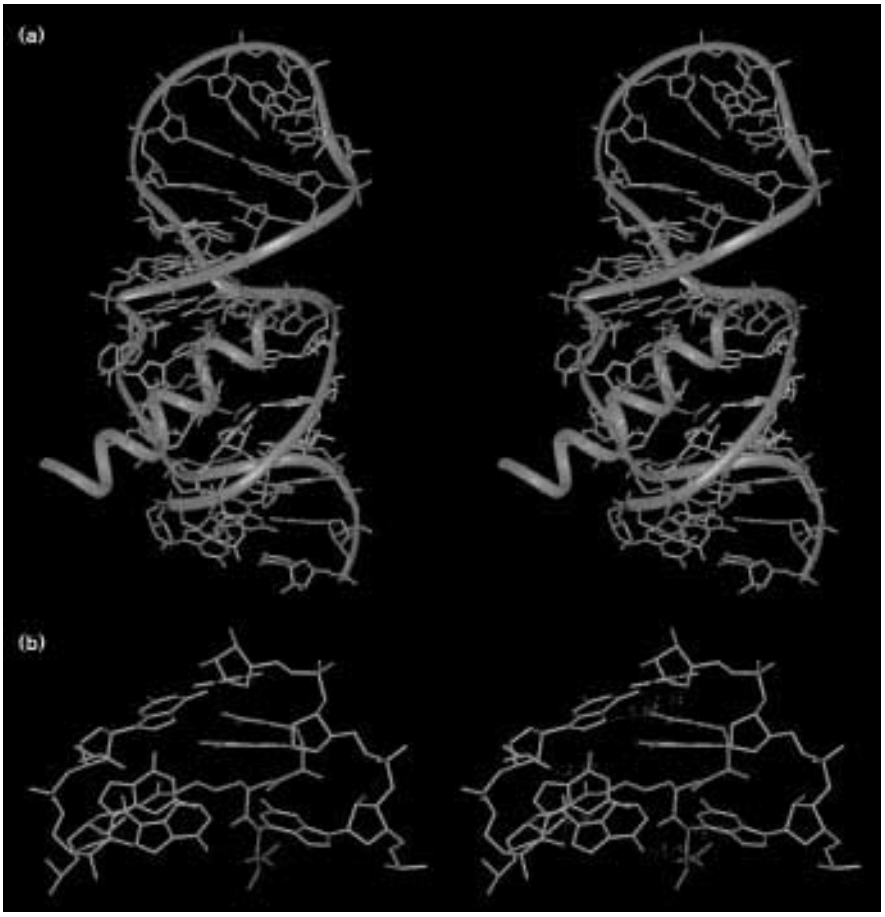
Model A–NL predicts specific interactions between residues Arg2, Arg5, and Arg11 of the peptide and the conserved nucleotides of the RBE internal loop (Table 3). The sidechain of Arg2 interacts specifically with the O6 of G67 (base paired to C51). Arg11 interacts with the N7 and phosphate of G46 (base paired to C74). The guanidino sidechain of Arg5 takes advantage of its multivalent hydrogen bonding capacity by making major groove contacts with both the O6 and N7 of G70 (Fig. 4b). It should be noted that G70 is a critical residue which is invariant in both the RBE and artificially selected sequences [16]. This pattern of interactions implies that there is no one critical arginine, but rather that the interactions of residues Arg2, Arg5 and Arg11 with the invariant residues of the RNA are vital to complex formation, as suggested previously from the results of binding studies with alanine substitutions of the Rev peptide or protein [17,24].

The hydrophilic residue Asn7 contacts the O6 of G47, which forms a noncanonical base pair with A73. In addition to this interaction, snapshot structures from the *in vacuo* dynamics simulations show that a long-range hydrogen bonding interaction between the carbonyl sidechain of Asn7 and the N6 amino group of A73 are possible. The G47–A73 noncanonical base pair is again invariant among the RBE and aptamer sequences [6]. Conversely, this interaction is absent in the complexes A–NL_a, A–NO and S–NO, reinforcing the conclusion that the modelled interaction between the peptide and the RBE is optimal.

