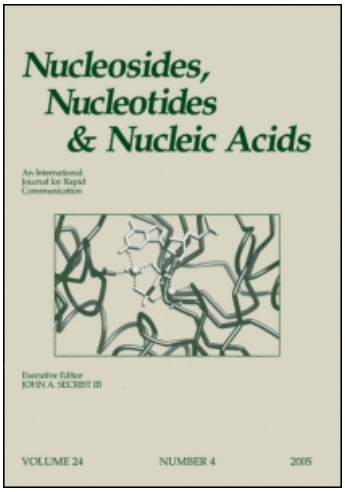

Nucleosides, Nucleotides and Nucleic Acids
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597286

ANALYSIS OF RELATIVE POSITIONS OF RIBONUCLEOTIDE BASES IN A CRYSTAL STRUCTURE OF RIBOSOME

Akitsugu Takasu[ab]; Kimitsuna Watanabe[bc]; Gota Kawai[a]
[a] Department of Industrial Chemistry, Chiba Institute of Technology, Narashino-shi, Chiba, Japan [b] Department of Chemistry and Biotechnology, Graduate School of Engineering, University of Tokyo, Bunkyo-ku, Tokyo, Japan [c] Department of Integrated Biosciences, Graduate School of Frontier Sciences, University of Tokyo, Kashiwa-shi, Chiba, Japan

Online publication date: 10 July 2002

To cite this Article Takasu, Akitsugu , Watanabe, Kimitsuna and Kawai, Gota(2002) 'ANALYSIS OF RELATIVE POSITIONS OF RIBONUCLEOTIDE BASES IN A CRYSTAL STRUCTURE OF RIBOSOME', Nucleosides, Nucleotides and Nucleic Acids, 21: 6, 449 — 462
To link to this Article: DOI: 10.1081/NCN-120014817
URL: http://dx.doi.org/10.1081/NCN-120014817

MARCEL DEKKER, INC. • 270 MADISON AVENUE • NEW YORK, NY 10016

NUCLEOSIDES, NUCLEOTIDES & NUCLEIC ACIDS
Vol. 21, Nos. 6 & 7, pp. 449–462, 2002

ANALYSIS OF RELATIVE POSITIONS OF RIBONUCLEOTIDE BASES IN A CRYSTAL STRUCTURE OF RIBOSOME

Akitsugu Takasu,[1,2] Kimitsuna Watanabe,[2,3] and Gota Kawai[1,*]

[1]Department of Industrial Chemistry, Faculty of Engineering, Chiba Institute of Technology, 2-17-1 Tsudanuma, Narashino-shi, Chiba 275-0016, Japan
[2]Department of Chemistry and Biotechnology, Graduate School of Engineering, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
[3]Department of Integrated Biosciences, Graduate School of Frontier Sciences, University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa-shi, Chiba 277-8562, Japan

ABSTRACT

Relative positions of bases to bases in a crystal structure of ribosome were analyzed extensively. It was found that there is no clear relation between bases apart more than 15 Å and, thus, the relative location of bases can be analyzed within 15 Å of the reference bases. As for base pairing, major positioning was found to be due to the Watson-Crick type base pairs. Some other positions corresponding to non-Watson-Crick type base pairs were also found in some extents. As for base–base stacking, it was observed that the bases stacked to adenine base are dispersive. It was found that less non-Watson-Crick base pairs was found close to the protein binding site, suggesting that the protein components have a tendency to bind to the regular stem structures. The database of

*Corresponding author. Fax: +81–47–478–0425; E-mail: gkawai@ic.it-chiba.ac.jp

DOI: 10.1081/NCN-120014817 1525-7770 (Print); 1532-2335 (Online)

relative location of bases must be useful for improvement of structural determination and structural modeling systems.

Key Words: Base pair; Base stacking; Ribosome; RNA; Structure

A recent topic in the field of RNA structural biology was a series of the crystal structures of ribosome.[1–8] Remarkably high resolution structure have been solved for a large subunit from *Haloarcula marismortui*, at 2.4 Å, including 2828 nucleotide residues.[4] Concurrently, a small subunit from *Thermus thermophilus* was analyzed with high resolution.[5,6]

Even with a crystal structure of ribosome small subunit, huge amount of structural information for RNA structure can be obtained and this amount already larger than all the information obtained so far. In fact, the size of RNA structural database suddenly increased by 8 times after the ribosome syndrome. Thus, now we have a useful source of RNA structure and it is required to analyze statistically the structures to build a database of RNA structural elements. The RNA structure in the *Haloarcula marismortui* ribosome was analyzed in detail to find new RNA motifs, A-minor and kink-turn motifs.[9,10] Furthermore, the RNA motifs and non-canonical interactions were classified to organize databases.[11,12] However, there is no comprehensive analysis for relative base location in known RNA structures so far, even though we have enough amount of structural information to perform statistically significant analysis.

