Ribosome structure —A milestone of single crystal X-ray analysis—

Akihito Yamano*

1. Introduction

A ribosome is a vast assembly consisting of proteins and RNAs. The molecular mass adds up to approximately 2.5 MDa, and it is the largest asymmetric entity whose structure has been solved in human history. The first step of X-ray structure analysis is to acquire well-ordered crystals of the target material. Generally speaking, crystallization becomes more difficult as the size of the molecule increases. After this obstacle is overcome, one would then face the phase problem, a principle problem in crystallography. For such a huge complex, conventional phasing methods such as MAD and MIR may not work. Taking all of these issues into account, one can readily imagine that the structure determination of the ribosome must have been an extremely difficult task.

The 2009 Nobel Prize in Chemistry was awarded to those researchers[1] who determined ribosome structures, not only because of the difficulty of the structure determination but also because the resultant structure answered many biological questions—such as underlying chemistry in protein synthesis—and is providing the basis for new drug discovery. Only X-ray structure analysis can yield the atomic structure of such an enormous complex.

2. What is a ribosome?

Genetic information stored in DNA is transcribed to messenger RNA (mRNA) and then translated to proteins. This flow of genetic information is called the *central dogma*. All living organisms follow this rule except for retroviruses. The ribosome is the major player in the latter half of the central dogma; that is, translation of genetic information to proteins (Fig. 1).

DNA consists of two strings of bases. Each string is composed of four types of bases: adenine (A), guanine (G), cytosine (C) and thymine (T). A and G are purines, with two carbon–nitrogen rings. C and T are pyrimidines, with only one carbon–nitrogen ring. Due to the size of the molecules and the number of hydrogen bonds, A always pairs with T and G always pairs with C. DNA is usually double-stranded by the complementary parings of these bases. When it comes to protein production, a part of the DNA corresponding to the target protein splits open and the exposed gene is transcribed to a single-stranded messenger RNA (mRNA). All types of RNA contain the same bases as DNA except uracil (U), which substitutes for T.

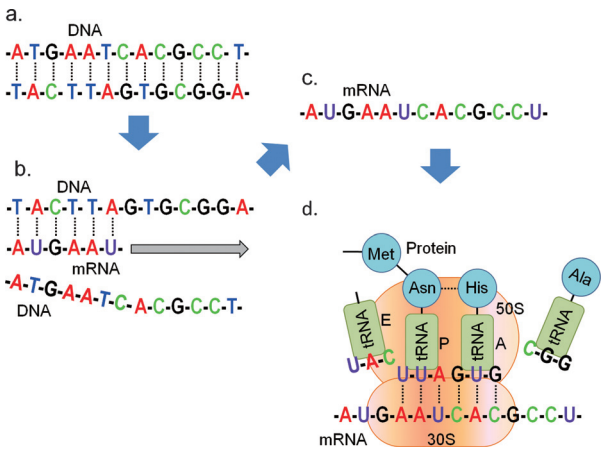

Fig. 1. The role of the ribosome in the "central dogma". (a) A double-stranded DNA. Genetic information is stored in DNA. DNA usually exists as a double-strand with complementary bases forming hydrogen bonds. (b) Transcription of genetic information from DNA to mRNA. The mRNA is synthesized using the exposed single-strand DNA as a template. DNA splits open prior to transcription. (c) The transcribed mRNA moves to a ribosome. (d) The codon and anticodon matching of mRNA and tRNA occurs at the 30S subunit of ribosome and an amino acid is added to the nascent protein by the dehydration-condensation reaction.

The genetic code in mRNA consists of three bases called codons. One codon stands for one amino acid. Proteins are composed of combinations of 20 amino acids. Therefore, three bases are sufficient ($4^3=64>20$) to code for a specific amino acid. For example, AUU in mRNA corresponds to isoleucine, AUU to asparagine. Some surplus codons encode the same amino acids and the "wobbling hypothesis" was proposed to explain this redundancy.

