

ARTICLE

A New Approach for Building an Atomic Model from a Three-Dimensional Electron Microscopy Data

Atsushi MATSUMOTO*

Japan Atomic Energy Agency, 8-1-7, Umemidai, Kizugawa, Kyoto 619-0215 Japan

We have developed a new approach to build an atomic model from a low-resolution three-dimensional electron microscopy data for analyzing the structures of 70S ribosome. In this approach, a huge number of atomic models with different conformations are prepared first, and then the best-fitting atomic model is selected from them for each electron microscopy structure. The rigid-body fitting calculation, in which the atomic model is moved as a rigid body by translational and rotational manipulation, is most intensively performed. However, the calculation takes quite long if all the six rigid-body parameters are treated as variables at the same time. Thus, we perform two different calculations alternatively, in which only three parameters are treated as variables and the others are fixed. We have ascertained that the calculation is fast and accurate enough.

KEYWORDS: *ribosome, normal mode analysis, rigid-body fitting, electron microscopy, atomic model, elastic network model*

I. Introduction

Ribosome is a huge macromolecular complex responsible for protein synthesis in all organisms. There are four essential steps during protein synthesis in ribosome; initiation, elongation, termination, and recycling. In the initiation step, large 50S subunit and small 30S subunit come together, bind to mRNA and form an initiation complex of the 70S ribosome. In the elongation step, ribosome synthesizes proteins by translating the base sequence information on mRNA to the amino-acid sequence. In the termination step, protein synthesis is terminated and the newly synthesized peptide chain is released. Finally, in the recycling step, ribosome dissociates into two subunits for a new cycle of protein synthesis.

To understand the molecular mechanism of the protein synthesis, it is important to know the three-dimensional (3D) structures of ribosome in functional states. So far, the structures of ribosome in a few functional states are determined at the atomic resolution by X-ray crystallography. However, they are not enough for understanding the whole molecular mechanism of protein synthesis.

Recent progress in electron microscopy (EM) has made it possible to capture 3D structures of ribosome at various functional states. From the comparison of these EM structures, functionally important conformational changes have been identified.¹⁾ However, because of the low resolution nature of the 3D-EM structures, it is not so easy to compare them directly. The atomic model, which is described by the positions of constituting atoms, is more suited for structural comparison. Thus, we have built an atomic model for each EM structure of ribosome. The results about the comparison

of the 42 structures were reported in our previous paper.²⁾ In this paper, the methodology for building the atomic model is described in detail.

II. Method

1. Building EM Structure from EM Density Data

The data for 3D-EM structure is available from the Electron Microscopy DataBank (EMDB, <http://emdatbank.org>) at the European Bioinformatics Institute (EBI). It is given as the density ρ at each voxel, which occupies a space (x, y, z) , where $i_x p \leq x < (i_x + 1)p$, $i_y p \leq y < (i_y + 1)p$, $i_z p \leq z < (i_z + 1)p$, and i_x , i_y , and i_z are integers, and p is the voxel size given as an experimental value. From the data, we build an EM structure as follows. We define that each EM structure is composed of voxels whose density ρ is higher than a threshold value ρ_0 . The threshold value ρ_0 is determined so that the volume occupied by the EM structure ($= N_v p^3$, where N_v is the number of voxels with a density higher than ρ_0) is equivalent to the molecular volume of the X-ray crystal structure of the 70S ribosome,³⁾ calculated by a program VOIDOO,⁴⁻⁶⁾ in which the probe radius is set to 1.4 Å.

2. New Approach for Building Atomic Model from EM Structure

Our approach to build an atomic model from each EM structure consists of two steps. In the first step, we build many atomic models with different conformations as a library by deforming the X-ray crystal structure. In the second step, the best-fitting atomic model is selected from them for each EM structure by rigid-body fitting calculation.

In the usual approach, which we call the “tailor-made” approach, time-consuming calculation for deforming the

*Corresponding author, E-mail: matsumoto.atsushi@jaea.go.jp

X-ray crystal structure so that it fits better into the EM structure is performed for each EM structure. On the other hand, in our “ready-made” approach, once the library of the deformed atomic models is prepared, only the rigid-body fitting calculation is necessary which require much less computational time. When we study many EM structures of the same target, our approach can save computational time.