In the present study, we analyze the relative positions of nucleotide bases as the characteristics of local structure of RNA molecules. We produced a database for the relative location of nucleotide bases and analyze its pattern. Such database must be useful for improvement of structural determination and structural modeling systems.

METHODS

The 5S and 23S RNA parts of the large subunit from *Haloarcula marismortui*[4] (Protein Data Bank: 1FFK) were used in this study, as well as the P4-P6 domain of the group I intron[13] (1GID), yeast $tRNA^{Phe}$ [14] (1EHZ) and the RNA part of hammerhead ribozyme[15] (1HMH). All calculations were performed on O2 workstation (Silicon Graphics) and programs were written by, mainly, C language. The resulting maps were drawn by ACE/gr system.[16]

Calculation of Base Position and Orientation

In order to define the relative position between two nucleotide bases, the center position and the orientation of a base were defined as shown in

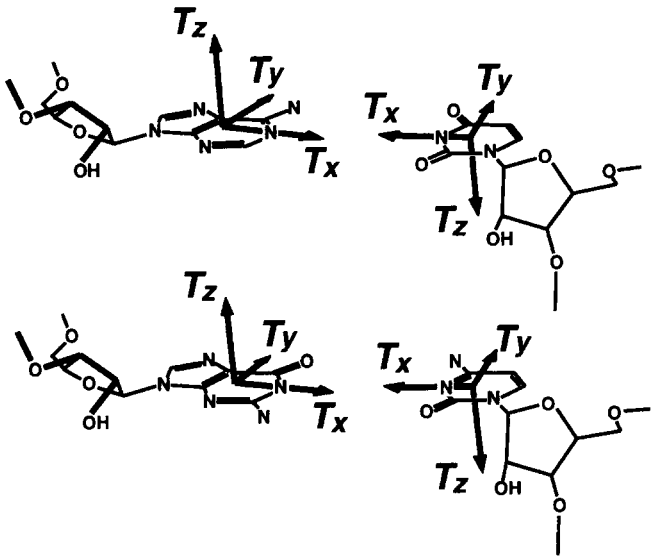

Figure 1. The definition of the center position and the orientation of each base. The center position of a base was simply defined as the mean position of all non-hydrogen atoms in the base. The orientation of a base was defined as three unit vectors, T_x, T_y and T_z.

Fig. 1. The center position was defined as the mean position of all non-hydrogen atoms in a base. The orientation of a base was defined as three vectors; the normal of the base plane (T_z), the direction from the center of the base to N1 for purine or N3 for pyrimidine (T_x), and the direction orthogonal to both the two directions and for the major groove side (T_y). Positions and orientations for all the 2828 residues were calculated by a program COB.

Calculation of Relative Position of a Target Base to a Reference Base

Based on the position and orientation, the relative positions of bases (target) to each of bases (reference) were calculated by a program RELP. Relative position of a target base was defined as a coordinate on the reference base coordinate system with the center position and orientation. Inter-base distance was defined as the inter-center position distances. A database was prepared for all the set of target and reference bases with the inter-base distance less than 30 Å. The database also includes distances from each base to its nearest C_α atom of protein components. Relative angle of a target base to a reference base was defined as the angle between the T_z vectors.

RESULTS

Distance and Angle Distribution of Bases

Figure 2 shows the relative location of bases in function of inter-base distance and the relative angle for the 5S and 23S RNA parts of the large subunit from *Haloarcula marismortui*[4] (1FFK). There are two types of peaks on the map: one group are almost parallel, with the angle less than 30 degrees, indicating most of the two bases belong to the same strand and the other group are almost anti-parallel, with the angle larger than 150 degrees, indicating most of the two bases belong to different strands to each other. For the bases belonging to the same strand, the nearest neighbor bases, located about 4 Å, are well concentrated (indicated by the arrow 'a' in the figure). However,