Genetic information presented as codons is matched to amino acids through transfer RNA (tRNA). The single chain of tRNA is folded into an L-shape by the DNA-like base paring within the chain, forming double-stranded regions and loops. An amino acid is attached at the beginning of the "L" and three bases corresponding to the amino acid are placed at the end of the "L". Those three bases, called an anticodon, are complementary to the codon of mRNA, and this is how the genetic information coded in DNA is precisely reflected in the protein. It is the ribosome that gets all of these components together to produce proteins, precisely

* Application Laboratories, Rigaku Corporation.

following genetic information originally coded in the DNA.

The ribosome performs some fairly complicated procedures during translation. It binds to mRNA, accepts a correct tRNA by codon–anticodon paring, and adds amino acids to the end of the nascent protein with high fidelity and at high speed, 20 amino acids per second. Therefore, the ribosome has to be a gigantic complex composed of numerous proteins and ribosomal RNA (rRNA) units.

3. Structure of ribosome

There are two basic types of ribosomes: prokaryotic (70S) and eukaryotic (80S). Prokaryotes are roughly equivalent to bacteria (except for eumycetes), and eukaryotes correspond to all other species. This manuscript refers to the 70S type because that is the only ribosome structure that has been determined by X-ray crystallography at present.

The 70S ribosome consists of two subunits, one large and one small. The 50S large subunit consists of 5S rRNA with 120 nucleotides, 23S rRNA with 2900 nucleotides, and 34 proteins. The 30S small subunit consists of 16S rRNA with 1540 nucleotides and 21 proteins. The shape of the 50S subunit resembles an armchair, and the 30S subunit rests on the 50S subunit.

Even when the crystal structure determination of the ribosome seemed like a distant dream, biologists attacked the ribosome structure using biophysical and biochemical methods: cryo-EM, neutron diffraction, ultra centrifuge, cross linking and so on. The ribosome structure was so important that people could not wait for the X-ray structure. Cryo-EM elucidated an overall shape of the ribosome at 20 Å resolution. This information was used to calculate phases for the first low resolution X-ray structure[2]. Neutron diffraction can distinguish deuterium from hydrogen, which can be used to measure intermolecular distances. Target ribosomal proteins labeled with deuterium were incorporated into a ribosome by reconstruction and small angle neutron scattering was performed to determine distances between labeled protein units. Ultra centrifuge can also be used to estimate approximate distances among proteins, along with molecular weights. In 1988, Capel and Moore presented a 3D map of all 21 proteins determined by neutron diffraction[3]. In the cross linking technique, cross linkers, are applied to the ribosome before it was chemically disassembled to its components. The resulting material was analyzed by ultra centrifuge to identify proteins having sedimentation speeds different from those of the original proteins. Proteins coupled by a linker should be located relative to each other in the ribosome within the length of the linker. Even before the X-ray structure analysis was done, the combination of these methods provided the relative arrangement of the ribosomal proteins. However, detailed chemistry of ribosome function, and the backbone of the ribosome structure, whether protein or rRNA, remained unknown. A number of questions about the ribosome were answered only when the atomic resolution structure of the ribosome was determined by X-ray crystallography.

4. High resolution X-ray structure of the ribosome

The pioneering work in ribosome crystallization was carried out by Yonath's group[4]. In 1980, they succeeded in obtaining crystals of the 50S large subunit for the first time. Yonath and co-workers used the 50S subunit from thermophile, *Bacillus stearothermophilus*. They observed the 45 Å structure under an electron microscope. Subsequently, they explored crystallization conditions and searched for bacteria that had ribosomes likely to yield good crystals. They gradually but steadily extended the maximum resolution: 9–18 Å in 1984[5], 6 Å in 1987[6] and finally 3 Å in 1991[7]. They used the 50S subunit of halobacteria, *Haloarcula marismortui.* In 2000, Yonath's group published the structure of the 30S subunit of *Thermus thermophilus*[8] along with Steitz and Ramakrishnan. Yonath succeesfully determined the 50S subunit structure in 2001[9], but her most significant contribution was to prove that ribosomes could be crystallized. To form a crystal, the target material must be homogeneous in solution. In the 1980s, it was believed that there were multiple subtypes of ribosomes due to its various functions. Yonath's group attempted the crystallization of the ribosome and, after a large amount of effort, they succeeded, inspiring other researchers devoted to the structure determination of the ribosome.

