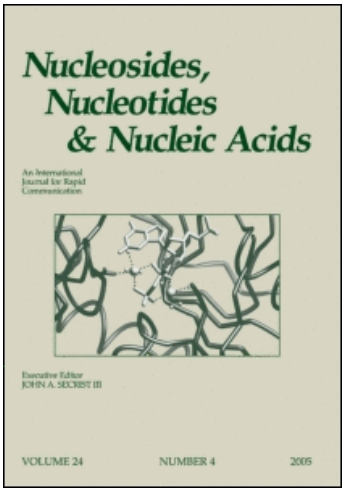

Nucleosides, Nucleotides and Nucleic Acids
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597286

Analysis of Local Convergence in NMR Structure Calculation for RNA by a Classification System for Nucleic Acid Structure (CSNA)

Tatsuhiko Someya[a]; Taiichi Sakamoto[a]; Akitsugu Takasu[ab]; Gota Kawai[a]
[a] Department of Industrial Chemistry, Faculty of Engineering, Chiba Institute of Technology, Chiba, Japan [b] Protein Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama, Japan

Online publication date: 05 June 2004

To cite this Article Someya, Tatsuhiko , Sakamoto, Taiichi , Takasu, Akitsugu and Kawai, Gota(2004) 'Analysis of Local Convergence in NMR Structure Calculation for RNA by a Classification System for Nucleic Acid Structure (CSNA)', Nucleosides, Nucleotides and Nucleic Acids, 23: 4, 691 — 700
To link to this Article: DOI: 10.1081/NCN-120037509
URL: http://dx.doi.org/10.1081/NCN-120037509

NUCLEOSIDES, NUCLEOTIDES & NUCLEIC ACIDS
Vol. 23, No. 4, pp. 691–700, 2004

Analysis of Local Convergence in NMR Structure Calculation for RNA by a Classification System for Nucleic Acid Structure (CSNA)

Tatsuhiko Someya,[1] Taiichi Sakamoto,[1] Akitsugu Takasu,[1,2] and Gota Kawai[1,*]

[1]Department of Industrial Chemistry, Faculty of Engineering, Chiba Institute of Technology, Chiba, Japan
[2]Protein Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama, Kanagawa, Japan

ABSTRACT

We are developing a program system, CSNA, to classify a set of structures into groups sharing similar structural characters. In the present study, CSNA was applied to the analysis of NMR structures obtained by the simulated annealing calculation to elucidate local convergences. A 34-mer RNA, U6-34, having a bulge-out region that is derived from the human U6 snRNA is used as a target molecule in the present study. Although the structure calculation was not converged with the conventional method, it was found by the CSNA analysis that the two stem regions in the molecule were converged well. Furthermore, one strand of the bulge-out region (A7–A11) was found to form a continuously stacked structure in two-thirds of calculated structures. In

*Correspondence: Gota Kawai, Department of Industrial Chemistry, Faculty of Engineering, Chiba Institute of Technology, 2-17-1 Tsudanuma, Narashino-shi, Chiba 275-0016, Japan; Fax: (81)-47-478-0425; E-mail: gkawai@ic.it-chiba.ac.jp.

DOI: 10.1081/NCN-120037509 1525-7770 (Print); 1532-2335 (Online)

conclusion, CSNA can be a novel tool to elucidate the local convergence of the NMR structure calculations.

Key Words: RNA structure; NMR; Classification; Hydrogen bond; Base–base stacking.

INTRODUCTION

Nuclear magnetic resonance (NMR) spectroscopy is a powerful tool for studying the structure and dynamics of biomolecules in solution and their interactions with ligands. In the case of RNA structure determination, NMR is more favorable compared to X-ray crystallography because crystallization of RNA molecules is difficult in many cases. In the process of structure determination of RNA molecules using NMR, one of the important parts is the structure calculation. Usually, hundred structures are calculated using restrained molecular dynamics protocols and at least ten converged structures having low overall energy are selected as NMR structures. However, due to the lack of information or intrinsic flexibility of RNA, the output structures are not converged well. In some cases, it is hard to know what happened by just checking the total energy and NMR violations. On the other hand, it may happen that, even though the overall structures are not converged, some parts are converged locally, or more than one converged structures are simultaneously obtained. Thus, the classification of structures according to their similarities will be useful for analyzing the character of a set of calculated structures.