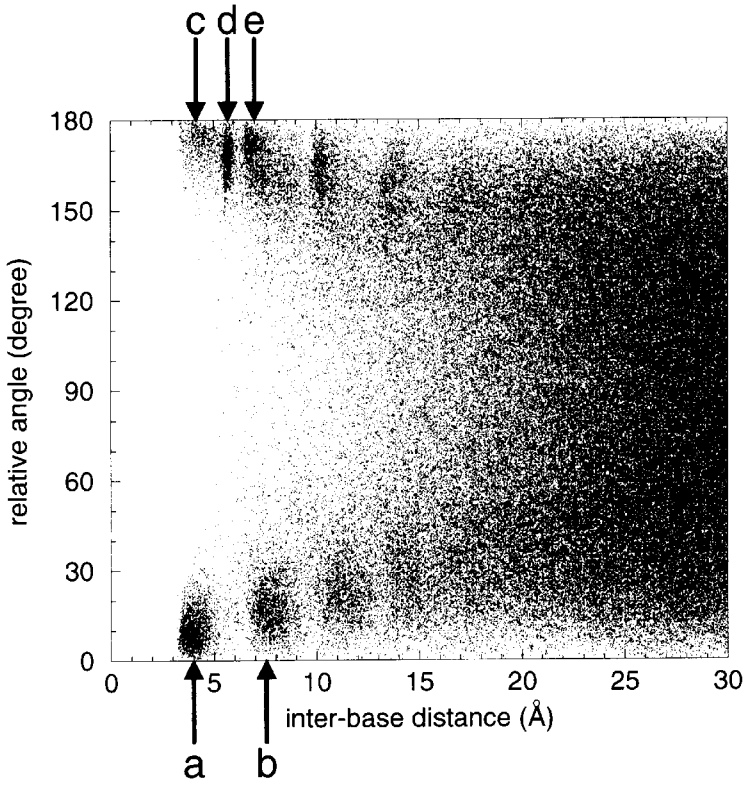

Figure 2. The relative location of bases in function of inter-base distance and the relative angle for the 5S and 23S RNA parts of the large subunit from *Haloarcula marismortui*[4] (1FFK). Inter-base distance was defined as the inter-center position distances. Relative angle was defined as an angle between the T_z vectors. The arrows indicate the spots corresponding to (a) the nearest neighbor bases, (b) the second nearest bases in the same strand with the reference base, and (c) the 3′ neighbors, (d) the bases forming base pairs with the reference bases, (e) the 5′ neighbors in the opposite strand.

the second nearest bases are dispersive (arrow b) and the third are almost dispersed, probably indicating a degree of bending of the stem structure. For the bases belonging to the different strands, the bases forming base pairs with the reference bases are located in the inter-base distance of 5.8 Å, forming a clear spot in the map (arrow d). The 5′ and 3′ neighbors of the paired bases are located in the distance of 7 and 4 Å, respectively, forming a dispersive spot (arrows e and c). The second neighbors of the paired base are almost dispersed. It was clearly found that there is no clear relation between bases apart more than 15 Å and, thus, the relative location of bases can be analyzed within 15 Å of the reference bases. The background in the distance between 5 to 15 Å must contain information concerning non-stem structure such as loop or tertiary interactions. It should be noted that few bases located close and perpendicular to the target base. Hereafter, analysis was performed for the reference bases which exist in a 20 Å cube centering on each reference base.

Base Distribution Around a Target Base

Figure 3 shows the distribution of relative positions of bases for the 5S and 23S RNA (Fig. 3a), P4-P6 domain of the group I intron (Fig. 3b), yeast $tRNA^{Phe}$ (Fig. 3c) and the RNA part of hammerhead ribozyme (Fig. 3d). Each three panel shows the projection along T_x (y-z plane), T_y (x-z plane) and T_z (x-y plane). Because the T_x is corresponding to the orientation of Watson-Crick type base pairing, a concentrated spot was clearly found in the center of the panel, indicating that the canonical base pair is abundant. Furthermore, periodically narrow islands were observed upper and lower sides of the center spot, showing the base–base stacking. In general, distribution patterns are almost same for the four RNA structures, and the distribution of bases around reference bases in the x-z plane is schematically shown in Fig. 4: bases close to the reference base are well concentrated and vice versa. The projection along T_y also shows the canonical base pairs with the distance of 5.2 Å and stacked bases in upper and lower sides. The projection along T_z, still shows the abundance of canonical base pair and this panel is further extracted as described below. The database produced from the 5S and 23S RNA parts was used for further analysis.

Extraction of Each Base Plane

In order to analyze the spatial relation between nearest neighbor bases, each base plane was extracted along planes parallel to the reference base: Fig. 5a and 5c show the location of the 3′ and 5′ nearest neighbor base pairs of the reference base, respectively, and Fig. 5b shows a cross section at the reference base. Clearly, in the Fig. 5b, a dense spot was observed at the

