Theor Chem Acc (1999) 101:131–137
DOI 10.1007/s002149800m85

Theoretical
Chemistry Accounts

Regular article

MCSS-based predictions of RNA binding sites*

Fabrice Leclerc[1], **Martin Karplus**[1,2]

[1] Harvard University, Department of Chemistry and Chemical Biology, 12 Oxford Street, Cambridge, MA 02138, USA
[2] Université de Strasbourg I, Institut le Bel, Laboratoire de Chimie Biophysique, F-67000 Strasbourg, France

Received: 24 April 1998 / Accepted: 4 August 1998 / Published online: 7 December 1998

Abstract. The diversity of RNA tertiary structures provides the basis for specific recognition by proteins or small molecules. To investigate the structural basis and the energetics which control RNA-ligand interactions, favorable RNA binding sites are identified using the MCSS method, which has been employed previously only for protein receptors. Two different RNAs for which the structures have been determined by NMR spectroscopy were examined: two structures of the TAR RNA which contains an arginine binding site, and the structure of the 16S rRNA which contains an aminoglycoside binding site (paromomycin). In accord with the MCSS methodology, the functional groups representing the entire ligand or only part of it (one residue in the case of the aminoglycosides) are first replicated and distributed with random positions and orientations around the target and then energy minimized in the force field of the target RNA. The Coulombic term and the dielectric constant of the force field are adjusted to approximate the effects of solvent-screening and counterions. Optimal force field parameters are determined to reproduce the binding mode of arginine to the TAR RNA. The more favorable binding sites for each residue of the aminoglycoside ligands are then calculated and compared with the binding sites observed experimentally. The predictability of the method is evaluated and refinements are proposed to improve its accuracy.

Key words: MCSS – Ligand Design – RNA – Binding

1 Introduction

Nucleic acids make logical targets for drug design, since all enzymes and receptor proteins depend on RNA for their synthesis. Unlike DNA, RNA may fold into complex and diverse molecular shapes which constitute attractive targets for specific and selective binding. A number of X-ray or NMR structures of RNA and its complexes with drugs, peptides, or proteins have been recently determined and provide the opportunity to better understand RNA-ligand interactions [6]. Nevertheless, only a few studies on modeling RNA-ligand interactions have been published [7, 8].

A primary step towards the design of drugs against macromolecule receptors is the identification of potential binding sites for ligand fragments (functional groups). Such an approach has been applied in computational and laboratory combinatorial ligand design for protein receptors [2, 9–11]. The TAR RNA of HIV-1 and the bacterial 16S rRNA, for which the structures have been determined by NMR spectroscopy [3–5], represent two interesting targets because of their biological role in the regulation of the HIV cycle and in the protein synthesis of pathogenic bacteria, respectively. They are used here as macromolecule targets with the MCSS method to predict favorable binding sites for arginine in the case of the TAR RNA and for aminoglycoside moieties in the case of the 16S rRNA. The MCSS method [1] is divided into two steps. The first step involves the replication of pre-defined functional groups and their distribution in random positions and orientations in a binding region delimited by a spherical or rectangular boundary. The second step involves their energy minimization in the force field of the receptor [1], and the selection of the local minima based on an energy cutoff. The search for minima is performed by an iterative process. Details of the method are given by Miranker and Karplus [1].

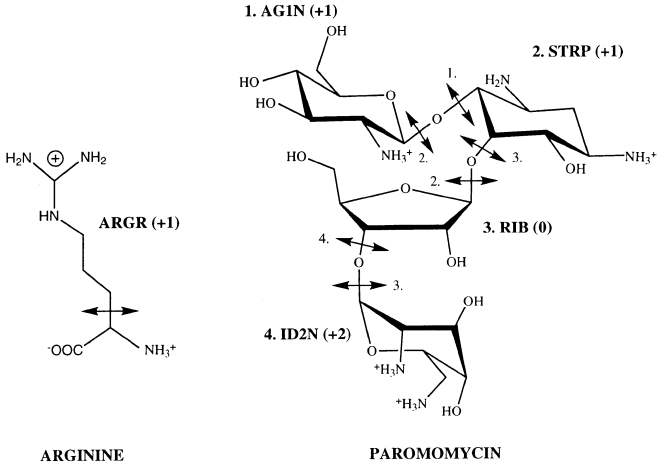
2 Specific approach

Two sets of force field parameters were used to identify and select the MCSS minima. In the first set (Model 1), the non-bonded interactions between the replicated functional groups (replica) and the receptor were calculated based on the CHARMM force field using the recent parameters for nucleic acids [12]. Since these parameters are designed for use with explicit solvents

* Contribution to the Proceedings of Computational Chemistry and the Living World, April 20–24, 1998, Chambery, France