We have reported a system of computer programs, ''Classification System for Nucleic Acid structure determination (CSNA)'' to extract the hydrogen bond or base–base stacking information and classify structures according to the information.[1] In the previous study, CSNA was applied to the classification of the results of two individual

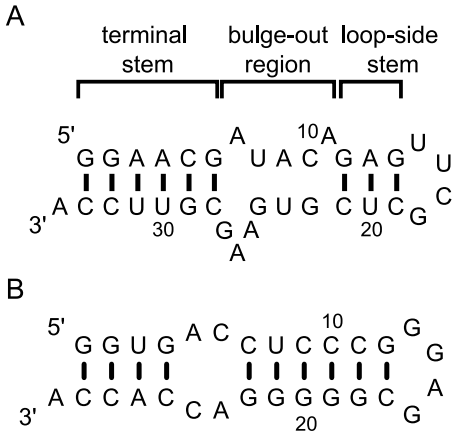

Figure 1. Secondary structures of RNAs used in this study. A: 34-mer fragment of human U6 snRNA, U6-34. U6-34 consists of a terminal stem (G1–G6/C28–C33), a loop-side stem (G12–G14/C19–C21), a bulge-out region (A7–A11/G22–G27) and a closing UUCG loop as indicated in the figure. B: 29-mer fragment of human SRP RNA, HE6.

MARCEL DEKKER, INC.
270 Madison Avenue, New York, New York 10016

restrained molecular dynamics calculations and was found to be useful for the extraction of the structures with lower energy without checking any energy term. The extracted structures were as well converged as the lowest energy structures. In this study, we applied CSNA to analyze local conformation of calculated structures.

A 34-mer RNA fragment U6-34 (Fig. 1A) derived from human U6 small nuclear RNA (snRNA) is the target of the present study. U6-34 consists of a terminal stem (G1–G6/C28–C33), a loop-side stem (G12–G14/C19–C21), a bulge-out region (A7–A11/G22–G27) and a closing UUCG loop and was subjected to the structure determination by the conventional NMR method (for review see Refs. [2–4]). We applied CSNA to analyze local conformation of the calculated structures of U6-34, comparing to a case of well-converged structures; a 29-mer RNA (HE6; Fig. 1B) corresponding to helix 6 of the human signal recognition particle (SRP) RNA[5] Utility of CSNA for the analysis of local conformation was discussed.

MATERIALS AND METHODS

Structure Determination of U6-34

For U6-34, RNA oligonucleotides were enzymatically synthesized by an in vitro transcription reaction using T7 RNA polymerase and purified by polyacrylamide gel electrophoreses. Stable isotopically labeled RNA samples were also prepared by using [*U*-^{13}C,^{15}N]NTPs (Nippon Sanso) and NMR spectra were analyzed with the conventional methods.[2–4] Experimental details were described elsewhere (Someya et al., submitted).

A set of 100 structures was calculated with the InsightII/Discover package (Accelrys). The simulated annealing protocol is the same with that used for structure calculation of HE6[5] and details will be submitted elsewhere.

Classification of Structures by CSNA

CSNA was applied to classify the 100 structures of U6-34 or HE6, which have been obtained by NMR structural calculations. For each of U6-34 and HE6, an appropriate value of threshold was chosen to obtain the best group containing about 10 structures. For each structural group, the averaged pair wise root mean square deviation (r.m.s.d.) was calculated.

Analysis of the Hydrogen Bonding and Base–Base Stacking Pattern

In the classification procedure of CSNA, two types of lists are produced; one is the ''complete list'' of the hydrogen bonds and base–base stackings and the other is the individual list (''bit list'') for each structure showing the existence of the hydrogen bonds and base–base stackings in the complete list (for details see Ref. [1]). By using these lists, the frequencies of the hydrogen bonds and base–base stackings for each pair of residues were calculated. This part of analysis was performed by a combination of VisualBasic and Excel (Microsoft) in the present study and will be rewritten with the C and awk languages to be included in the web system.