Role of Thr1 and Arg6: indirect read-out

The sidechain hydroxyl group of Thr1 is predicted to hydrogen bond with the 3′-phosphate of G48. This portion of the model accords with previous suggestions that the non-Watson–Crick pair between positions 48 and 71 serves to position the phosphate backbone for precise recognition by the Rev peptide [16]. The threonine may additionally serve as an end-capping residue to stabilize the α-helical structure of the peptide, since binding affinities of short Rev peptides correlate to their percent helicity in solution [25]. The sidechain of arginine Arg6 also interacts with a phosphate group, in this case 3′ to the invariant guanosine residue G70. In addition to these interactions the three arginine residues, Arg8, Arg9, and Arg13, form electrostatic contacts with phosphates that lie along the major groove lining the RBE internal loop (Table 3). Finally, dynamics simulations indicate that the

Figure 4



Stereoview of model A-NL. (a) Stereoview of model A–NL of the $Rev_{34–50}$–RBE complex. The internal loop region of RBE (nucleotides G46, G47 and G48 on the right side and G70, G71 and A73 on the left side) is highlighted in yellow. Amino acids Arg2, Arg5, Arg6 and Arg11 are in green, Asn7 and Thr1 in orange, and Trp12 in violet. (b) Close-up stereoview of important hydrogen bonds. Nucleotides G47, G48 and C49 on the right side and G70, G71 and A73 on the left side are in yellow. Arg5 (on the top) is in purple as is Arg6 (on the bottom). Asn7 is in orange. Hydrogen bonds are indicated by dashed lines.

looped-out nucleotide A68 has a preference for stacking interactions with the tryptophan residue Trp12 of $Rev_{34–50}$.

Experimental data and the A–NL model

While some experimental data were used to guide the model-building process, one of the best tests of our model was the extent to which it was consistent with experimental data not used in model building. In this respect, model A–NL of the complex between $Rev_{34–50}$ and RBE can largely rationalize data from protection experiments and modification interference analyses. For example, residues Arg2, Arg5, and Arg6 are buried in the major groove of the RBE, thereby providing a rationale for their observed resistance to proteolysis [20]. Residue Arg13, on the other hand, is only partially enveloped by the RNA, which may explain its greater reactivity towards proteolysis. Model A–NL does not explain the inaccessibility of Arg17 [20], although a plausible reason for this discrepancy is that the experimental studies were performed on the complex of Rev and RRE whereas our models use minimal binding domains of the protein and the RNA.

Interference analyses with the RNA also support our model. The phosphate groups of nucleotides U45, G46, G47 and U66 cannot be ethylated [10] or converted to phosphonates [22] without loss of binding activity. In A–NL, the phosphate groups adjacent to U45, G46 and G47 are predicted to be involved in specific hydrogen bonding interactions with peptide sidechains, as indicated in Table 3. Further, the phosphate groups adjacent to G67 and U66 are involved in electrostatic interactions with Arg2. Similarly, many of the guanine bases within the internal loop, in particular those phosphate groups associated with residues G46, G67, G70 and G71, cannot be modified without loss of binding activity [10] and are predicted to be involved in interactions with peptide sidechains (Table 3).

In contrast, N7 and O6 of G48 can be chemically modified without loss of binding [15]. This result can be explained by the A–NL model, in which no direct interactions between G48 and the peptide are observed. Similarly, for both the ‘A’ and ‘S’ structures of the RBE, the N7 of G71 is located in the minor groove while that of G48 is in the

Table 3

The hydrogen bonding interactions between the residues of the RBE and the sidechains of the $Rev_{34–50}$ peptide predicted by the A–NL model.