It was Steitz's group that first succeeded in determining the structure of the 50S subunit of the ribosome[2]. Crystals of the 50S subunit of *H. marismortui* diffracted to 3 Å at the NSLS X12c and X12b beamlines, but data collection was limited to 7 Å, probably because they had to use a long crystal-to-detector-distance because of the large unit cell. As a result, the high angle reflections were beyond the detector aperture. Additionally, they had to use a long X-ray wavelength in order to measure anomalous scattering. Initial phases were determined by the MIRAS method. The most popular method for phase determination of novel proteins is the MAD or SAD methods, which use selenomethionine biochemically introduced into the target protein. However, about 2/3 of the ribosome is rRNA and ribosomal protein has low methionine content. Preliminary calculations revealed that it was impossible to determine initial phases by the MAD method, therefore they had to use the MIRAS method. To calculate phases by MIRAS, it is necessary to determine heavy atom position. They first used the low-resolution cryo-EM structure to locate heavy atoms, and then used the resulting structure to calculate phases to obtain heavy atoms on a difference Fourier map. The resulting electron density map was 9 Å resolution; that is, the resolution jumped to 9 Å from the 20 Å cryo-EM structure.

Crystallization of the 30S subunit was pursued mainly by two groups: Trakhanov's group in Russia[10] and

Yonath's group. Trakhanov and co-workers succeeded in crystallizing both the 30S subunit and the 70S ribosome of *T. thermophilus*. Yonath's group crystallized the 30S subunit, the 50S subunits and the 70S ribosome of *E. coli, B. stearothermophilus* and *H. marismortui*[11]. Crystallization procedures used to obtain high-resolution X-ray structure were established basically by Trakhanov's group and Yonath's group. In the field of structural biology, those who crystallize a protein first usually report the structure first. However, the ribosome structure was an exception. The structure of 50S was first reported by Steitz's group at Yale and that of 30S was published by Ramakrishnan's group from MRC.

In August of 1999, the 5 Å 30S subunit structure of *T. thermophilus* and 5.5 Å 50S subunit structure of *H. marismortui* were published in the same issue of Nature[12][13]. The following year, 2000, was memorable for ribosome structure. Steitz et al. succeeded in determining the 2.4 Å resolution structure of the 50S subunit[14] and Ramakrishnan et al. did a 3 Å structure of the 30S subunit[15]. Twenty years after the first crystallization of the 50S ribosome by Yonath's group, the ribosome structure was determined at atomic resolution.

In 2001, Yusupov et al. determined the entire 70S ribosome structure to 5.5 Å resolution[16] and in 2005 they improved the resolution to 3.5A[17]. In 2006, the complex among tRNA, mRNA and 70S ribosome was elucidated at 2.8 Å resolution[18]. This structure clearly presents the interaction among ribosome, mRNA and tRNA at atomic level (Fig. 2).

If the scientific developments necessary between 1980 and 2000 were limited to ribosome crystallization, we would have seen the ribosome structure sooner. Fundamental technologies to tackle extremely difficult structures were also dramatically improved during this period, including improvements to synchrotron sources, area detectors, cryo-crystallography, MIR methods using metal cluster, structure analysis software and computers. In 1980, these were either not available or limited compared to the present level. In the 1980 paper, Yonath used a 1.5 kW sealed tube X-ray generator to check the durability of the 50S subunit crystal against X-rays[4].