Correspondence to: M. Karplus
e-mail: marci@brel.u-shasbg.fr

and the calculations are done in vacuo for the sake of efficiency, a modified set (Model 2) was derived from the first one by scaling down the atomic charges on the phosphate groups to mimic the presence of counterions [13]. In both sets, the solvent screening is modeled using a distance-dependent dielectric $\epsilon(r) = (c \cdot r)$ where c varies from 1 to 4. The search for MCSS minima is first carried out in a restricted binding region defined by a sphere of 12 Å of radius centered on the position of the bound arginine residue. A more extensive search is then performed in a box including the whole RNA. In the case of the 16S rRNA, the first search is carried in a box centered on the aminoglycoside binding site (8250 $Å^3$). The MCSS functional group for the arginine side-chain is represented by a polar hydrogen model [1, 14]. The four MCSS functional groups corresponding to the individual residues of paromomycin are represented with the all-hydrogen model. The chemical structures of the RNA ligands, arginine, and paromomycin are shown in Scheme 1; the functional groups are separated by arrows (the number attached to the arrow indicates where the bond is broken for the corresponding residue) and the net charge per group is indicated in parentheses.



Scheme 1.

The minima are sorted based on the energy of interaction, which is defined by the sum of the van der Waals and electrostatic contributions. The improved MCSS program (version 2.1) developed by Erik Evensen (unpublished), a complete reimplementation of the original MCSS method with increased efficiency, flexibility, and ease-of-use, and the CHARMM program [15] are used for the calculations.

3 Results

The strong arginine binding site of the TAR RNA is used to calibrate the method by adjusting the force field parameters to account for the implicit presence of counterions and/or the screening effect of solvent. The two classes of model were tested to reproduce the position of the arginine side-chain in the TAR RNA binding site. For Model 1 the MCSS minima are closer to the phosphate backbone, while they are closer to the nucleic acid bases for the Model 2. In both models, the increase of the dielectric constant produces more favorable MCSS minima in the known arginine binding site. Model 2, which takes into account both the implicit presence of counterions and the screening effect of solvent ($c = 3$), gave the minima with the best score and the lowest root mean square deviation (RMSD) with respect to the position of the arginine residue in the NMR structure (see Table 1). In what follows, we describe the results for Model 2.

The results are presented in Fig. 1: the MCSS minima identified in a restricted searched region around the arginine binding site (Fig. 1A) and on the entire RNA surface (Fig. 1B). The minima exhibit three different binding modes: the first one involves arginine fork-like structures with nucleic acid bases (guanine) [4], the second one corresponds to non-specific interactions with phosphate groups, and the last one to stacking interactions with the base of bulge nucleotides

Table 1. Prediction accuracy of RNA binding sites[a]

RNA	Ligand	MCSS functional group	Model 1			Model 2		
			RMSD	Score	Occupancy[b]	RMSD	Score	Occupancy
HIV-1 TAR[c]	Arginine	Arginine side-chain	0.6 Å	100	7%	0.6 Å	100	8%
HIV-1 TAR[d]	Arginine	Arginine side-chain	1.1 Å	100	6%	1.0 Å	100	15%
16S rRNA	Paromomycin	1st residue (AG1N)	1.0 Å	65	1%	1.0 Å	76	3%
	Paromomycin	2nd residue (STRP)	≥4	≤1	–	2.8 Å	1.2	–
	Paromomycin	3rd residue (RIB)	≥4	≤1	–	≥4	≤1	–
	Paromomycin	4th residue (ID2N)	0.3 Å	74	2%	0.3 Å	100	4%
	Paromomycin	Residues 1 and 2 (NEA2)	–	–	–	0.3 Å/1.5 Å	34/92	3%

[a] The prediction accuracy is evaluated by two criteria: the root mean square deviation (RMSD) and the score. The RMSD is measured for all non-hydrogen atoms between the positions of the MCSS minima and the ligands observed experimentally. The score (min = 0, max = 100) is given by the following expression: $100(1 - |(E_{max} - E_i)/(E_{max} - E_{min})|)$, where E_{max} and E_{min} are the maximum and minimum energies of interaction, respectively, and E_i the energy of interaction for the corresponding MCSS minimum. A score of 100 indicates that the minimum has the best energy of interaction

[b] The occupancy is defined as the proportion of MCSS minima similar to the ligand observed experimentally (RMSD ≤ 1.5 Å)

[c] HIV-1 TAR RNA structure determined by Puglisi et al. [4]

[d] HIV-1 TAR RNA structure determined by Aboul-ela et al. [3]