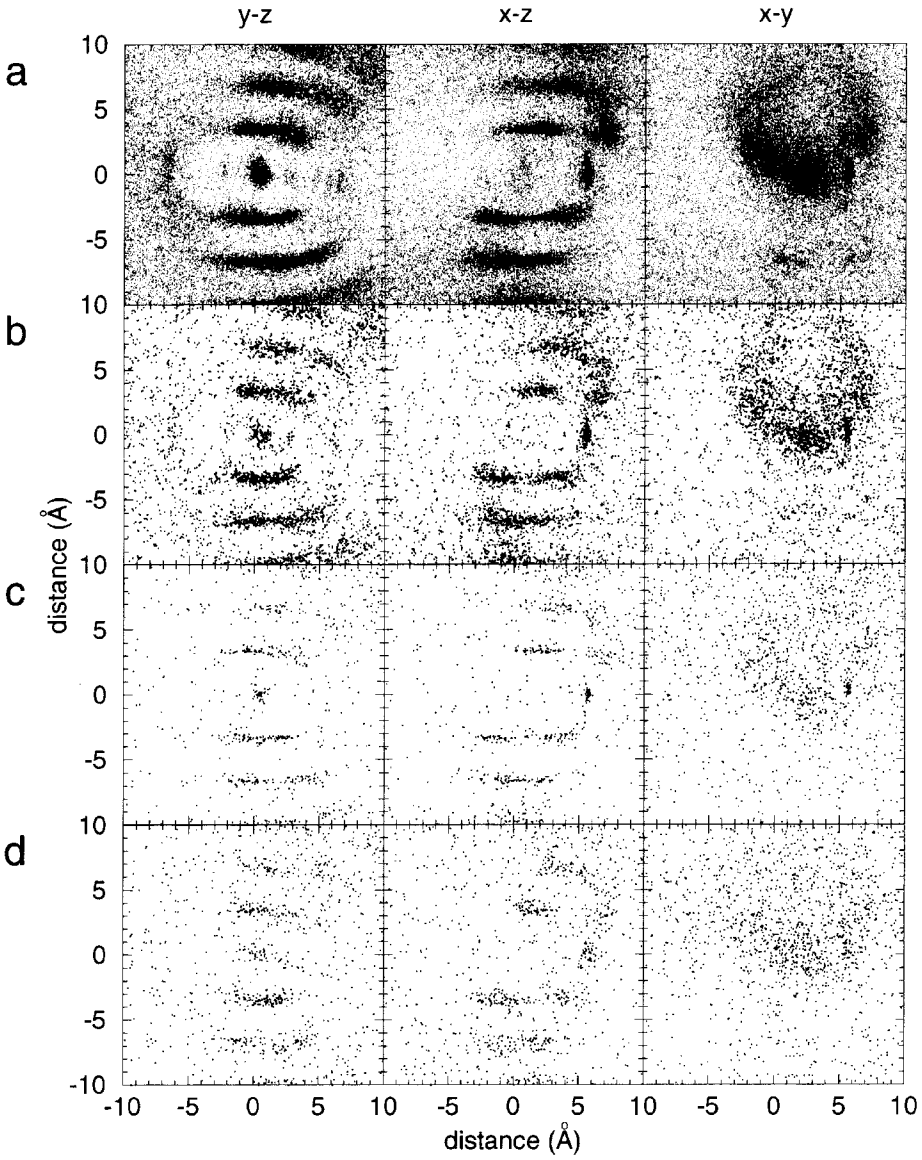

Figure 3. The distribution of relative positions of bases for (a) the 5 S and 23 S RNA[4] (1FFK), (b) P4-P6 domain of the group I intron[13] (1GID), (c) yeast $tRNA^{Phe}$ [14] (1EHZ) and (d) the RNA part of hammerhead ribozyme[15] (1HMH). Each three panels show the projection along T_x (y-z plane), T_y (x-z plane) and T_z (x-y plane) with a range of ± 10 Å from the reference base. The horizontal and vertical axes are y and z for the y-z plane, x and z for the x-z plane and x and y for the x-y plane.

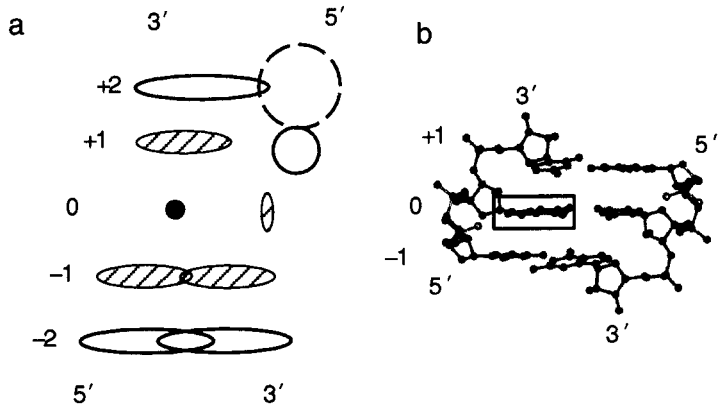

Figure 4. Schematic drawing of the distribution of bases corresponding to the projection along T_y (x-z plane) in Fig. 3 (a) and a molecular structure of the RNA-A type helix (b). Filled circle indicates the position of the reference base. Hatched and open circles indicate the first and second nearest bases, respectively. The dashed circle indicates the third nearest base. A box in the panel b indicates the reference base.

position corresponding to the Watson-Crick type base pair. The elliptic shape of the spot indicates that the distance between the Watson-Crick type base pair is well fixed, but the relative orientation is slightly swinging. As described below, the ideal position for the base forming the Watson-Crick type base pair is different for purine and pyrimidine. However, this difference is smaller than the positional fluctuation. Several weak spots are observed around the reference base indicating existence of non-Watson-Crick type base pairs all around. This was analyzed in detail as described below.