MARCEL DEKKER, INC.
270 Madison Avenue, New York, New York 10016

Analysis of the Bulge-Out Region of U6-34

The coordinates of residues A7–A11 are extracted from the set of NMR structures of U6-34 and subjected to CSNA to classify the structure of this part. Based on the grouping of the region, the best group obtained by the whole structure analysis was further analyzed.

RESULTS AND DISCUSSION

Analysis of NMR Spectra and the Structure Calculation of U6-34

Imino proton and non-exchangeable proton signals were assigned and it was confirmed that U6-34 forms a stem-loop structure with a bulge-out region as shown in Fig. 1A. Sharp imino proton signals were not observed for residues in the bulge-out region including U8, G22, U23, G24 and G27. Details for spectral analysis was described elsewhere (Someya et al., submitted). A total of 420 NOE distance restraints, 171 dihedral restraints, and 26 hydrogen bonding restraints were used for conventional structural calculation. The 14 structures with lowest total energy were chosen from a hundred structures and the r.m.s.d. was calculated to be 2.6 Å. The lowest energy structures contain neither any NOE restraint (> 0.2 Å) nor dihedral restraint (>5°) violations. The superimposition of the lowest energy structures was shown in Fig. 2A. It is clear that the structure calculation of U6-34 is not converged well, resulting in the fluctuation of the global structures. This is due to the intrinsic flexibility of U6-34 and/or lack of structural information. The long-range structural information obtained from the residual dipolar couplings may improve the convergence of the global structures. However, it is still true that structural calculation is not always converged for RNA molecules due to the conformational flexibility.

Classification by CSNA

The 100 structures of U6-34 described above and HE6 obtained by Sakamoto et al. were subjected to CSNA. For U6-34, 468 hydrogen bonds and 64 stackings were detected and, with the threshold of 9, the best group contained 14 structures. The best group represents the most commonly appeared structures in the set of structures. The threshold defines the maximum structural difference for structures to be included in a group. Thus, the smaller value of threshold means that the structural similarity among the selected structures is higher. The averaged pair wise r.m.s.d. for the best group was 3.3 Å and the superimposition is shown in Fig. 2B. For HE6, 287 hydrogen bonds and 47 stackings were detected and, with the threshold of 6, the best group contained 12 structures. The r.m.s.d. for HE6 was 1.1 Å for the 12 lowest energy structures and 1.6 Å for the best group. The superimpositions of the lowest energy structures and the best group are shown in Fig. 2C and 2D, respectively. In the case of HE6, the calculation was converged well and the smaller value of r.m.s.d. was obtained with the smaller value of threshold compared to the case of U6-34.

The most obvious difference between the U6-34 and HE6 was observed in the number of hydrogen bonds in the complete list detected by CSNA; 14 per residue for