	RBE groups	Peptide groups
1	G48:O1P	Thr1*
2	G67:O6	Arg2
3	G70:O6,N7; C69:O1P	Arg5
4	G70:O1P	Arg6
5	A73:N6	Asn7†
6	G47:O6	Asn7‡
7	G46,G47:O1P	Arg8
8	A68:O1P; C69:O2P	Arg9
9	U45:O1P; G46:O1P,N7	Arg11
10	G71:O1P	Arg13

*Hydroxyl group. †Carbonyl group. ‡Amide group.

major groove. The A–NL model predicts that this non-Watson–Crick base pair should not be involved in any direct interactions. This prediction is consistent with binding studies in which G48 and G71 were substituted with 7-deazaguanosine [15] and further buttresses the conclusion that the non-Watson–Crick pairing aids in establishing a matrix of negatively charged phosphates that are contacted by positively charged arginine residues.

Probing interactions between $Rev_{34–50}$ and the RBE with nucleotide substitutions

In order to experimentally verify the model A–NL, interactions between $Rev_{34–50}$ and several RBE sequence variants were assessed. Sequence substitutions in the RBE were chosen to probe the model because individual hydrogen bonds between the peptide sidechains and nucleotide bases could be assessed. In addition, the effects of the nucleotide substitutions on the structure could be quickly and easily modelled and the modelling results compared with experiment.

The model predicts that the O6 position of G67 interacts with the guanidino group of Arg2. A G67→A nucleotide substitution replaces O6, a hydrogen bonding acceptor, with an exocyclic amine, a hydrogen bond donor. As expected, binding of this RBE variant to $Rev_{34–50}$ is severely compromised (Table 4). Obviously, this result could be due to disruption of the C51–G67 base pair and a more global alteration of the RBE structure rather than the loss of a particular intermolecular hydrogen bond. Therefore, to minimize possible structural disruptions the double substitution G67→A; C51→U was also synthesized. This variant also failed to bind $Rev_{34–50}$, emphasizing that a critical sequence-specific interaction had been affected. Similar predictions were made and confirmed for interactions between Arg5 and RNA with single G70→A and the double G70→A; C49→U nucleotide substitutions.

Table 4

Summary of gel shifts measured with the $Rev_{34–50}$ peptide and the RBE mutants.

RNA	% Binding of $Rev_{34–50}$
RBE	100%
G46→A; C74→U	72%
G48→A	74%
G48→A; G71→A	110%
G48→A; G71→A; U72→C	100%
G67→A	16%
G67→A; C51→U	49%
G50→A; A68→C; C69→A	150%

The numbers represent the average percentages of RNA that had a modified mobility compared to the wild-type complex in duplicate experiments. The single mutants G70→A and G46→A and the double mutant G70→A; C49→U showed no shifts with $Rev_{34–50}$.

In contrast, residue G46 is predicted to interact with Arg11 of $Rev_{34–50}$ via its N7 position and with Arg8 via an adjacent phosphate. Interactions with a G46→A variant are thus expected to hinge on whether or not this substitution results in a conformational alteration of the RBE, rather than the disruption of a particular hydrogen bond. Modelling of the G46→A nucleotide variant suggests that the formation of an A–C pair between G46→A and C74 would modify the local backbone conformation of this base pair, resulting in a loss of hydrogen bonding to N7 and the adjacent phosphate. Reversal to a canonical Watson–Crick base pair should restore local geometry and hence binding activity. In agreement with our expectations, the G46→A single substitution variant cannot bind $Rev_{34–50}$, whereas the G46→A; C74→U double substitution has more than half the binding activity of the wild-type RBE (Table 4). This pattern of suppression is consistent with our model, but contrasts with that seen for G70 and G67, where the lack of the guanosine O6 atom cannot be corrected by a double nucleotide substitution.

Not all nucleotide substitutions compromise binding activity. The triple nucleotide substitution G50→A; A68→C; C69→A, originally identified by *in vitro* selection [16], shows improved binding activity both *in vitro* and *in vivo*. In accord with these findings, the Rev peptide also binds to this triple substitution (Table 4). Also, the series G48→A, G48→A; G71→A and G48→A; G71→A; U72→C do not appreciably affect binding. These results again accord with the fact that none of these residues is predicted to be involved in base-specific interactions with the peptide.