Analysis of Base Pairing

Figure 6b shows the extraction of the cross section shown in Fig. 5b according to the kinds of bases. The UA panel that shows relative positions of target adenine bases to reference uridine bases includes a clear spot corresponding to the Watson-Crick type U-A base pair. The CG, GC and AU panels include spots corresponding to the Watson-Crick type base pairs as well. On the other hand, some of panels contain spots corresponding to wobbling base pairs. The UG and GU panels contain clear spots indicating the existence of the G-U wobble base pairs. The spots corresponding to the G-U pair is well concentrated. The GA and AG panels contain spots corresponding to the shared G-A pairs. In this case, the spots are not well concentrated. The UU panel contains two spots, even though weak, corresponding to the U-U wobble base pairs. The AA panel contains three spots; two spots include the shared A-A pair and one spot includes the head-to-head A-A pair. The base density around the adenosine as the reference base is

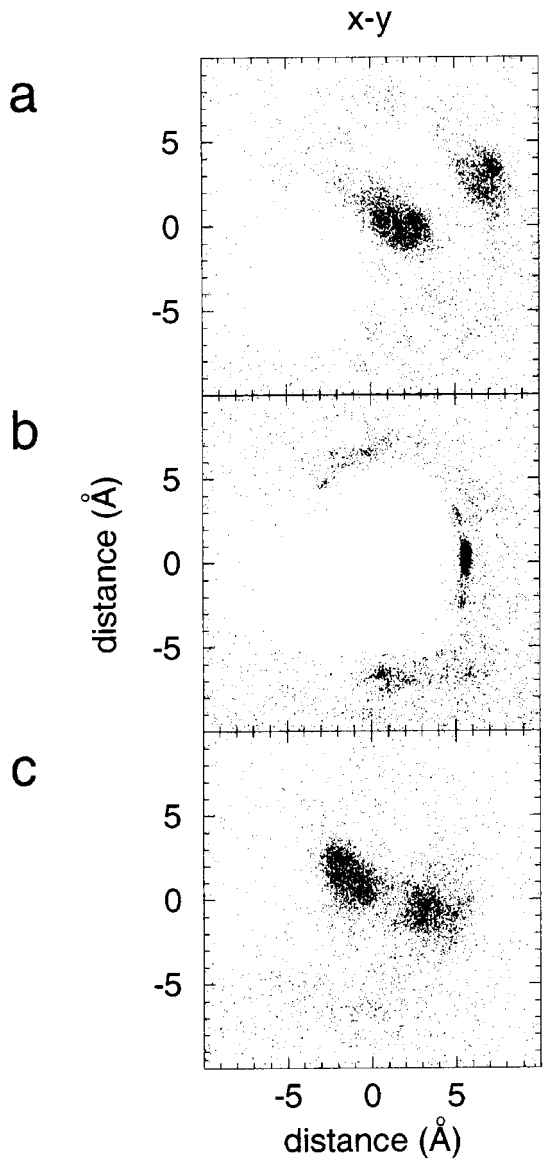

Figure 5. Distribution of bases in the planes parallel to the reference base for the 5 S and 23 S RNA. (a) The plane in the 3′ side of the reference base. (b) The plane including the reference base. (c) The plane in the 5′ side of the reference base. Each panel is a projection of 2.5 Å in the T_z direction.