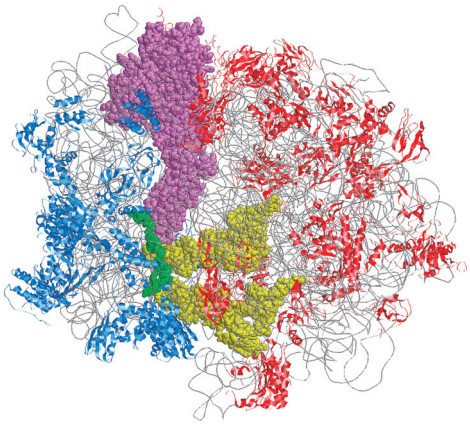

Fig. 2. The structure of a 70S ribosome complexed with mRNA (green) and tRNA (yellow). The rRNAs, proteins of the 30S subunit and those of the 50S subunit are shown in gray, blue and red, respectively. 2/3 of ribosome is occupied by rRNA and the backbone of the ribosome structure is formed by the rRNAs. X-ray structure analysis confirmed the ribosome structure derived from other analytical methods. Additionally the underlying chemistry of catalysis, proofreading to discriminate wrong tRNAs and mechanisms of the "wobbling hypothesis" are also elucidated at the atomic level.

5. What the ribosome structure elucidated

The crystallographic results confirmed that it was not proteins but rRNA that form the backbone of the ribosome structure. Proteins fill the gaps in the folded rRNA; therefore, ribosomal proteins often have a spherical domain and an extended domain to stabilize themselves by sticking the rod part into rRNA. Some proteins have multiple extended domains.

Another feature is that the interface between the 30S and 50S subunits, especially the part that binds mRNA and tRNA, is free of proteins (Fig. 3). Additionally, the variation of structural motifs seems to be limited, because the secondary structure motifs seen in the 50S rRNA are also seen in the 30S rRNA. The most outstanding difference between the 30S and 50S subunits is the structures of the rRNA. The 16S rRNA of the 30S subunit has 4 localized domains, while the 6 domains of the 23S rRNA of the 50S subunit encompass the entire subunit, mutually interlocking[19].

Ribosomes catalyze the polymerization of amino acids, and the peptide transferase reaction center is located on the 50S subunit. Since the ribozyme (RNA having enzymatic activity) was found, rRNA had been thought to catalyze the peptide transfer reaction. By analyzing the structure of 50S ribosome complexed with a substrate analogue, Steitz and co-workers proved that the peptide transferase reaction is catalyzed only by tRNA[20]. The ribosome was truly a ribozyme. The catalytic mechanism was understood as the reverse reaction of the deacylation of the serine protease, with the serine residue breaking peptides.

Ribosomes synthesize protein precisely following genetic information on mRNA. The accuracy of the translation is known to be higher than that achieved by

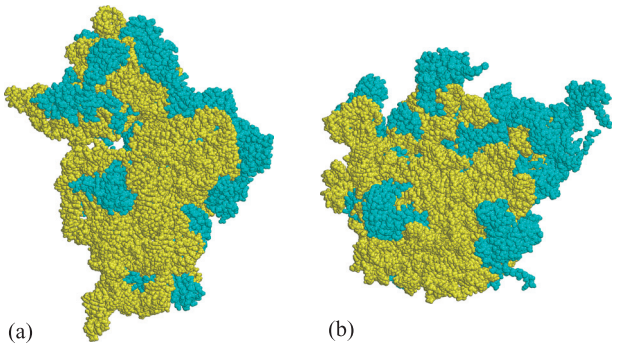

Fig. 3. The overall arrangement of rRNAs and proteins at the intersubunit interface between the 30S (a) and the 50S (b) subunits.