Conclusions

We have used computer docking and energy minimization procedures to model the molecular recognition between a peptide from BIV Tat and the Tat-responsive element, and between the $Rev_{34–50}$ peptide and the Rev-binding element. In so doing, we have established an energy criterion to evaluate the orientation and register of peptide–RNA complexes. While this approach is not able to distinguish between closely related structures and may not accurately predict the precise conformation of either the isolated RNA or the isolated peptide, it does provide experimentally verified insights into base-specific and backbone-specific interactions.

As observed in our earlier study [11], a wide major groove forced open by noncanonical base pairs in the internal loop of RBE (Figs 1,3) can accommodate the α-helix with amino acids having long sidechains (Fig. 4). In addition, functional groups presented in the wide major groove, and correspondingly shielded in the regular A-form RNA helix, serve as sites for specific hydrogen bonding to the peptide. Our model is consistent with protection and interference data for the peptide–RNA complex and roughly predicts the relative strength of binding interactions between the peptide and variants of the RBE.

The model that we propose for the $Rev_{34–50}$–RBE complex was constructed solely with the information attributed in this paper. In spite of our confidence in this model, we must recognize the fact that it is based to a large extent on the forcefield parameters for molecular simulation procedures and the conditions used for the inclusion of solvent and salt effects which are well known to be approximate.

During the final preparations of this manuscript, an NMR-generated structure for a Rev peptide–RBE complex appeared [26]. Although the primary sequence of both RNA and peptide are slightly different from that in this report, the overall peptide orientation and register predicted by our model are identical to that in the NMR structure. Specific interactions between the asparagine and arginine residues in the NMR structure are duplicated in our model, i.e. the contacts between Asn7 and the two nucleotides G47 and A73 and between the residue Arg2 and nucleotide G67. In addition, most phosphate backbone contacts that we have predicted are in good agreement with the NMR structure, while the prediction of the van der Waals' contacts are, perhaps expectedly, less successful. Both represent the peptide conformation as close to an ideal α-helix. Encouraged by the global and local structural similarity between these two structures, we are presently assessing the robustness of our model through a detailed comparison of local interactions between the RNA and peptide against those determined by this NMR structure.

Materials and methods

Materials

The listed sequences represent the nontemplate strand of a double-stranded DNA oligomer; all of these sequences were fused to a T7 RNA polymerase promoter (5′-GGTAATACGACTCACTATA-3′), the position of which is indicated by an X in the following sequences. The underlined positions represent substitutions relative to the wild-type RBE.

RBE	5′-XGGGTGGGCGCACGTTCGCGTGACGGTA-CACC-3′
G46→A	5′-XGGGTAGGCGCACGTTCGCGTGACGGTA-CACC-3′
G46→A;C74→T	5′-XGGGTAGGCGCACGTTCGCGTGACGGTAT-ACC-3′
G50→A;A68→C;C69→A	5′-XGGGTGGGCACACGTTCGCGTGCAGGTA-CACC-3′
G70→A	5′-XGGGTGGGCGCACGTTCGCGTGACAGTA-CACC-3′
C49→T;G70→A	5′-XGGGTGGGTGCACGTTCGCGTGACAGTA-CACC-3′
G48→A	5′-XGGGTGGACGCACGTTCGCGTGACGGTA-CACC-3′
G48→A;G71→A	5′-XGGGTGGACGCACGTTCGCGTGACGATA-CACC-3′
G48→A;G71→A;T72→C	5′-XGGGTGGACGCACGTTCGCGTGACGACA-CACC-3′
G67→A	5′-XGGGTGGGCGCACGTTCGCGTAACGGTA-CACC-3′
G67→A;C51→T	5′-XGGGTGGGCGTACGTTCGCGTAACGGTA-CACC-3′

The sequence of the $Rev_{34–50}$ peptide was NH_2-RQARRNRRRRWR-ERQR-$CONH_2$ (carboxamide). The peptide was prepared as previously described by Xu and Ellington [27] and its purity was determined by HPLC and mass spectrometry. The amount of peptide to be used in gel-shift reactions was determined by HPLC analysis of dissolved samples.