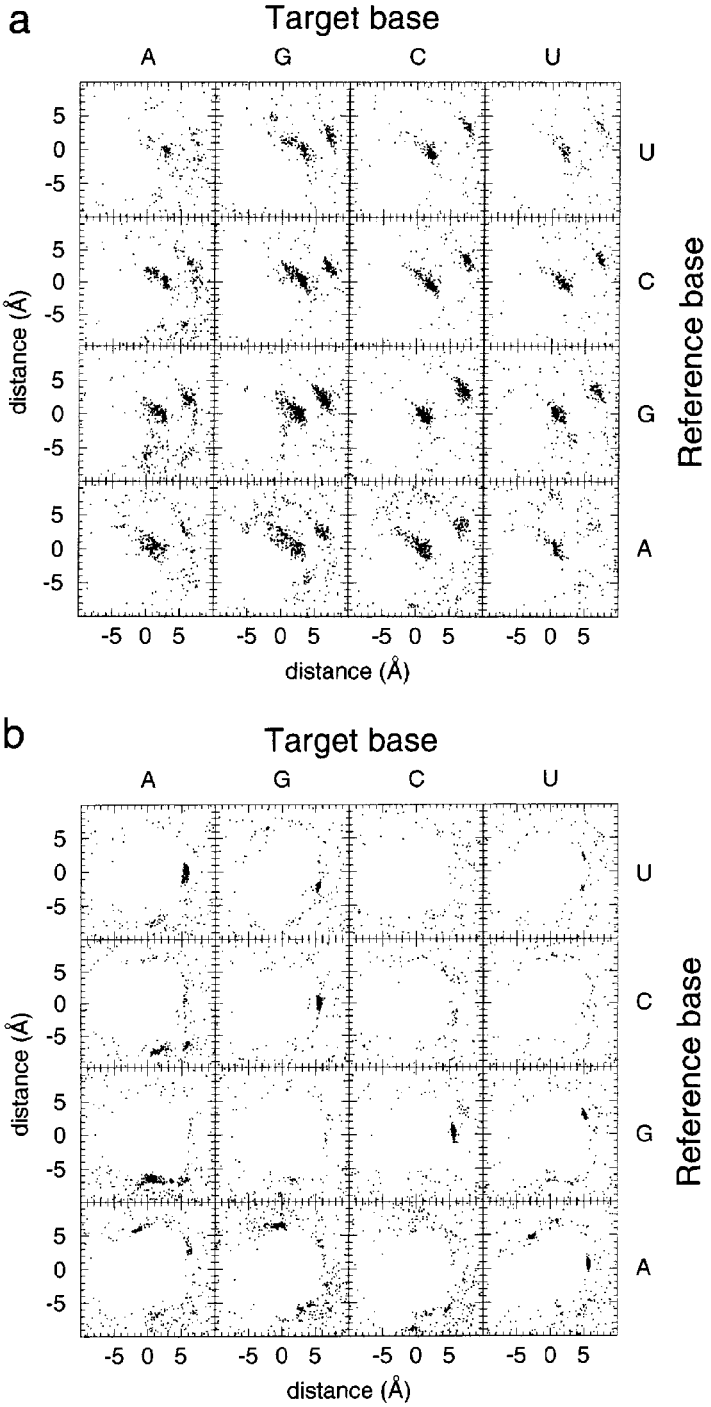

Figure 6. Extraction of each reference and target bases from Fig. 5. (a) The plane in the 3′ side of the reference base. (b) The plane including the reference base. (c) The plane in the 5′ side of the reference base.

(*continued*)

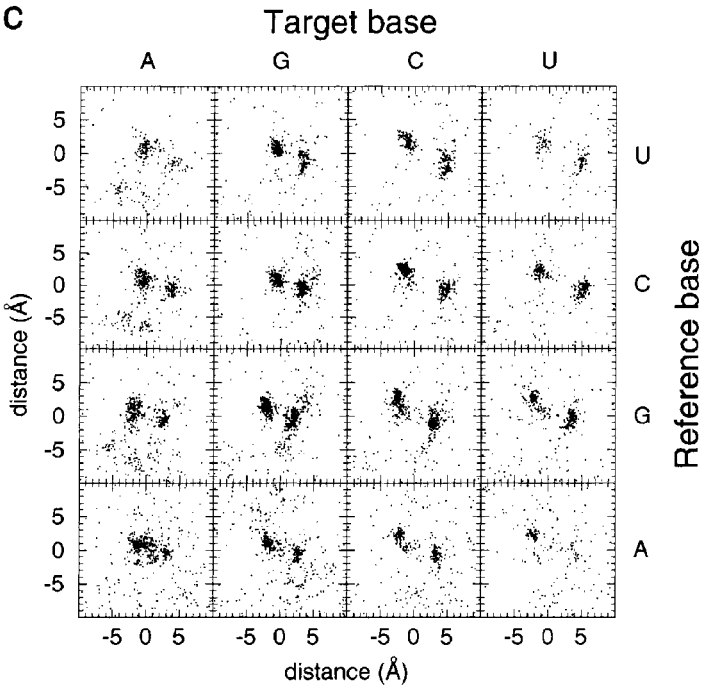

Figure 6. Continued.

clearly higher that other three nucleotides, indicating that adenosine residue is frequently involved in non-regular interactions.

Analysis of Base–Base Stacking

Figure 6a shows the extraction of 3′ stacked bases (Fig. 5a). Clearly, it was observed that the bases stacked to A or U are dispersive whereas the bases stacked to G or C are well concentrated. This character of stacking should be considered to build and/or evaluate RNA structures. Similar character was observed for the bases 5′ stacked to A (Fig. 6c). It is noted that, for Fig. 6a, stacked base of the same strand is well concentrated and that of the opposite strand is dispersed as found in the Fig. 3b.