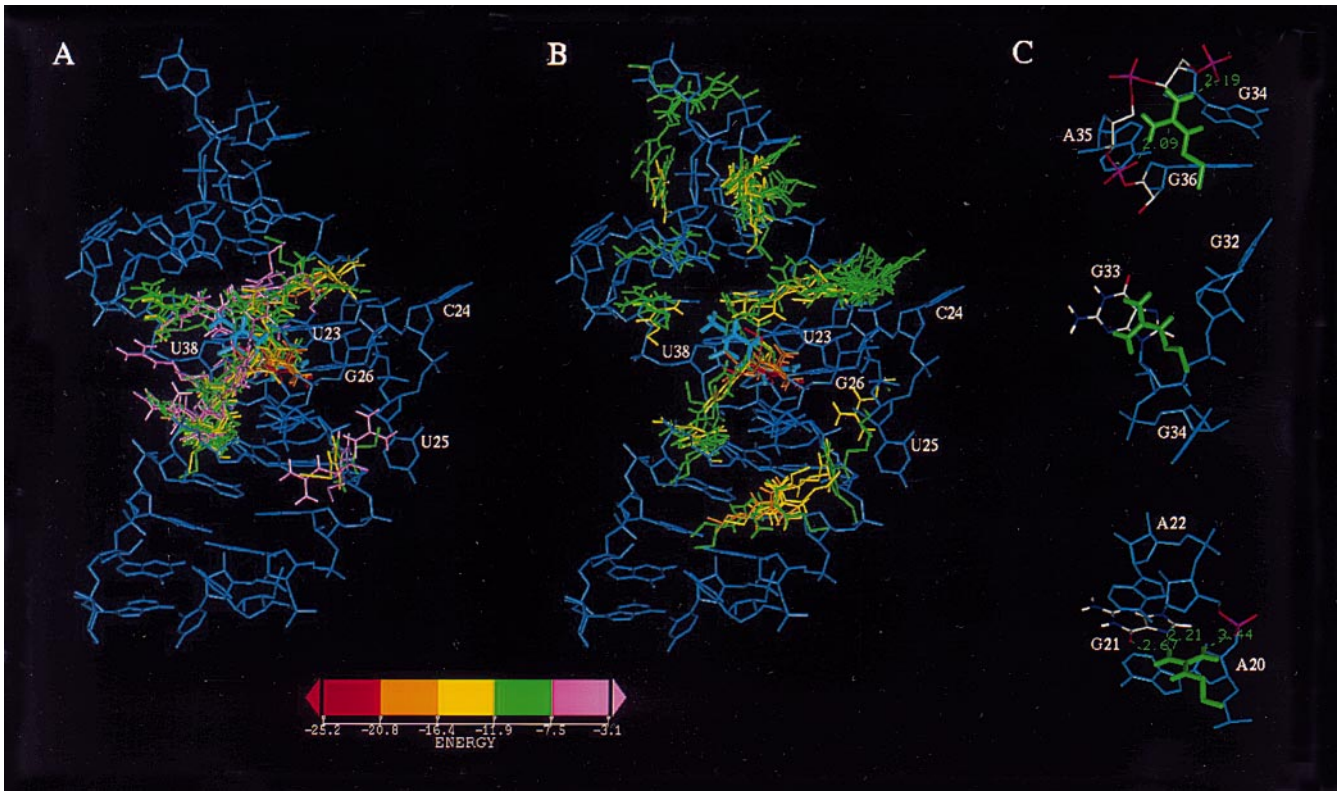


Fig. 1A–C. Distribution of MCSS minima at the surface of the HIV-1 TAR RNA [3]. **A** Minima identified within a sphere of 12 Å centered on the structure of the arginine residue. **B** Minima identified on the entire surface of the RNA. The RNA target, colored in *blue*, corresponds to the bound TAR conformation; its bound arginine is colored in *light blue*. The score of each minima is given by the color scheme at the bottom representing the energy of interaction. **C** Detail of the three arginine binding modes (*top*: interactions with the phosphate backbone; *middle*: stacking interactions with the bulge nucleotide G33; *bottom*: arginine fork motif). The hydrogen bonds are represented by *dashed lines*. The program InsightII was used for the graphical representation (MSI, San Diego)

(Fig. 1C). The minima with high scores largely overlap with the arginine residue in the NMR structure of the complex. They also correspond to the more favorable binding sites on the entire surface of the RNA. The details of the interaction between the minima with the best score and the RNA binding site are represented on Fig. 2. The best minima (with the 10 highest scores) reproduce accurately the position of the arginine side-chain in the two NMR structures of the arginine-TAR RNA complexes, but with a lower RMSD in the case of the structure determined by Puglisi et al. [3, 4] (see Table 1 and Fig. 2).

The distribution of the minima corresponding to the four paromomycin residues are shown in Fig. 3A and 3B. The global distribution for the AG1N (first residue), STRP (second residue), RIB (third residue), and ID2N (fourth residue) groups is similar: for all the residues (Fig. 3C), independently of the charge and shape, three different favorable binding regions are found. The minima with high scores are concentrated mostly in one region (Fig. 3B). The position of the best AG1N and ID2N minima are in good agreement with that of the first and fourth residues of paromomycin (Fig. 4); the minima for the STRP and RIB groups, which share some overlap with the actual positions of the corresponding paromomycin residues, have very low scores. These latter could be due to the fact that these residues occupy a sub-optimal position in the binding site as part of the paromomycin molecule. To test this hypothesis, an additional search was performed using a larger functional group (NEA2) including both the first and second residues in order to determine the more favorable positions of the STRP group when merged to AG1N (see Scheme 1). The results demonstrate that the binding site of the second residues can be better predicted in the context of a larger functional group (see Table 1). In the case of aminoglycoside moieties, the computational cost of the MCSS search remains linearly dependent on the size of the functional group (on a DEC 3000 AXP 500, around 4 h of computer time for the AG1N and STRP groups and around 10 h of computer time for NEA2). A general view of the paromomycin binding site and the positions of some minima are shown in Fig. 5. The accuracy of the predicted binding sites for each functional group is summarized in Table 1.