Docking and simulations

The initial structure of the $Rev_{34–50}$ peptide was the canonical α-helical form generated by the MSI graphics software INSIGHT II (version 2.3.5, MSI Inc, San Diego, CA). The C terminus of the peptide was modified to a carboxamide group. Amino acid sidechains were initially represented in their extended conformations, and arginine and glutamic acid sidechains were represented in their fully charged forms. The RBE structure 'A' (PDB ID 163D) was generated by energy optimization of the RBE model previously derived using constraint satisfaction program MC-SYM [11]. In the 'A' structure, nucleotide A68 was bulged out of the RNA helix and the nucleotide G71 was in the *anti* conformation. A similar RBE structure, 'S', was obtained by energy optimization of the RBE model using the constraint satisfaction approach but specified a *syn* conformation for the G71 nucleotide. It should be noted that this conformation was also observed during the dynamics simulations of 'A'. As stated in the text, the two initial RBE structures were used to evaluate the conformational preference of the G71 nucleotide in the context of the complex.

The electrostatic potential around the RBE was calculated using the finite difference solution to the nonlinear Poisson–Boltzmann equation implemented in the Delphi program of the MSI software package. The dielectric constant of solvent (represented as a continuum model) and the RNA were included in these calculations. Using the electrostatic

potential map around the RBE, models for the peptide–RNA complex were constructed using the MSI software module DOCKING. Electrostatic and van der Waals' interactions between the peptide and the RNA were minimized while orienting the peptide in the field of the RNA. From several possible juxtapositions of the peptide with respect to the RNA, preferred conformations were selected based first on a low nonbonded interaction energy and then on an optimum electrostatics energy for the entire assembly. The models R-C1, R-C2 and I for the BIV Tat–TAR peptide–RNA complex were derived in the same way starting with the coordinates of the complex [6] provided to us by JD Puglisi (University of California at Santa Cruz).

The docked assemblies, A–NL, A–NO, S–NL and S–NO, derived from the initial RBE structures 'A' and 'S' and the $Rev_{34–50}$ peptide were subsequently optimized using the energy minimization procedure in the molecular mechanics package Discover (version 2.96, MSI Inc) interfaced with the AMBER 2.0 forcefield [28,29]. Energy minimizations were performed *in vacuo* using the conjugate gradient minimizer and a convergence criterion of 0.5 kcal mol^{-1} $Å^{-2}$ maximum derivative. A distance-dependent dielectric constant, $\epsilon = 4r$, was used to simulate the screening effects of the solvent and a nonbonded cut-off distance of 15 Å was used in all energy calculations. Distance constraints were imposed on all base-paired hydrogen bonding atoms within the RNA and on atom pairs within the UUCG tetraloop based on its NMR structure [29]. The distance constraint was a harmonic potential function with a force constant of 20 kcal mol^{-1} $Å^{-2}$.

In order to improve interactions between the peptide and the RNA, *in vacuo* dynamics simulations of the minimized form of the complex model A–NL were performed using the Discover package and the AMBER forcefield parameters. The minimized complex was gradually heated to 300 K in steps of 50 K over 500 steps with a timestep of 1 fs. The dielectric constant, nonbonded cut-offs and distance constraints used were the same as in the minimization calculations. In addition to the hydrogen bonding constraints within the RNA, weak harmonic distance restraints between the peptide and RNA were enforced with a force constant of 10 kcal mol^{-1} $Å^{-2}$. A 20 ps trajectory of the complex at 300 K was run during which coordinates were stored every 0.5 ps. A single average structure of the last 15 ps of the trajectory was used to identify interactions between the $Rev_{34–50}$ peptide and the RBE.