Effect of Protein Interaction on the Distribution of Bases

Figure 7 shows the extraction of the cross section shown in Fig. 5b according to the distance between proteins and reference bases. Fig. 7a, 7b and 7c show the target base distributions around the reference bases located

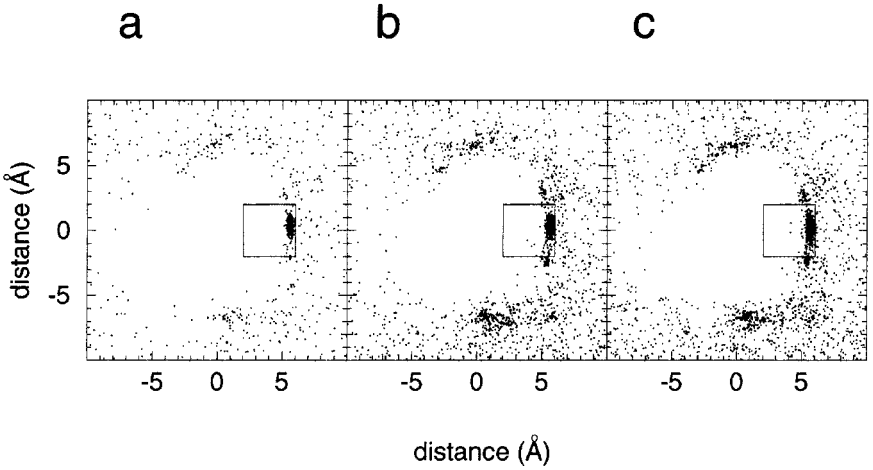

Figure 7. The extraction of the cross section shown in Fig. 5b according to the distance between proteins and reference bases. (a) The target base distributions around the reference bases located less than 8 Å, (b) between 8–12 Å and (c) larger than 12 Å to protein components. Distances between the center positions of nucleotide bases and the Cα of amino acid residues were used. The boxed regions corresponding to the Watson-Crick type base pairs were used to produce Table 1.

Table 1. Populations of Watson-Crick and Non-Watson-Crick Base Pairs Corresponding to Fig. 7a, 7b and 7c. The Numbers of Watson-Crick Type and Non-Watson-Crick Type Base Pairs Were Calculated as the Number of Points Inside and Outside the Small Boxes in the Fig. 7

	Watson-Crick	Non-Watson-Crick	Ratio
Near	406	738	1.82
Medium	591	1383	2.34
Far	573	1208	2.11

less than 8 Å, between 8–12 Å and larger than 12 Å to protein components, respectively. Table 1 shows the populations of Watson-Crick and non-Watson-Crick base pairs corresponding to Fig. 7a, 7b and 7c. It was found that population of non-Watson-Crick base pairs is decreased around protein components, suggesting that the protein components have a tendency to bind to the regular stem structures.

DISCUSSION

In the present study, the relative locations of bases in the ribosome crystal structure are analyzed in detail. Especially, the existence and

population of the wobble base pairs are well characterized. Furthermore, the stability of the base pairs and stacking interactions can be estimated by the distributions of the relative positions. For example, as described above, the relative location of the stacked A to the reference A is dispersed, probably indicating the instability of the A-A structure. Detailed analysis of the distribution patterns will give more information about the energy of the interactions which can be compared with the previous experimental and theoretical analyses, and can add more information to the database of RNA base pairs.[12,17]

By the analysis of ribosome crystal structures, new structural motifs, A-minor motif and kink-turn, have been found.[9,10] These motifs consist of more than two nucleotides and, thus, it seems to be difficult to be discriminated by the present system which analyzes the pair of bases. To find out such motifs, it will be useful to generate database of relative positions of bases to specific pair of bases such as Watson-Crick base pair rather than a base.

It was found that less non-Watson-Crick base pairs were found close to the protein-binding site, suggesting that the protein components have a tendency to bind to the regular stem structures. Further analyses, by combining the distribution of amino acid residues around nucleotide bases,[18–20] might be useful for elucidating the insight of RNA recognition by protein. Thus, the present database can be used for many aspects. For example, it will be useful to extract information concerning the tertiary interaction between stems and loops.

There is a possibility that the structure used in this analysis is biased by the force field used in the structural determination. However, this effect must be small and may not influence the results of the present analysis. To check the effect of structure determination systems, it might be useful to analyze different crystal structures from different groups. On the other hand, the database produced in the present study can be used for additional force field to calculate RNA structures. In fact, Kuszewski et al. introduce a database potential to NMR structure calculation of DNA.[21] The database is being prepared for downloading at http://www.ic.it-chiba.ac.jp/relp/.

ACKNOWLEDGMENTS

This work was supported by Grants-in-Aid for Scientific Research on Priority Areas (09278206) and High Technology Research from Ministry of Education, Science, Sports and Culture, Japan, and the ''Research for the Future'' Program (JSPS-RFTF97L00503) from the Japan Society for the Promotion of Science.