MARCEL DEKKER, INC.
270 Madison Avenue, New York, New York 10016

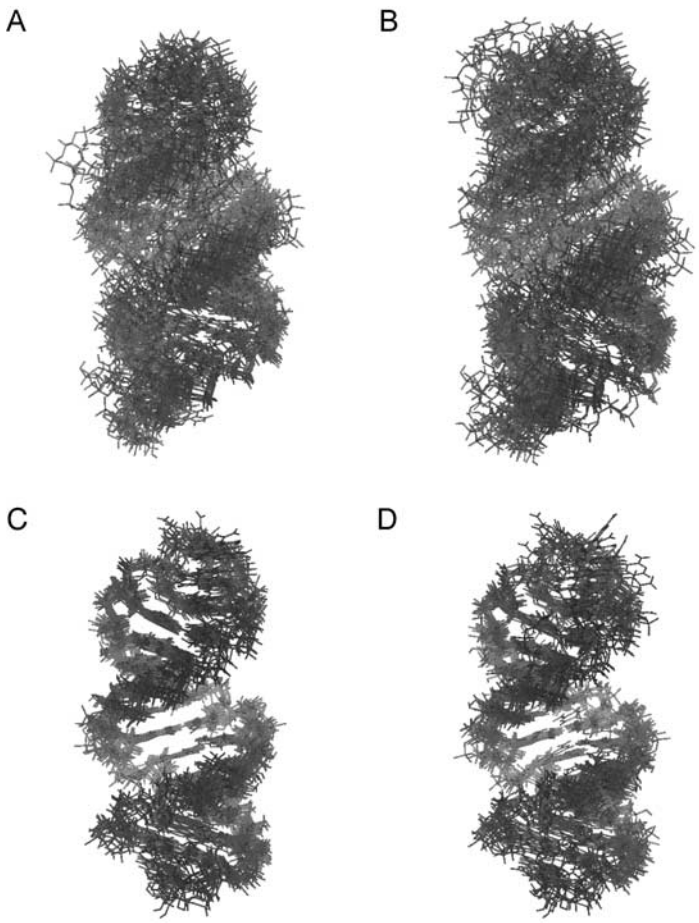

Figure 2. Superposition of U6-34 (A, B) and HE6 (C, D). A, C: The lowest energy structures. B, D: The CSNA best groups. The best group represents the most commonly appeared structures in the set of structures. All sets of structures are shown by superimposing all residues. (*View this art in color at www.dekker.com.*)

U6-34 whereas 10 for HE6. Probably, this reflects the difference in convergence of the structure calculation. Because the calculation for U6-34 was not well converged, each structure takes different conformations, resulting in the increase of the number of hydrogen bond. Then, we further analyzed the structures of U6-34 to elucidate the local convergence.

Pattern Analysis of U6-34 and HE6

The patterns of the existence of hydrogen bonds and stackings for each pair of residues were then analyzed to elucidate the local structural character for U6-34 as well as HE6. Figure 3A shows the pattern of hydrogen bonds for U6-34. The anti-diagonal