Free energy calculations

The total electrostatic free energy of a molecule is represented as the sum of the two contributions:

$$\Delta G = \Delta G^{coul} + \Delta G^{RF} \quad (1)$$

where ΔG^{coul} is the coulombic energy of all the charges in the molecule and ΔG^{RF} is the reaction field energy, i.e. the energy arising from polarizing the environment of the molecule [31]. These two components of the electrostatic energy were calculated for the starting conformations of RBE, 'A' and 'S', and for the $Rev_{34–50}$ peptide by solving their nonlinear Poisson–Boltzmann equation. Their overall electrostatic energies are referred to as ΔG_{RNA} and $\Delta G_{peptide}$, respectively. The electrostatic energies of the four $Rev_{34–50}$–RBE complexes were calculated in the same manner using the Delphi module [31,32] of the MSI software. All calculations involved the use of a 65^3 lattice and a 0.75 Å spacing between grid points for each of the molecules. The potential at the boundaries of the grid were calculated using the Debye–Hückel term with AMBER charges assigned to each atom in the molecule. In all calculations, the charges were treated as embedded in a low dielectric medium ($\epsilon = 2$) made up of the volume enclosed by the solvent-accessible surface of the molecule (calculated with a probe radius = 1.4 Å [33]). In each case, the solvent was treated as a continuum model ($\epsilon = 80$) with a 2.0 Å ion exclusion radius.

During docking, since the peptide approached the RNA from an infinite separation, the binding free energy of the models for the complex was calculated according to equation 2 [33,34]:

$$\Delta G_{binding} = \Delta G_{complex} - \{\Delta G_{RNA} + \Delta G_{peptide}\} \quad (2)$$

where ΔG is determined from the Delphi calculations according to equation 1. Since the binding process involved the association of two oppositely charged centres, we chose the electrostatic component of the binding free energy as a useful index for characterizing the probable juxtaposition of the peptide in the complex.

The electrostatic free energy of the original BIV Tat–TAR complex [6], different incorrect models and the uncomplexed TAR RNA and Tat peptide were calculated in a similar manner.

Generation of electrostatic surfaces

The electrostatic potential energy surfaces of $Rev_{34–50}$ and the RBE were generated with the INSIGHT module of the MSI software via potential energy grids that were in turn calculated based on the finite difference calculations mentioned above. Potential energy surfaces were mapped using a colour ramp.

Transcription

DNA oligomers and their complements were heated to 95°C for 5 min in 10 mM Tris, pH 7.4, 100 mM NaCl, 1 mM EDTA and allowed to anneal by cooling to room temperature over 30 min. Transcription reactions were carried out using an Ampliscribe kit (Epicentre Technologies, Madison, WI) according to the manufacturer's instructions; 1 μl of radiolabel (α-[^{32}P]UTP, 3000 Ci $mmol^{-1}$, 10 mCi ml^{-1}) was incorporated into each 10 μl reaction. Transcripts were purified on a denaturing polyacrylamide gel. The amount of RNA that was to be used in gel-shift reactions was determined by measuring the absorbance of the sample at 260 nm.

Gel shift

The ability of RNAs to bind to free peptide was assayed by gel-mobility shift. Radiolabelled RNA samples were thermally equilibrated in 50 mM HEPES, pH 7.4; 100 mM NaCl by heating to 75°C for 3 min and then cooling to ambient temperature over 5 min. RNAs (0.5 μg, 1.2 μM) were mixed with peptides (25 ng; 1.0 μM) in 10 μl total volume of 1 × Shift Buffer (10 mM HEPES, pH 7.4; 100 mM KCl; 1 mM DTT; 1 mM $MgCl_2$; 1 mM EDTA; 50 μg ml^{-1} tRNA; 5% glycerol). The binding reaction was incubated on ice for 30 min. Peptide–RNA complexes were separated from free RNA by electrophoresis on a 12% polyacrylamide gel (29:1 acrylamide : bisacrylamide; 1 × TBE) at 8 W for 3 h. The temperature of the gel remained at 4–7°C throughout the run. Radiolabelled bands were visualized and quantitated using a Phosphorimager (Molecular Dynamics, Sunnyvale, CA).

Acknowledgements

This work was supported by the Medical Research Council of Canada. R Cedergren is a Fellow of the Canadian Institute for Advanced Studies programme in evolutionary biology.