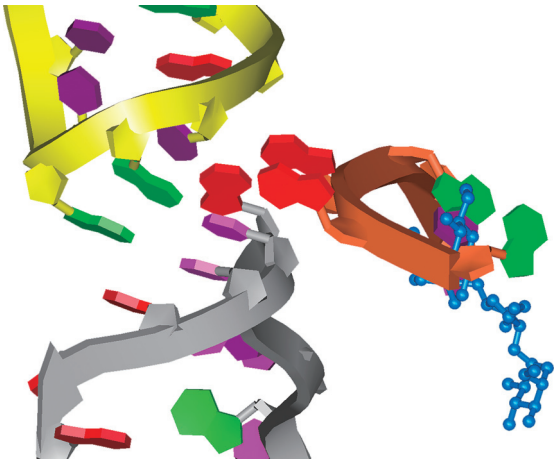

Fig. 4. The structure in the vicinity of the codon–anticodon pairng site in the 16S rRNA of the 30S subunit (brown), mRNA (gray), tRNA (yellow) and paromomycin (blue) complex. Adenine, guanine, cytosine and uracil are in red, green, pink and purple, respectively. The side chain of A1492 and A1493 residues proofreading the codon–anticodon matching are protruding toward A–C–C and U–G–G due to the binding of the paromomycin molecule.

the codon–anticodon interaction between mRNA and tRNA. This suggested that ribosomes have a proofreading mechanism. This mechanism became clear from the structure of a ribosome complexed with an antibiotic, paromomycin[19]. Paromomycin is a medicine used to treat intestinal infections and binds to the decoding site of the 16S ribosomal RNA of the 30S subunit. It was known that the fidelity of translation deteriorates when paromomycin binds to ribosomes. When a correct tRNA binds to mRNA, A1492 and A1493 change their conformation to participate in the codon–anticodon recognition for proofreading. However, when paromomycin is bound, those residues were fixed to the proofreading position. As a result, a similar but different tRNA can form hydrogen bonds; therefore, the precision of translation worsens (Fig. 4).

Ribosomes have three tRNA binding sites: A (acceptor), P (peptidyl) and E (exit). Interactions among ribosome, tRNA and mRNA were also understood at an atomic level[18]. The matching and proofreading of the codon–anticodon of mRNA and tRNA occurs at the A site. The tRNA judged as a correct one is allowed to move to the P site. The amino acid binding part of the tRNA goes into the peptidyltransferase site and an amino acid is added to the carboxyl terminal of the nascent protein. Though the codon and anticodon are still bound, the ribosome no longer proofreads but rather enforces capturing the correct tRNA. The existence of the E site was doubted. The clear observation of electron density corresponding to a deacylated tRNA in the high-resolution X-ray structure analysis of the 70S and tRNA complex ended the argument. The major component of the E site is protein instead of rRNA, and the strong interaction between tRNA and 16S rRNA seen at the A and P sites is absent. The codon–anticodon bond is proved to be impossible by checking interatomic distances.

A number of other mechanisms are elucidated from the X-ray structure of the ribosome. For example, the "wobbling hypothesis"—that is, the fact that more than one codon codes for one amino acid—is understood at an atomic level[21].

6. Summary

The high-resolution X-ray structure of the ribosome elucidated detailed mechanism of protein synthesis at an atomic level. This is truly a milestone of X-ray structure analysis and an example of the technique's strength. It is beyond imagination how far X-ray structure analysis is going to contribute to structural biology, attempting to elucidate not only the structure and function of the ribosome but also the mechanism of life. The 40S small subunit, the 60S large subunit and eventually the entire 80S ribosome structure of eukaryotes will probablybe published in the future.

However, this brilliant achievement of the ribosome structure analysis is only a small part of the entire biological phenomena. For example, why a child resembles its parents remains unknown. Molecular genetics and structural biology have given answers only to a small portion of life. There remains an unlimited number of questions that X-ray structure analysis needs to answer.

References

(1) http://nobelprize.org/nobel_prizes/chemistry/laureates/2009/press.html for example.