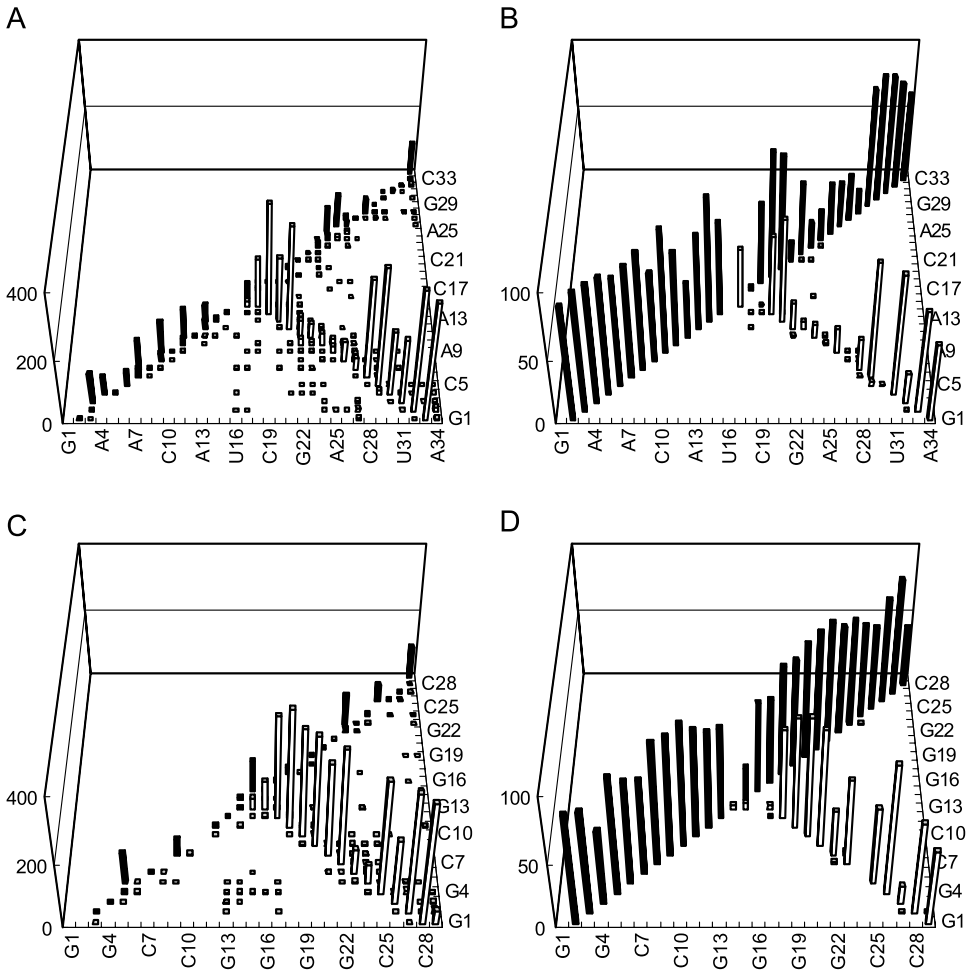

Figure 3. Mapping of the hydrogen bond and base–base stacking analyzed by CSNA of U6-34 (A, B) and HE6 (C, D). A, C: Mapping of hydrogen bond. B, D: Mapping of base–base stacking. Information of intra-residue or adjacent residues is indicated by filled bars, and the rest information is indicated by open bars.

values reflect the formation of base pairing and it can be seen that almost no hydrogen bonding was detected in the bulge-out region. Figure 3B shows the pattern of stackings for U6-34 and, again, almost no stacking was detected in the one strand of the bulge-out region, G22–C28 (the near diagonal values). These are the reasons why the overall structure of the best group or the lowest energy group was not well converged. However, some parts of U6-34 show the formation of hydrogen bonds and stackings including the two stems and a strand of the bulge-out region (A7–A11). Figure 4 shows the superimposition on each of the two stems of the best group of U6-34. As indicated by the CSNA analysis (Fig. 3A and 3B), each stem is well converged and the r.m.s.d. values of

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

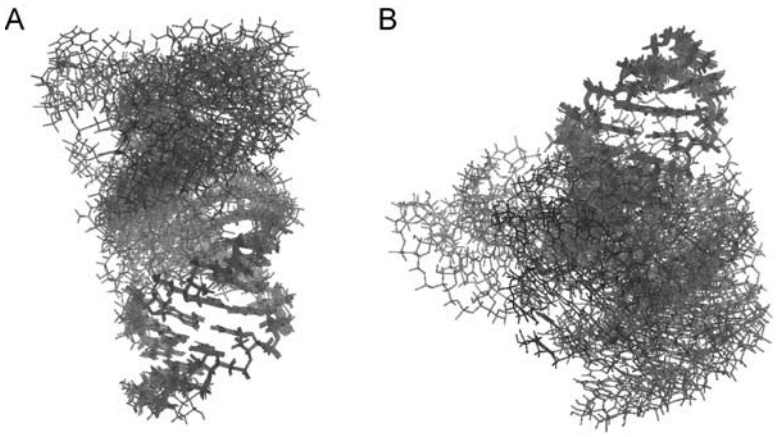

Figure 4. Local convergence shown in the superposition of the best group for U6-34. The structures of the best group were superimposed by the (A) terminal stem or (B) loop-side stem. (*View this art in color at www.dekker.com.*)

the terminal stem and loop-side stem were 0.3 and 0.4 Å, respectively. Further analysis was performed for the bulge-out region as described below.

In contrast, Fig. 3C shows the formation of base pairs throughout the stem of HE6 including two A:C pairs in the middle of stem, which is confirmed by the graph (the anti-diagonal values). The hydrogen bond for each A:C pair in the middle of stem was detected in more than 60% of the structures. It should be noted that only one hydrogen bond is formed in the A:C pair whereas three hydrogen bonds were formed in the G:C pair and these are reflected in the numbers of hydrogen bonds detected by CSNA. Figure 3D also shows the stacking interactions throughout the stem. These results agree with the tertiary structure of HE6 determined by NMR methods [5] as well as the X-ray crystallography.[6] It is noted that CSNA detected the stacking in the GGAG loops representing the character of the tetraloop; the continuous stacking of G14–G16 and the sharp turn between the G13 and G14. Furthermore, it is noted that CSNA detected some inter-strand stackings (the anti-diagonal values).