3. Atomic Model Library

To build deformed atomic models, we employ normal mode analysis. In the normal mode analysis, molecular deformation can be described by a sum of normal modes;

$$\mathbf{r} - \mathbf{r}_0 = \sum_k \alpha_k \mathbf{v}_0^k, \quad (1)$$

where \mathbf{r} and \mathbf{r}_0 are $3N$ -dimensional vectors, where N is the number of atoms in the molecule, describing the deformed and energy-minimized atomic structures, respectively, \mathbf{v}_0^k is the k th normal mode vector for the structure \mathbf{r}_0 , and α_k the coefficient. In this study, we use Hookean pairwise potential as an energy function E ;

$$E = \frac{1}{2} \sum_{(i,j)} C (d_{ij} - d_{ij}^0)^2, \quad (2)$$

where d_{ij} is the distance between atom i and j , the zero superscript indicates the initial structure, and C is the coefficient. Note that the energy function is defined in such a way that the initial structure has the minimum energy. The summation is taken over all pairs of atoms (i, j) , whose distances in the initial structure are less than a cutoff distance, which depends on the atom types. Due to the form of this energy function, the model is called an elastic network model.

Equation (1) is a good approximation only when the deformation α_k is small. When it is large, the molecule does not deform along the low-energy path. Using an elastic network model, Miyashita et al.⁷⁾ showed that atomic structures with large deformations along the low-energy path could be obtained by applying small deformations along the normal mode iteratively. In a similar way, we obtain a series of deformed atomic structures \mathbf{r}_n^k along the k th normal mode of the X-ray crystal structure as follows; We start from the X-ray crystal structure \mathbf{r}_0 and perform the normal mode analysis, in which the structure \mathbf{r}_0 is regarded as the initial structure in Eq. (2). Using the normal mode vector \mathbf{v}_0^k ($|\mathbf{v}_0^k| = 1$) for the k th lowest frequency normal mode, a deformed atomic structure \mathbf{r}_1^k is obtained as $\mathbf{r}_0 + \alpha \mathbf{v}_0^k$, where $\alpha (> 0)$ is a constant and determines how far the structure is deformed. If we further regard the deformed structure \mathbf{r}_1^k as the initial structure in Eq. (2), another normal mode analysis can be performed and normal mode vectors are obtained. Among them, the one $\mathbf{v}_1^{k_1}$ which is most similar to \mathbf{v}_0^k ($\mathbf{v}_0^k \cdot \mathbf{v}_1^{k_1} \square 1$) is used to build the next structure \mathbf{r}_2^k ($= \mathbf{r}_1^k + \alpha \mathbf{v}_1^{k_1}$). Note that k_1 is not always equal to k , because the order of the normal modes could be different for the deformed structure. The calculation is repeated, and a series of structures $\{ \mathbf{r}_n^k; n = 1, \dots, n_+ \}$, where

$\mathbf{r}_n^k = \mathbf{r}_0 + \alpha \mathbf{v}_0^k + \alpha \mathbf{v}_1^{k_1} + \alpha \mathbf{v}_2^{k_2} + \dots + \alpha \mathbf{v}_{(n-1)}^{k_{(n-1)}}$, is obtained. Another series of structures is built for the opposite direction, that is, starting from the X-ray crystal structure \mathbf{r}_0 , a deformed structure \mathbf{r}_{-1}^k ($= \mathbf{r}_0 - \alpha \mathbf{v}_0^k$) is built. The calculation is repeated and another series of structures $\{ \mathbf{r}_{-n}^k; n = 1, \dots, n_- \}$ is obtained. We refer to these two series of structures (\mathbf{r}_n^k and \mathbf{r}_{-n}^k) as deformed structures along the k th normal mode. In this study, we set the value of the constant α to 0.1, which means that the root-mean-square displacement (RMSD) between \mathbf{r}_{n-1}^k and \mathbf{r}_n^k is 0.1 Å.