4 Discussion

We show that a simple model that makes use of the MCSS method with a modified CHARMM force field (reduced phosphate charges) can simulate RNA-ligand interactions and predict RNA binding sites. The predic-

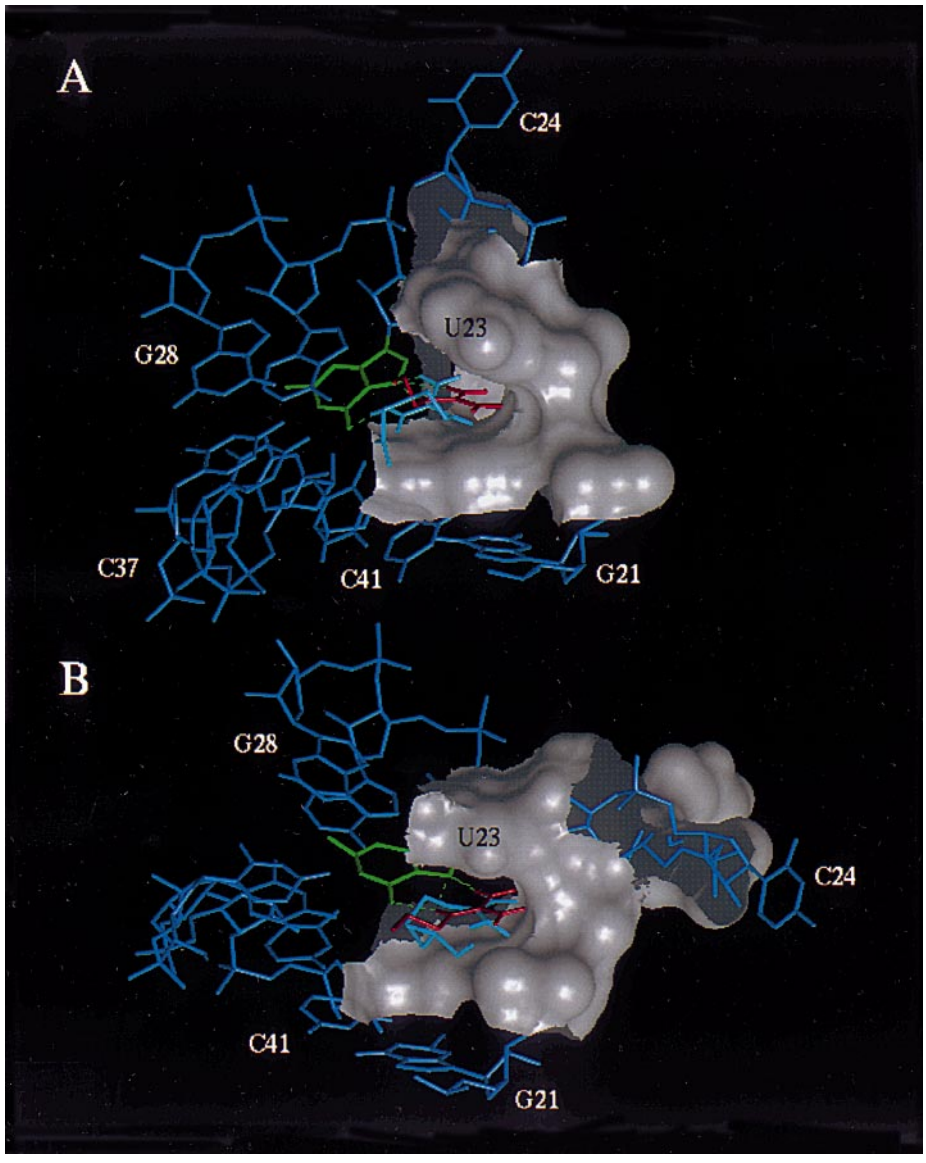


Fig. 2A, B. Close-up view of the MCSS minima within the TAR arginine binding site. The structure of the arginine-RNA complex is colored as in Fig. 1 (RNA in *blue*, arginine in *black*). The binding site is represented by a solvent accessible surface. **A** NMR structure by Aboul-ela et al. [3]. **B** NMR structure by Puglisi et al. [4]. The MCSS minimum with the highest score is shown in *red*, the guanine forming the arginine fork motif is in *green*