Analysis of the Bulge-Out Region of U6-34

As shown in Fig 3A and 3B, A7–A11 in the bulge-out region forms stacking and, thus, the region was extracted to be subjected to CSNA. Sixty-three hydrogen bonds and four stackings were detected and 8 groups were obtained with the threshold value of 2 (Fig. 5A). The best group containing 65 structures shows the r.m.s.d. value of 1.5 Å. Figure 5B shows the superimposition of these structures and it was found that the five nucleotides A7–A11 are almost continuously stacked although the conformational fluctuation is rather high. Figure 3A shows the existence of small number of hydrogen bonds within the bulge-out region and this may contribute to the formation of the stacked structure of A7–A11. The structures of the best group obtained by the whole structure analysis described above are indicated by filled bars in Fig. 5A, showing that the 10 out of 14 structures in the best group of the whole structure analysis were

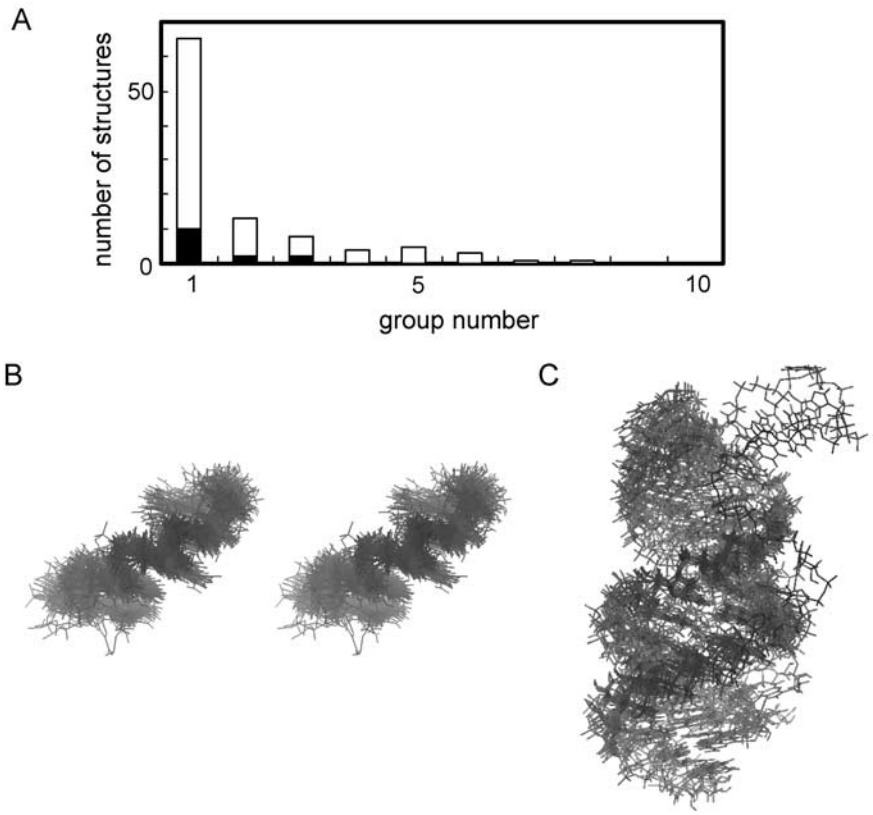

Figure 5. Classification of structure of A7–A11 in the bulge-out region of U6-34. A: Distribution of the number of structures. The value of threshold was 2. The first group (number 1) represents the best group. The structures in the best group of the whole structure analysis are indicated by filled bars. B: Stereo view of the superposition of structure in the best group of the local (A7–A11) analysis. C: The structures included in both of the best groups of the whole and local analyses. The structures are superimposed by A7–A11. (*View this art in color at www.dekker.com.*)

included in the best group of the local structure analysis. Figure 5C shows the superimposition by A7–A11 of the 10 structures, showing that the overall structures are not converged even though each of the three parts of the molecule, two stems and one strand of the bulge region, is converged.