Using the deformed atomic structures \mathbf{r}_n^k , we replace Eq. (1) with

$$\mathbf{r} - \mathbf{r}_0 = \sum_k (\mathbf{r}_{n_k}^k - \mathbf{r}_0) \quad (3)$$

for large molecular deformation. Note that \mathbf{r} in Eq. (1) is determined by real numbers $(\alpha_1, \alpha_2, \alpha_3, \dots)$, while \mathbf{r} in Eq. (3) is by integers (n_1, n_2, n_3, \dots) . Thus, a ready-made atomic structure \mathbf{r} in Eq. (3) is specified by the combination of the integers (n_1, n_2, n_3, \dots) .

4. Scoring Function

To build the best-fitting atomic model from each EM structure, a scoring function is necessary. The scoring function is an index to describe quantitatively how well each atomic model is fitted into the EM structure. We define the scoring function as a ratio of the number of atoms inside the space occupied by the EM structure to the total number of atoms. Usually, more refined scoring functions are used in the EM fitting calculations.⁸⁾ Our simple scoring function is chosen mainly for the fast rigid-body fitting calculation.

The scoring function is used for two purposes. Given an EM structure and an atomic model, the scoring function is maximized by the rigid-body fitting calculation to find the optimum position and orientation of the atomic model with respect to the EM structure. Then, the atomic model with the highest fitting score among the atomic models in the library is regarded as the best-fitting atomic model for the EM structure.

5. Rigid-Body Fitting Calculation

In the rigid-body fitting calculation,⁸⁾ the EM structure is fixed in space and only the atomic model is moved by rotational and translational manipulation. We perform two different calculations successively for the rigid-body fitting. In the first calculation, the atomic model is moved so that the principal axes of inertia for the atomic model and the EM structure overlap each other. If the two rigid bodies were exactly the same, this manipulation would make them overlap each other perfectly. Of course, they are not the same, and the position and orientation of the atomic model is further refined in the second calculation, where the atomic model is iteratively rotated and translated so that the fitting score is maximized. The details of the two calculations are described below.

(1) Rigid-Body Fitting Using Principal Axes of Inertia

The inertial tensor \mathbf{I} for an atomic model is described as follows;

$$\mathbf{I} = \begin{pmatrix} \sum m_i(y_i^2 + z_i^2) & -\sum m_i x_i y_i & -\sum m_i x_i z_i \\ -\sum m_i y_i x_i & \sum m_i(x_i^2 + z_i^2) & -\sum m_i y_i z_i \\ -\sum m_i z_i x_i & -\sum m_i z_i y_i & \sum m_i(x_i^2 + y_i^2) \end{pmatrix}, (4)$$

where (x_i, y_i, z_i) is the position of atom i , and m_i the mass. The origin of the coordinate system \mathbf{o} is set on the center of mass of the atomic model. The principal axes of inertia $(\mathbf{e}_x, \mathbf{e}_y, \mathbf{e}_z)$ is obtained as the eigenvectors of the inertia tensor \mathbf{I} , that is, $\mathbf{I}\mathbf{e}_x = I_1\mathbf{e}_x$, $\mathbf{I}\mathbf{e}_y = I_2\mathbf{e}_y$, and $\mathbf{I}\mathbf{e}_z = I_3\mathbf{e}_z$, where I_1, I_2 and I_3 are inertial moments and we assume $I_1 \leq I_2 \leq I_3$. The inertial tensor \mathbf{I}' for an EM structure is similarly described, where (x'_i, y'_i, z'_i) is the position of the center of each voxel, and we assume m'_i is the same for all voxels. The origin \mathbf{o}' is set on the center of mass, and the principal axes of inertia $(\mathbf{e}'_x, \mathbf{e}'_y, \mathbf{e}'_z)$ satisfies $\mathbf{I}'\mathbf{e}'_x = I'_1\mathbf{e}'_x$, $\mathbf{I}'\mathbf{e}'_y = I'_2\mathbf{e}'_y$, and $\mathbf{I}'\mathbf{e}'_z = I'_3\mathbf{e}'_z$, where $I'_1 \leq I'_2 \leq I'_3$.