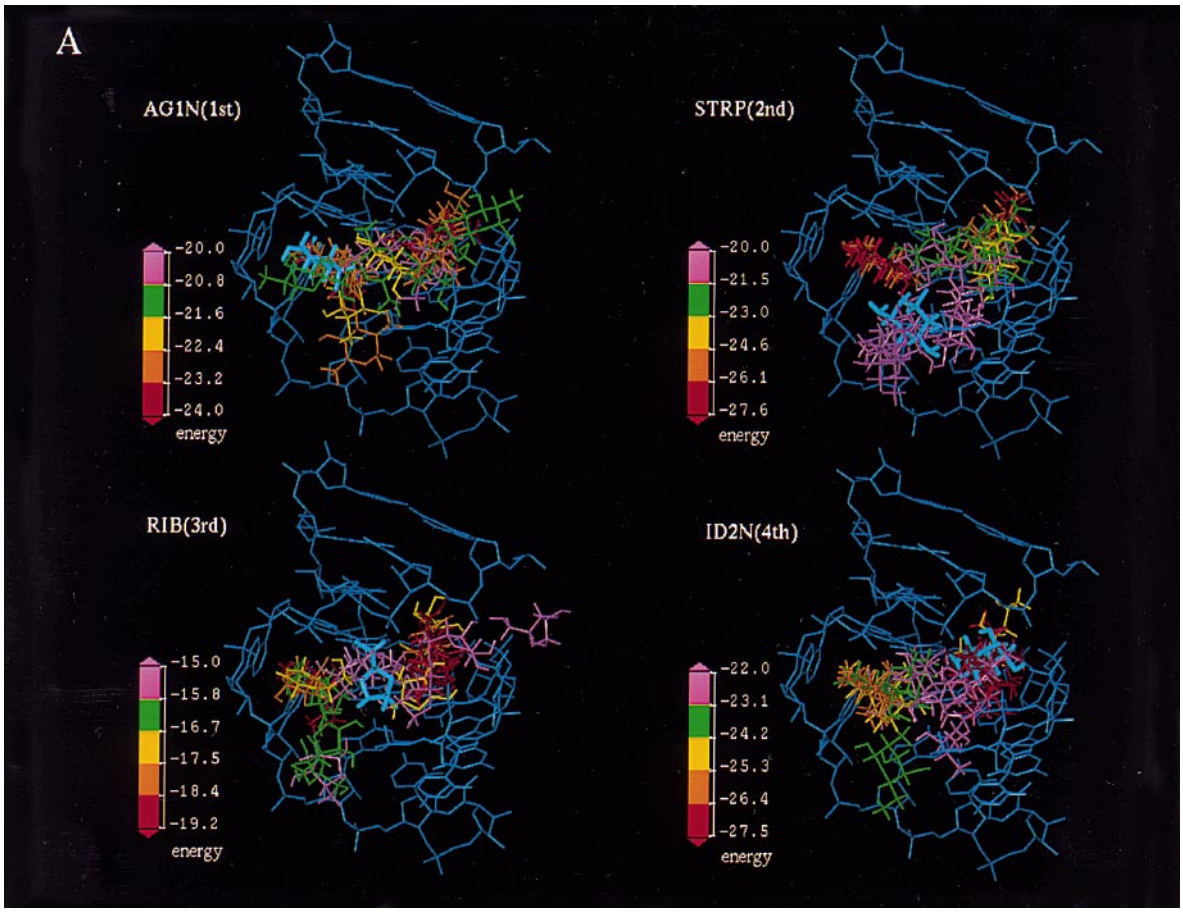
tion accuracy is dominated by the geometry and electrostatic properties of the RNA binding site.