References

1. Strynadka, N.C.J., *et al*., & James, M.N.G. (1996). Molecular docking programs successfully predict the binding of a β-lactamase inhibitory protein to TEM-1 β-lactamase. *Nat. Struct. Biol.* 3, 233–239.
2. Kuntz, I.D., Meng, E.C. & Shoichet, B.K. (1994). Structure-based molecular design. *Acc. Chem. Res.* 27, 117–123.
3. Grootenhuis, P.D.J., Kollman, P.A., Seibel, G.L., DesJarlais, R.L. & Kuntz, I.D. (1990). Computerized selection of potential DNA-binding compounds. *Anti-Cancer Drug Des.* 5, 237–242.
4. Grootenhuis, P.D.J., Roe, D.C., Kollman, P.A, & Kuntz, I.D. (1994). Finding potential DNA-binding compounds by using molecular shape. *J. Comp. Mol. Des.* 8, 731–750).
5. Draper, D.E. (1995). Protein RNA recognition. *Annu. Rev. Biochem.* 64, 593–620.
6. Puglisi, J.D., Chen, L., Blanchard, S. & Frankel, A.D. (1995). Solution structure of a bovine immunodeficiency virus Tat–TAR peptide–RNA complex. *Science* 270, 1200–1203.
7. Ye, X., Kumar, R.A. & Patel, D.J. (1995). Molecular recognition in