The fitting calculation is performed by translational and rotational manipulation of the atomic model so that the origin \mathbf{o} agrees with \mathbf{o}' , and $(\mathbf{e}_x, \mathbf{e}_y, \mathbf{e}_z)$ with $(\mathbf{e}'_x, \mathbf{e}'_y, \mathbf{e}'_z)$. Because the directions of the principal axes cannot be defined uniquely (if \mathbf{e} is an eigenvector of inertia tensor \mathbf{I} , $-\mathbf{e}$ is also the eigenvector), there are four possible orientations for the atomic model, if we assume $(\mathbf{e}_x, \mathbf{e}_y, \mathbf{e}_z)$ and $(\mathbf{e}'_x, \mathbf{e}'_y, \mathbf{e}'_z)$ make right-handed systems. We calculate the fitting scores for the four orientations and choose the one with the highest score.

(2) Refinement of Rigid-Body Fitting by Iterative Rotation and Translation

In this calculation, the atomic position $\mathbf{r}_i (= (x_i, y_i, z_i))$ is moved iteratively so that the fitting score is maximized. The atomic position \mathbf{r}_i is described as

$$\mathbf{r}_i = \mathbf{r}_i^0 + \boldsymbol{\omega} \times \mathbf{r}_i^0 + \boldsymbol{\delta}, (5)$$

where \mathbf{r}_i^0 is the initial position, $\boldsymbol{\omega}$ a rotation vector, around which each atom is rotated, and $\boldsymbol{\delta}$ a translation vector. The vectors $\boldsymbol{\omega}$ and $\boldsymbol{\delta}$ are the variables and we perform the search for the optimum position and orientation of the atomic model by looking for the rotation and translation vectors which give the highest fitting score.

We start from the position \mathbf{r}_i^0 determined by the calculation using principal axes of inertia with the vectors $\boldsymbol{\omega}$ and $\boldsymbol{\delta}$ set to zero. At each repetition, the vectors $\boldsymbol{\omega}$ and $\boldsymbol{\delta}$ are changed to $\boldsymbol{\omega} + \Delta\boldsymbol{\omega}$ and $\boldsymbol{\delta} + \Delta\boldsymbol{\delta}$, respectively. We describe $\Delta\boldsymbol{\omega}$ and $\Delta\boldsymbol{\delta}$ as

$$\begin{aligned} \Delta\boldsymbol{\omega} &= a (l_1, l_2, l_3), \\ \Delta\boldsymbol{\delta} &= b (m_1, m_2, m_3), \end{aligned} (6)$$

where a and b are the step sizes, and $(l_1, l_2, l_3, m_1, m_2, m_3)$ are integers. In this study, the range of $-3 \leq l_q, m_q \leq 3$ ($q=1, 2, 3$) is searched in each repetition. The combination of the integ-

ers which gives the highest fitting score is selected and the vectors $\boldsymbol{\omega} + \Delta\boldsymbol{\omega}$ and $\boldsymbol{\delta} + \Delta\boldsymbol{\delta}$ are used as $\boldsymbol{\omega}$ and $\boldsymbol{\delta}$ in the next repetition. If all the selected integers are zero ($\Delta\boldsymbol{\omega} = \Delta\boldsymbol{\delta} = \mathbf{0}$), the step sizes a and b are made smaller. The iteration is terminated when a and b become small enough.

III. Results and Discussion

1. Atomic Model Library

We use three series of deformed structures \mathbf{r}_n^k ($k=5, 6, 7$) for building atomic models by Eq. (3). The reason for choosing the three normal modes ($k=5, 6, 7$) is described in our previous paper.²⁾ We use 101 conformations ($-50 \leq n \leq 50$) for $k=5$ and 7 , and 151 conformations ($-50 \leq n \leq 100$) for $k=6$. Thus, the atomic model library consists of 1.54 million ($=101 \times 101 \times 151$) atomic models with different conformations. For each EM structure, the best-fitting atomic model is selected from this library by the rigid-body fitting calculation.

2. Rigid-Body Fitting Calculation

In Table 1, the fitting scores obtained by various rigid-body fitting methods are shown for the 42 EM structures. In the second column, the scores S_θ for the X-ray crystal structure fitted into the EM structures by using the principal axes of inertia are shown. They are quite dependent on the EM structures and tend to be low for the EM structures with low resolution (data not shown).