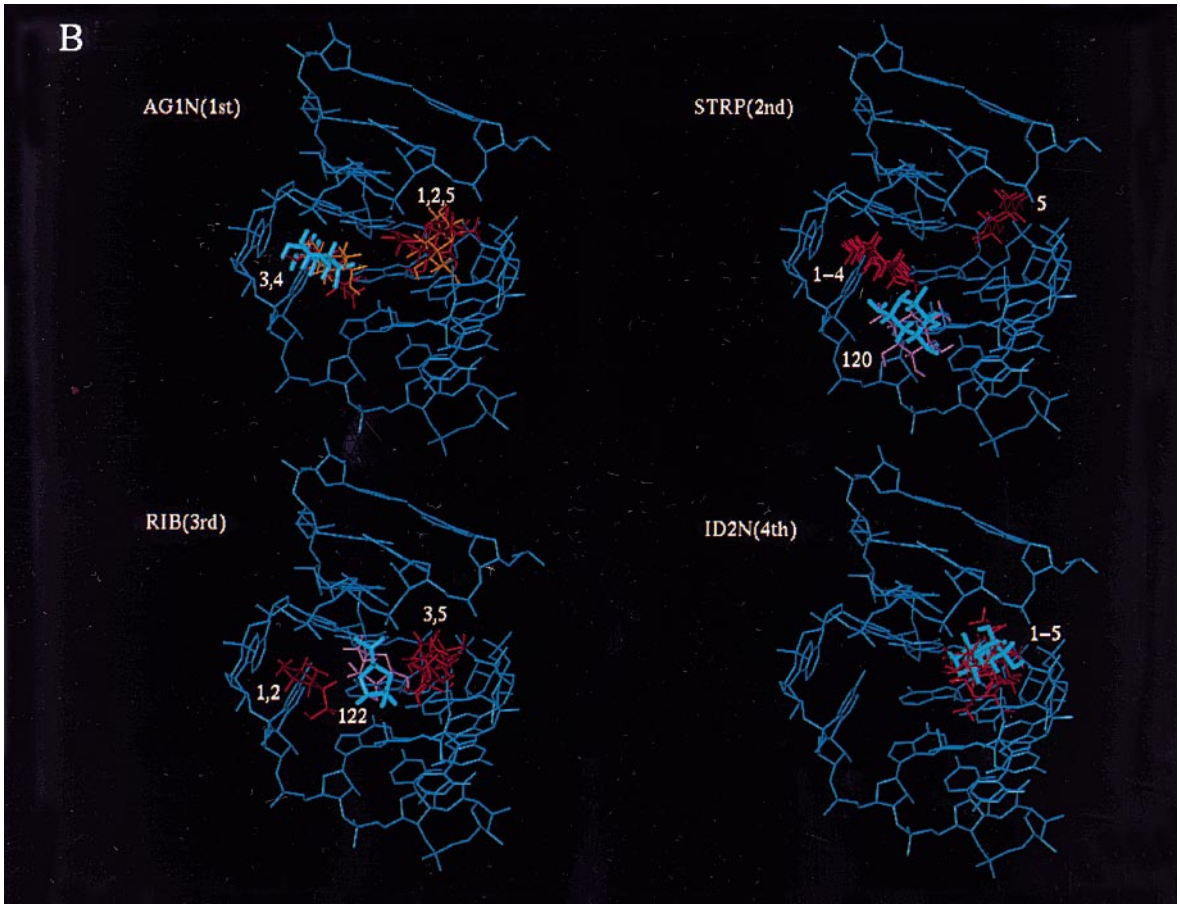
In the case of the TAR RNA, the accuracy is sensitive to the NMR structure used as the target (see differences in RMSD and score in Table 1). In the first TAR RNA structure [4], the bulge nucleotide U23 forms a base-triple with the Watson-Crick base pair A27 · U38, a structural feature absent from the second structure [3]. The presence of the base triple restrains the positions of the MCSS minima to a well-defined cavity where the guanidinium group of the arginine residue is perfectly stacked over the bulge nucleotide U23 and paired to the guanine G26 according to the arginine fork motif [4] (see Fig. 2). In the second structure [3], the MCSS minima tend to stack over U23 close to the phosphate backbone and are more poorly paired to G26.

In the case of the 16S rRNA, the predictions depend strongly on the geometry and electrostatic properties of the RNA binding site. Independently of the functional group used to represent one of the four residues of paromomycin, two regions of the RNA appear to be very favorable for binding. One binding region is a well-defined cavity in which the first residue of paromomycin fits by stacking interactions with the nucleotide A1492 (see Fig. 5). This region is also an attractive pole for monocationic residues like the first and second residue of paromomycin (see Fig. 3A, B). The second region is also

►

Fig. 3. **A** Distribution of MCSS minima at the surface of the 16S rRNA. The four panels represent the distribution for each of the functional groups corresponding to the four paromomycin residues (first residue: AG1N; second residue: STRP; third residue: RIB; fourth residue: ID2N) colored in *light blue*. A color scheme indicates the score of the minima for each functional group. **B** Distribution of high score MCSS minima. The five minima with the highest scores are shown for each functional group. The minima closer to the corresponding paromomycin residue are minimum 4 for the first residue (AG1N), minimum 122 for the second residue (STRP), minimum 122 for the third residue (RIB), and minima 1 and 2 for the fourth residue (ID2N)


A
AG1N(1st)
-20.0
-20.8
-21.6
-22.4
-23.2
-24.0
energy
STRP(2nd)
-20.0
-21.5
-23.0
-24.6
-26.1
-27.6
energy
RIB(3rd)
-15.0
-15.8
-16.7
-17.5
-18.4
-19.2
energy
ID2N(4th)
-22.0
-23.1
-24.2
-25.3
-26.4
-27.5
energy