(2) N. Ban, B. Freeborn, P. Nissen, P. Penczek, R. A. Grassucci, R. Sweet, J. Frank, P. B. Moore and T. A. Steitz: *Cell*, **93** (1998), 1105–1115.

(3) M. S. Capel and P. B. Moore: *J. Appl. Cryst*., **21** (1988), 823–827.

(4) A. Yonath, J. Mussig, B. Tesche, S. Lorenz, V. A. Erdmann and H. G. Wittmann: *Biochem Int.*, **1** (1980), 428–435.

(5) A. Yonath, H. D. Bartunik, K. S. Bartels and H. G. Wittmann: *J. Mol. Biol*., **177** (1984), 201–206.

(6) M. Shoham, H. G. Wittmann and A. Yonath: *J. Mol. Biol*., **193** (1987), 819–822.

(7) K. von Bohlen, I. Makowski, H. A. S. Hansen, H. Barlels, Z. Berkovitch-Yellin, A. Zaytzev-Bashan, S. Meyer, C. Paulke, F. Franceschi and A. Yonath: *J. Mol. Biol*., **222** (1991), 11–15.

(8) F. Schluenzen, A. Tocilj, R. Zarivach, J. Harms, M. Gluehmann, D. Janell, A. Bashan, H. Bartels, I. Agmon, F. Franceschi and A. Yonath: *Cell*, **102** (2000), 615–623.

(9) J. Harms, F. Schluenzen, R. Zarivach, A. Bashan, S. Gat, I. Agmon, H. Bartels, F. Franceschi and A. Yonath: *Cell*, **107** (2001), 679–688.

(10) S. D. Trakhanov, M. M. Yusupov, S. C. Agalarov, M. B. Garber, S. N. Ryazantsev, S. V. Tischenko and V. A. Shirokov: *FEBS Lett*, **220** (1987), 319–322.

(11) A. Yonath, C. Glotz, H. S. Gewitz, K. S. Bartels, K. von Bohlen, I. Makowski and H. G. Wittmann: *J. Mol. Biol*., **203** (1988), 831–834.

(12) W. M. Clemons Jr, J. L. C. May, B. T. Wimberly, J. P. McCutcheon, M. Capel and V. Ramakrishnan: *Nature*, **400** (1999), 833–840.

(13) N. Ban, P. Nissen, J. Hansen, M. Capel, P. B. Moore and T. A. Steitz: *Nature*, **400** (1999), 841–847.

(14) N. Ban, P. Nissen, J. Hansen, P. B. Moore and T. A. Steitz:

Science, **289** (2000), 905–920.
(15) B. T. Wimberly, D. E. Brodersen, W. M. J. Clemons, R. Morgan-Warren, C. von Rhein, T. Hartsch and V. Ramakrishnan: *Nature*, **407** (2000), 327–339.
(16) M. M. Yusupov, G. Z. Yusupova, A. Baucom, K. Lieberman, T. N. Earnest, J. H. D. Cate and H. F. Noller: *Science*, **292** (2001), 883.
(17) B. S. Schuwirth, M. A. Borovinskaya, C. W. Hau, W. Zhang, A. Vila-Sanjurjo, J. M. Holton and J. H. D. Cate: *Science*, **310** (2005), 827–834.
(18) M. Selmer, C. M. Dunham, F. V. Murphy, A. Weixlbaumer, S. Petry, A. C. Kelley, J. R. Weir and V. Ramakrishnan: *Science*, **313** (2006), 1935–1942.
(19) V. Ramakrishnan and P. B. Moore: *Curr Opin Struct Biol*, **10** (2001), 144–154.
(20) P. Nissen, J. Hansen, N. Ban, P. B. Moore and T.A. Steitz: *Science*, **289** (2000), 920–930.
(21) V. Ramakrishnan: *Biochem. Soc. Trans*., **36** (2008), 567–574.