bovine immunodeficiency virus Tat peptide–TAR RNA complex. *Chem. Biol.* 2, 827–840.
8. Zapp, M.L., Hope, T.J., Parslow, T.G. & Green M.R. (1991). Oligomerisation and RNA binding domains of the type 1 human immunodeficiency virus Rev protein: a dual function for an arginine-rich binding motif. *Proc. Natl. Acad. Sci. USA* 88, 7734–7738.
9. Cook, K.S., Fisk, G.J., Hauber, J., Usman, N., Daly, T.J. & Rusche, J.R. (1991). Characterization of HIV-1 Rev protein: binding stoichiometry and minimal RNA substrate. *Nucleic Acids Res.* 19, 1577–1583.
10. 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.
11. Leclerc, F., Cedergren, R. & Ellington, A.E. (1994). A three-dimensional model of the Rev-binding element of HIV-1 derived from analyses of aptamers. *Nat. Struct. Biol.* 1, 293–300.
12. Battiste, J.L., Tan, R., Frankel, A.D. & Williamson, J.R. (1995). Assignment and modelling of the Rev response element (RRE) RNA bound to a Rev peptide using 13C-heteronuclear NMR. *J. Biomol. NMR.* 6, 375–389.
13. Peterson, R.D., Bartel, D.P., Szostak, J.W., Horvath, S.J. & Feigon, J. (1994). 1H NMR studies of the high affinity Rev binding site of the Rev Responsive Element of HIV-1 mRNA: base pairing and core binding element. *Biochemistry* 33, 5357–5365.
14. 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.
15. Iwai, S., Pritchard, C., Mann, D.A., Karn, J. & Gait, M.J. (1992). Recognition of the high affinity binding site in Rev-responsive element RNA by the human immunodeficiency virus type-1 Rev protein. *Nucleic Acids Res.* 20, 6465–6472.
16. Giver, L., Bartel, D., Zapp, M., Pawul, A., Green, M. & Ellington, A.D. (1993). Selective optimization of the Rev-binding element of HIV-1. *Nucleic Acids Res.* 21, 5509–5516.
17. Tan, R., Chen, L., Buettner, J.A., Hudson, D. & Frankel, A.D. (1993). RNA recognition by an isolated α-helix. *Cell* 73, 1031–1040.
18. Daly, T.J., Doten, R.C., Rusche, J.R. & Auer, M. (1995). The amino terminal domain of HIV-1 Rev is required for discrimination of the RRE from nonspecific RNA. *J. Mol. Biol.* 253, 243–258.
19. Scanlon, M.J., Fairlie, D.P., Craik, D.J., Englebretsen, D.R. & West, M.L. (1995). NMR solution structure of the RNA-binding peptide from human immunodeficiency virus (type 1) Rev. *Biochemistry* 34, 8242–8249.
20. Jensen, T.H., Leffers, H. & Kjems, J. (1995). Intermolecular binding sites of human immunodeficiency virus type 1 Rev protein determined by protein footprinting. *J. Biol. Chem.* 270, 13777–13784.
21. Tan, R. & Frankel, A.D. (1994). Costabilization of peptide and RNA structure in an HIV Rev peptide–RRE complex. *Biochemistry* 33, 14579–14585.
22. Pritchard, C.E., *et al.*, & Gait, M.J. (1994). Methylphosphonate mapping of phosphate contacts critical for RNA recognition by the human immunodeficiency virus tat and rev proteins. *Nucleic Acids Res.* 22, 2592–2600.
23. Shu-Yun, L., Pattabiraman, N. & Maizel, J.V., Jr.(1994). RNA tertiary structure of the HIV RRE domain II containing non-Watson–Crick base pairs G:G and G:A: molecular modelling studies. *Nucleic Acids Res.* 22, 3966–3976.
24. Hammerschmid, M., *et al.*, & Hauber, J. (1994). Scanning mutagenesis of the arginine-rich regions of the human immunodeficiency virus type 1 rev *trans* activator. *J. Virol.* 68, 7329–7335.
25. Doig, A.J. & Baldwin, R.L. (1995). N- and C-capping preferences for all 20 amino acids in alpha-helical peptides. *Proteins Sci.* 4, 1325–1336.
26. Battiste, J.L., *et al.*, & Williamson, J.R. (1996). α-Helix–RNA major groove recognition in an HIV-1 Rev peptide–RRE RNA complex. *Science* 273, 1547–1551.
27. Xu, W. & Ellington, A.E. (1996). Anti-peptide aptamers recognize amino acid sequence and bind a protein epitope. *Proc. Natl. Acad. Sci. USA* 93, 7475–7480.
28. Weiner, S.J., *et al.*, & Weiner, P. (1984). A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* 106, 765–784.
29. Weiner, S.J., Kollman, P.A., Nguyen, D.T. & Case, D.A. (1986). An all atom force field for simulations of proteins and nucleic acids. *J. Comp. Chem.* 7, 230–252.
30. Cheong, C., Varani, G. & Tinoco, I. (1990). Solution structure of an unusually stable RNA hairpin, 5′GGAC(UUCG)GUCC. *Nature* 346, 680–682.
31. Klapper, I., Hägstrom, R., Fine, R., Sharp, K. & Honig, B. (1986). Focussing of electric field in the active site of Cu-Zn superoxide dismutase: effect of the ionic strength and amino acid modification. *Proteins* 1, 47–59.
32. Gilson, M.K. & Honig, B.H. (1988). Energetics of charge–charge interactions in proteins. *Proteins* 3, 32–53.
33. Zacharias, M., Luty, B.A., Davis, M.E. & McCammon, A.J. (1992). Poisson–Boltzmann analysis of the λ repressor–operator interaction. *Biophys. J.* 63, 1280–1285.
34. Jain, A. & Murcko, M.A. (1995). Computational methods to predict binding free energy in ligand–receptor complexes. *J. Med. Chem.* 38, 4953–4967.
35. Major, F., Turcotte, M., Gautheret, D., Lapalme, G., Fillion, E. & Cedergren, R. (1991). The combination of symbolic and numerical computation for three dimensional modelling of RNA. *Science* 253, 1255–1260.
36. Gautheret, D., Major, F. & Cedergren, R. (1993). Modelling the three dimensional structure of RNA using discrete nucleotide conformational sets. *J. Mol. Biol.* 229, 1049–1064.

Because *Folding & Design* operates a 'Continuous Publication System' for Research Papers, this paper has been published via the internet before being printed. The paper can be accessed from http://biomednet.com/cbiology/fad.htm – for further information, see the explanation on the contents page.