B
AG1N(1st)
1,2,5
3,4
STRP(2nd)
5
1–4
120
RIB(3rd)
3,5
1,2
122
ID2N(4th)
1–5

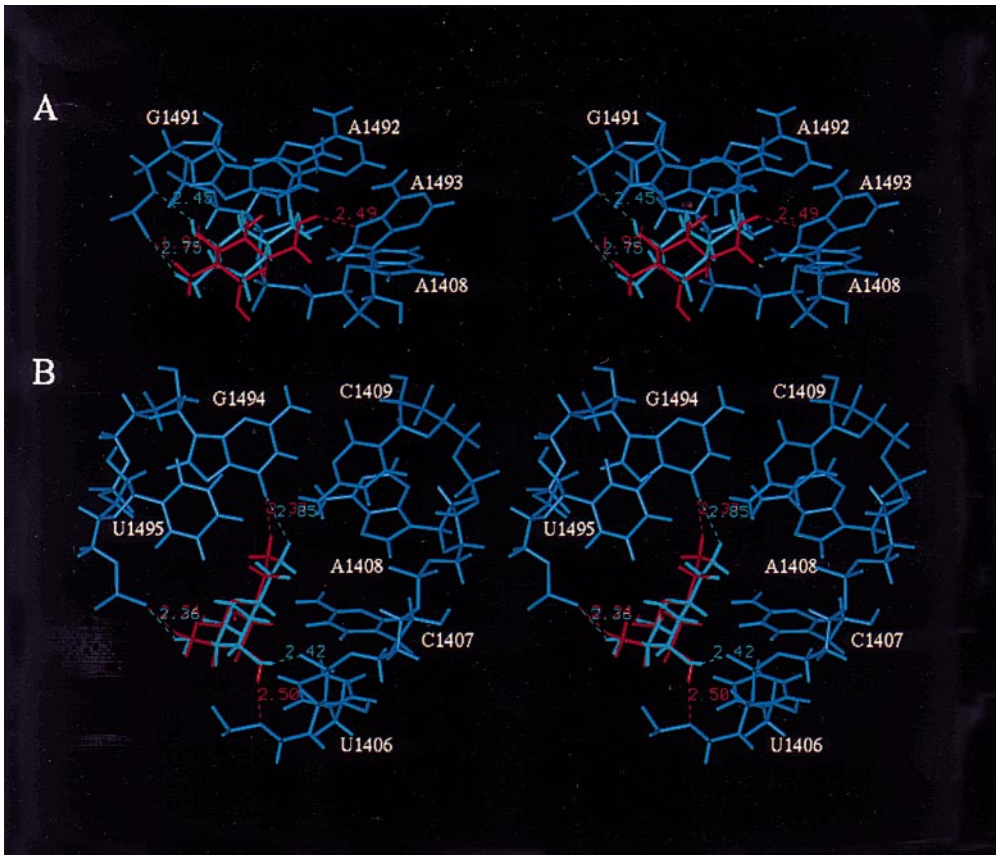
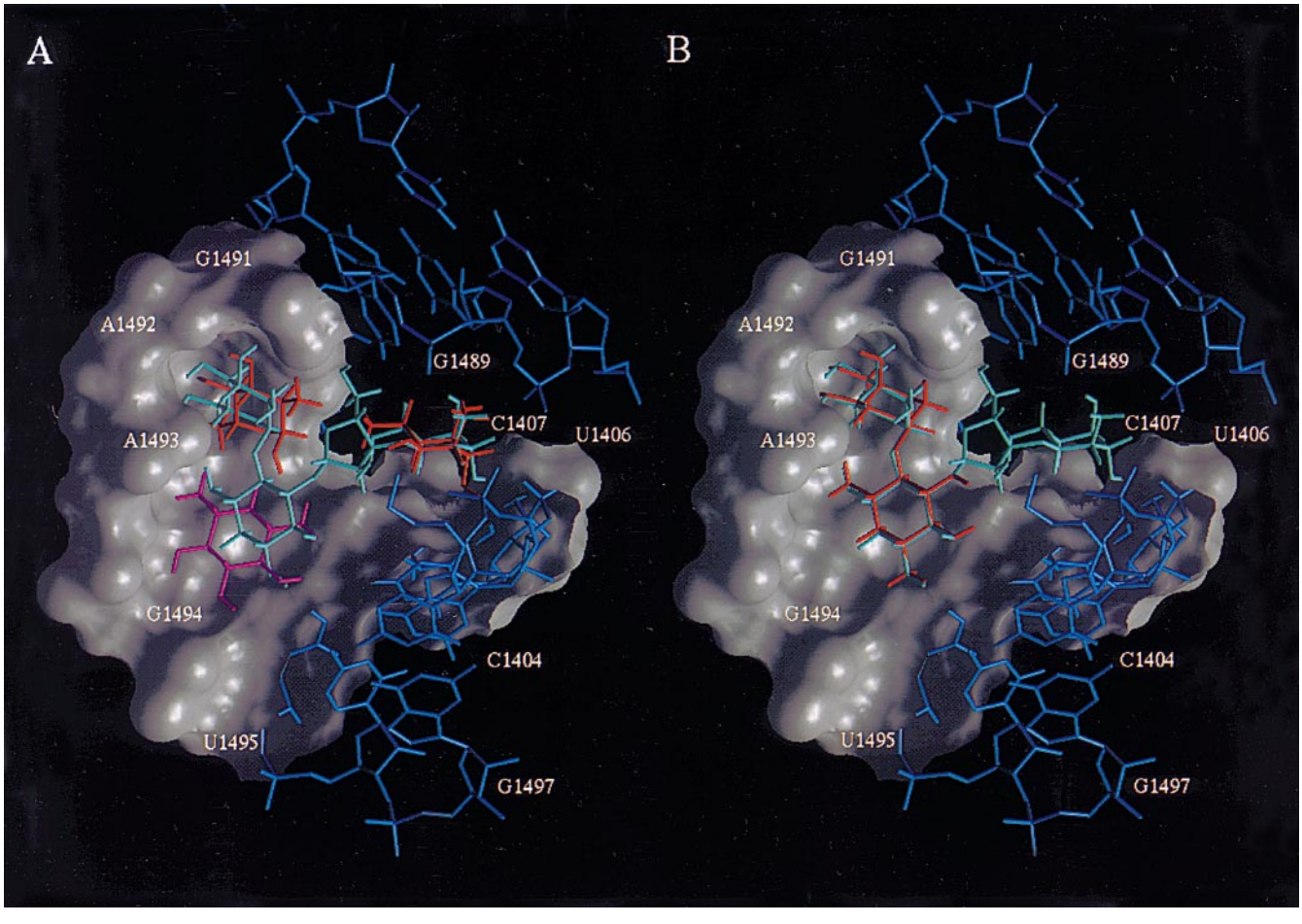