Equation (5) shows that the atomic positions are described by the six rigid-body parameters. Thus, in the refinement process of the rigid-body fitting, the search for the optimum position and orientation of the atomic model with respect to the EM structure is made in the six-dimensional space. However, if we allow all the six parameters to vary in each repetition, the computation takes quite long as shown in the last column of Table 1, where the CPU time T_6 required for fitting the X-ray crystal structure into each EM structure is listed. This is because the sampling space is large, that is, the number of combinations of the integers in Eq. (6) is large.

For fast computation, we perform two types of repetitions alternately. In one repetition, we allow only translational parameters to vary and fix the rotational parameters, and *vice versa* in another repetition. In this way, we make the computation faster by performing the search in the three-dimensional space at each repetition. The CPU time required for the computation is listed as T_3 in the 6th column of Table 1, which is much shorter than T_6 .

The fitting scores for the X-ray crystal structure by the two different iterative calculations are shown in Table 1. The fitting score S_3 is obtained by the iterative calculation performed in the three-dimensional space at all repetitions, while S_6 in the six-dimensional space. The latter is shown as $S_6 - S_3$ in the fourth column. As expected, S_6 tends to be higher than S_3 and the average difference is $(1.09 \pm 0.76) \times 10^{-4}$. It may seem strange that S_6 is smaller than S_3 for two cases (EM-1065 and 1363), because larger area is searched by the calculation in the six-dimensional space than that in the three-dimensional space. However, as long as the searching

Table 1 Results of rigid-body fitting calculations for 42 EM structures.

EM-ID	S_0	S_3	S_6-S_3 ($\times 10^{-4}$)	σ (\AA)	T_3 (sec)	T_6 (sec)
1003	0.803	0.867	1.42	0.533	45	4377
1004	0.509	0.735	0.08	0.108	58	4877
1005	0.521	0.697	1.38	0.347	55	5878
1006	0.743	0.833	0.69	0.180	52	4691
1007	0.761	0.842	1.82	0.234	49	5016
1008	0.770	0.839	0.53	0.222	40	5060
1045	0.790	0.852	1.09	0.371	44	4350
1055	0.634	0.806	0.36	0.112	47	4752
1056	0.651	0.813	0.00	0.028	50	4648
1064	0.470	0.666	2.31	0.267	50	5202
1065	0.508	0.605	-0.28	0.085	52	5630
1068	0.485	0.802	2.75	0.477	61	7133
1070	0.766	0.829	1.38	0.374	40	4446
1071	0.594	0.753	1.09	0.191	48	4449
1072	0.762	0.800	0.73	0.423	44	3736
1077	0.777	0.825	1.05	0.192	47	4457
1110	0.510	0.725	2.19	0.653	56	5560
1122	0.622	0.823	1.22	0.328	46	4269
1128	0.649	0.762	0.08	0.208	46	4657
1143	0.647	0.853	1.05	0.344	47	4752
1184	0.793	0.838	0.65	0.279	44	4406
1185	0.744	0.815	1.94	0.599	47	4680
1248	0.734	0.819	0.32	0.136	45	4406
1250	0.640	0.757	0.36	0.057	41	4209
1251	0.705	0.750	1.54	0.619	38	3963
1261	0.458	0.801	0.81	0.064	71	6129
1262	0.690	0.850	1.09	0.053	49	4735
1263	0.765	0.875	0.57	0.172	47	4526
1302	0.717	0.795	2.63	0.379	51	4609
1310	0.821	0.843	1.98	0.437	45	4129
1311	0.753	0.800	2.31	0.487	40	4137
1312	0.736	0.827	1.46	0.139	43	4763
1315	0.657	0.822	0.45	0.108	55	4940
1323	0.558	0.716	0.81	0.297	56	4752
1324	0.505	0.767	1.58	0.411	56	5344
1362	0.675	0.810	1.09	0.081	54	4382
1363	0.635	0.770	-0.24	0.327	46	3981
1365	0.623	0.784	1.50	0.494	47	5334
1366	0.747	0.814	0.53	0.070	38	4255
1370	0.688	0.803	1.86	0.227	48	4114
1391	0.571	0.736	0.69	0.549	48	5320
1395	0.781	0.829	0.85	0.079	45	4276

area is limited (otherwise, the calculation takes forever), we can not avoid this happen.