REFERENCES

1. Clemons, W.M., Jr; May, J.L.; Wimberly, B.T.; McCutcheon, J.P.; Capel, M.S.; Ramakrishnan, V. Structure of a Bacterial 30S Ribosomal Subunit at 5.5 å Resolution. Nature **1999**, *400*, 833–840.
2. Ban, N.; Nissen, P.; Hansen, J.; Capel, M.; Moore, P.B.; Steitz, T.A. Placement of Protein and RNA Structures into a 5 Å-resolution map of the 50 S Ribosomal Subunit. Nature **1999**, *400*, 841–847.
3. Culver, G.M.; Cate, J.H.; Yusupova, G.Z.; Yusupov, M.M.; Noller, H.F. Identification of an RNA-protein Bridge Spanning the Ribosomal Subunit Interface. Science **1999**, *285*, 2133–2136.
4. Ban, N.; Nissen, P.; Hansen, J.; Moore, P.B.; Steitz, T.A. The Complete Atomic Structure of the Large Ribosomal Subunit at 2.4 Å Resolution. Science **2000**, *289*, 905–920.
5. Schluenzen, F.; Tocilj, A.; Zarivach, R.; Harms, J.; Gluehmann, M.; Janell, D.; Bashan, A.; Bartels, H.; Agmon, I.; Franceschi, F.; Yonath, A. Structure of Functionally Activated Small Ribosomal Subunit at 3.3 Angstroms Resolution. Cell **2000**, *102*, 615–623.
6. Wimberly, B.T.; Brodersen, D.E.; Clemons, W.M., Jr; Morgan-Warren, R.J.; Carter, A.P.; Vonrhein, C.; Hartsch, T.; Ramakrishnan, V. Structure of the 30 S Ribosomal Subunit. Nature **2000**, *407*, 327–339.
7. Yusupov, M.M.; Yusupova, G.Z.; Baucom, A.; Lieberman, K.; Earnest, T.N.; Cate, J.H.; Noller, H.F. Crystal Structure of the Ribosome at 5.5 Å Resolution. Science **2001**, *292*, 883–896.
8. Harms, J.; Schluenzen, F.; Zarivach, R.; Bashan, A.; Gat, S.; Agmon, I.; Bartels, H.; Franceschi, F.; Yonath, A. High Resolution Structure of the Large Ribosomal Subunit from a Mesophilic Eubacterium. Cell **2001**, *107*, 679–688.
9. Nissen, P.; Ippolito, J.A.; Ban, N.; Moore, P.B.; Steitz, T.A. RNA Tertiary Interactions in the Large Ribosomal Subunit: The A-minor Motif. Proc. Natl. Acad. Sci. U.S.A. **2001**, *98*, 4899–4903.
10. Klein, D.J.; Schmeing, T.M.; Moore, P.B.; Steitz, T.A. The Kink-Turn: A New RNA Secondary Structure Motif. EMBO J. **2001**, *20*, 4214–4221.
11. Klosterman, P.S.; Tamura, M.; Holbrook, S.R.; Brenner, S.E. SCOR: A Structural Classification of RNA Database. Nucleic Acids Res. **2002**, *30*, 392–394.
12. Nagaswamy, U.; Larios-Sanz, M.; Hury, J.; Collins, S.; Zhang, Z.; Zhao, Q.; Fox, G.E. NCIR: A Database of a Non-canonical Interactions in known RNA Structures. Nucleic Acids Res. **2002**, *30*, 395–397.
13. Cate, J.H.; Gooding, A.R.; Podell, E.; Zhou, K.; Golden, B.L.; Kundrot, C.E.; Cech, T.R.; Doudna, J.A. Crystal Structure of a Group I Ribozyme Domain: Principles of RNA Packing. Science **1996**, *273*, 1678–1685.
14. Shi, H.; Moore, P.B. The Crystal Structure of Yeast Phenylalanine tRNA at 1.93 Å Resolution: A Classic Structure Revisited. RNA **2000**, *6*, 1091–1105.
15. Pley, H.W.; Flaherty, K.M.; McKay, D.B. Three-dimensional Structure of a Hammerhead Ribozyme. Nature **1994**, *372*, 68–74.
16. http://plasma-gate.weizmann.ac.il/Xmgr/
17. Leontis, N.B.; Westhof, E. Geometric Nomenclature and Classification of RNA Base Pairs. RNA **2001**, *7*, 499–512.

18. Allers, J.; Shamoo, Y. Structure-based Analysis of Protein-RNA Interactions Using the Program ENTANGLE. J. Mol. Biol. **2001**, *311*, 75–86.
19. Kuszewski, J.; Schwieters, C.; Clore, G.M. Improved Accuracy of NMR Structures of DNA by Means of a Database Potential of Mean Force Describing Base–base Positional Interactions. J. Am. Chem. Soc. **2001**, *123*, 3903–3918.
20. Jones, S.; Daley, D.T.A.; Luscombe, N.M.; Berman, H.M.; Thornton, J.M. Protein-RNA Interactions: A Structural Analysis. Nucleic Acids Res. **2001**, *29*, 943–954.
21. Nobeli, I.; Laskowski, R.A.; Valdar, W.S.J.; Thornton, J.M. On the Molecular Discrimination Between Adenine and Guanine by Proteins. Nucleic Acids Res. **2001**, *29*, 4294–4309.

Received April 23, 2002
Accepted June 3, 2002