Fig. 4A, B. Close-up view of the MCSS minima for the AG1N and ID2N groups in the paromomycin binding site. **A** MCSS minimum for the AG1N group. **B** MCSS minimum for the ID2N group. The MCSS minima are colored in *red*, the corresponding residues in the RNA-aminoglycoside complex in *light blue*. The hydrogen bonds between the RNA and the ligands (distance X-H $\leq$3.0 Å where X = O, N) are indicated by *dashed lines* according to the same color scheme. The MCSS minima represented correspond to those described in Table 1



a very attractive electrostatic pole for mono- or dicationic groups because of the proximity of two portions of the RNA backbone (the phosphates of nucleotides G1489 and C1407 in Fig. 5). The decomposition of the binding free energy on a per residue basis indicates that the group for which the binding position and conformations are predicted best corresponds to the residue that contributes the most to the free energy of binding (results not shown), in agreement with recent experimental data [16].

5 Conclusions

A preliminary study has been made which indicates that the MCSS method can be used to investigate functional group binding to nucleic acids. To improve the reliability of the predictions, models which give a more accurate description of the solvent and polyelectrolyte effects on RNA-ligand interactions are being evaluated. Reliable predictions of known RNA binding sites will open new perspectives for the application of combinatorial structure-based drug design to these molecules, which are emerging as attractive drug targets.

Acknowledgements. This work was supported in part by a grant from the Department of Energy to Harvard University.

◀

Fig. 5A, B. General view of the MCSS minima within the paromomycin binding site. The structure of the paromomycin-16S rRNA complex is colored as in Fig. 3 (RNA in *blue*, paromomycin in *light blue*). The binding site is represented by a solvent accessible surface. **A** The minima 1 and 4 (AG1N and ID2N) corresponding to the first and fourth residues of paromomycin are shown in *red*; the minimum 122 (STRP) corresponding to the second residue of paromomycin is shown in *pink* according to the score. **B** The minimum with the lowest RMSD (see Table1) corresponding to the residues 1 and 2 (NEA2) is shown in *red*

References

1. Miranker A, Karplus M (1991) Proteins 11:29
2. Caflisch A, Miranker A, Karplus M (1993) J Med Chem 36:2142
3. Aboul-ela F, Karn J, Varani G (1995) J Mol Biol 253:313
4. Puglisi JD, Tan R, Calnan BJ, Frankel AD, Williamson JR (1992) Science 257:76
5. Fourmy D, Recht MI, Blanchard SC, Puglisi JD (1996) Science 274:1367
6. Ramos A, Gubser CC, Varani G (1997) Curr Opin Struct Biol 7:317
7. Veal JM, Wilson WD (1991) J Biomol Struct Dyn 8:1119
8. Leclerc F, Cedergren R (1998) J Med Chem 41:175
9. Gordon EM, Barrett RW, Dower WJ, Fodor SP, Gallop MA (1994) J Med Chem 37:1385
10. Shuker SB, Hajduk PJ, Meadows RP, Fesik SW (1996) Science 274:1531
11. Caflisch A (1996) J Comput Aided Mol Des 10:372
12. MacKerell AD Jr, Wiorkiewicz-Juczera J, Karplus M (1995) J Am Chem Soc 117:11946
13. Tidor B, Irikura KK, Brooks BR, Karplus M (1983) J Biomol Struct Dyn 1:231
14. Neria E, Fischer S, Karplus M (1996) J Chem Phys 105:1902
15. Brooks BR, Bruccoleri RE, Olafson D, States DJ, Swaminathan A, Karplus M (1983) J Comput Chem 4:187
16. Alper PB, Hendrix M, Sears P, Wong C-H (1998) J Am Chem Soc 120:1965