In the fifth column, the average difference of the atomic position $\sigma (= \sqrt{\sum (\mathbf{r}_i^{(3)} - \mathbf{r}_i^{(6)})^2 / N}$) is shown, where $\mathbf{r}_i^{(3)}$ and

$\mathbf{r}_i^{(6)}$ are the i th atomic position of the fitted atomic model by the calculations in the three and six-dimensional space, respectively. It is generally small and shows that the two fitted models are in the close position and orientation.

To see whether the difference S_6-S_3 is negligible or not, slightly deformed atomic models \mathbf{r}_i^k ($k=3, 5, 6$), each of which is different from the X-ray crystal structure \mathbf{r}_0 by 0.1 \AA in RMSD, are fitted into each EM structure by iterative calculation in the six-dimensional space and the fitting scores $S_6(\mathbf{r}_i^k)$ are obtained. The average values of the magnitude of the difference $|S_6(\mathbf{r}_i^k) - S_6(\mathbf{r}_0)|$ for the 42 EM structures are $(1.71 \pm 1.00) \times 10^{-4}$, $(4.07 \pm 1.55) \times 10^{-4}$, and $(3.97 \pm 1.80) \times 10^{-4}$ for $k=3, 5$, and 6 , respectively. These values are generally larger than the difference between the scores obtained by the calculations in the six and three-dimensional spaces (fourth column in Table 1), suggesting that the iterative calculation performed in the three-dimensional space at each repetition is practically accurate enough for selecting the best-fitting atomic model from the atomic model library.

IV. Conclusion

We have developed a new approach to build the best-fitting atomic model from an EM structure and applied it for analyzing the EM structures of the 70S ribosome. In our new approach, the rigid-body fitting calculation is most intensively executed, which is performed using the six rigid-body parameters. However, it takes quite long if all six parameters are treated as variables at the same time. Thus, we alternatively perform two different calculations, in which only three parameters are treated as variables and the others are fixed. We have ascertained that the calculation is fast and accurate enough.

Acknowledgment

This work was supported by Strategic International Cooperation Program and Core Research for Evolutional Science and Technology (CREST), Japan Science and Technology Agency and by Grants-in-Aid for Scientific Research on Priority Areas (No. 17053029) from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

References

- 1) J. Frank, R. K. Agrawal, "A ratchet-like inter-subunit reorganization of the ribosome during translocation," *Nature*, **406**[6793], 318-322 (2000).
- 2) A. Matsumoto, H. Ishida, "Global conformational changes of ribosome observed by normal mode fitting for 3D Cryo-EM structures," *Structure*, **17**[12], 1605-1613 (2009).
- 3) L. Jenner, P. Romby, B. Rees, C. Schulze-Briese, M. Springer, C. Ehresmann, B. Ehresmann, D. Moras, G. Yusupova, M. Yusupov, "Translational operator of mRNA on the ribosome: how repressor proteins exclude ribosome binding," *Science*, **308**[5718], 120-123 (2005).
- 4) G. J. Kleywegt, T. A. Jones, "Biomacromolecular Speleology," *CCP4/ESF-EACBM Newsletter on Protein Crystallography* **29**,

- 26-28 (1993).
- 5) G. J. Kleywegt, T. A. Jones, "Detection, delineation, measurement and display of cavities in macromolecular structures," *Acta Crystallogr D Biol Crystallogr*, **50**[Pt 2], 178-185 (1994).
- 6) G. J. Kleywegt, J. Y. Zou, M. Kjeldgaard, T. A. Jones, "Around O," in *Crystallography of Biological Macromolecules*, edited by M.G. Rossmann, E. Arnold (Dordrecht: Kluwer Academic Publishers, The Netherlands, 2001), pp. 353-356, 366-367.
- 7) O. Miyashita, J. N. Onuchic, P. G. Wolynes, "Nonlinear elasticity, proteinquakes, and the energy landscapes of functional transitions in proteins," *Proc. Natl. Acad. Sci. U S A*, **100**[22], 12570-12575 (2003).
- 8) M. G. Rossmann, R. Bernal, S. V. Pletnev, "Combining electron microscopic with x-ray crystallographic structures," *J. Struct. Biol.*, **136**[3], 190-200 (2001).
-