CSNA as a Novel Tool for the Analysis of Local Convergence

In the course of the structure determination by NMR method especially for RNA, it is not easy to find out which part of the molecule is converged when overall structure is not converged. Present study shows that CSNA will be a useful tool to find out the local convergence of the NMR calculation. Recently, a novel method using the residual dipolar coupling was developed to improve the global structure of RNAs during the NMR structure calculation.[7] Nevertheless, RNA molecules sometimes have internal

flexibility that makes the NMR analysis difficult, and, in such cases, CSNA can tell us the conformational character of the set of structures that will be important for the elucidation of the function of RNA molecules.

In the beginning, CSNA was developed to classify the structures generated by the structural modeling program MC-SYM,[8] and then applied to the evaluation of NMR structure calculation.[1,9–11] Here, we showed the potential of CSNA to find out the local convergence and we are trying to apply CSNA to analyze the trajectory of the molecular dynamics simulations. The trial version of CSNA can be accessed on our web site (http://rna.le.it-chiba.ac.jp/).

ABBREVIATIONS

snRNA	small nuclear RNA
NMR	Nuclear magnetic resonance
r.m.s.d.	root mean square deviation

ACKNOWLEDGMENTS

This work was supported by ''Research for the Future'' Program (JSPS-RFTF97L00503) from the Japan Society for the Promotion of Science and by, in part, a Grant-in-Aid for High Technology Research from Ministry of Education, Science, Sports and Culture, Japan.

REFERENCES

1. Takasu, A.; Watanabe, K.; Kawai, G. Classification of RNA structures based on hydrogen bond and base–base stacking patterns: application for NMR structures. J. Biochem. **2002**, *132* (2), 211–215.
2. Pardi, A. Multidimensional heteronuclear NMR experiments for structure determination of isotopically labeled RNA. In *Methods in Enzymol.*; James, T.L., Ed.; Academic Press, Inc.: London, 1995; Vol. 261, 350–380.
3. Varani, G.; Aboul-era, F.; Allain, F.H.-T. NMR investigation of RNA structure. Prog. Nucl. Magn. Reson. Spectrosc. **1996**, *29* (1–2), 51–127.
4. Puglisi, E.V.; Puglisi, J.D. Nuclear magnetic resonance spectroscopy for RNA. In *RNA Structure and Function*; Simons, R.W., Grunberg-Manago, M., Eds.; Cold Spring Harbor Laboratory Press: New York, 1998; 117–146.
5. Sakamoto, T.; Morita, S.; Tabata, K.; Nakamura, K.; Kawai, G. Solution structure of a SRP19 binding domain in human SRP RNA. J. Biochem. **2002**, *132* (2), 177–182.
6. Wild, K.; Weichenrieder, O.; Leonard, G.A.; Cusack, S. The 2 Å structure of helix 6 of the human signal recognition particle RNA. Struct. Fold. Des. **1999**, *7* (11), 1345–1352.
7. McCallum, S.A.; Pardi, A. Refined solution structure of the iron-responsive element RNA using residual dipolar couplings. J. Mol. Biol. **2003**, *326* (4), 1037–1050.

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

8. Major, F.; Turcotte, M.; Gautheret, D.; Lapalme, G.; Fillion, E.; Cedergren, R. The combination of symbolic and numerical computation for three-dimensional modeling of RNA. Science **1991**, *253* (5025), 1255–1260.
9. Takasu, A.; Kawai, G.; Watanabe, K. Development of a system to classify 3D structural character of RNA. Nucleic Acids Symp. Ser. **1999**, *42*, 233–234.
10. Takasu, A.; Kawai, G.; Watanabe, K. Classification of 3D structural character of RNA by hydrogen bond and base stacking. Nucleic Acids Symp. Ser. **2000**, *44*, 227–228.
11. Baba, S.; Takasu, A.; Watanabe, K.; Kawai, G. Application of the RNA structure classification system, CSNA, to NMR structure determination. Nucleic Acids Res. Suppl. **2003**, *3*, 127–128.

Received December 29, 2003
Accepted February 23, 2004

Request Permission or Order Reprints Instantly!