Published online 6 September 2023 *Nucleic Acids Research, 2023, Vol. 51, No. 18* **9567–9575**
https://doi.org/10.1093/nar/gkad721

RIBFIND2: Identifying rigid bodies in protein and nucleic acid structures

Sony Malhotra[1,†], Thomas Mulvaney[2,3,4,†], Tristan Cragnolini[2,5], Haneesh Sidhu[5], Agnel P. Joseph[1], Joseph G. Beton[2,3] and Maya Topf[2,3,4,*]

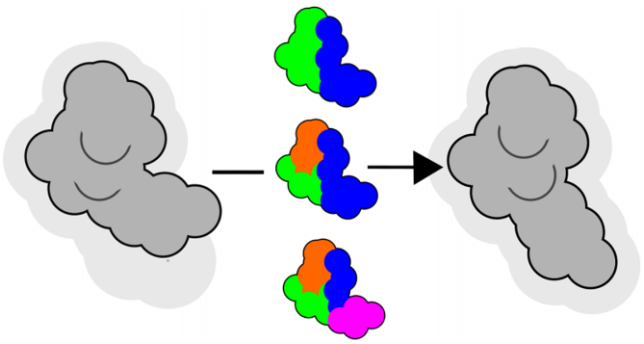
[1]Science and Technology Facilities Council, Scientific Computing, Research Complex at Harwell, Didcot OX11 0FA, UK, [2]Leibniz Institute of Virology, Hamburg 20251, Germany, [3]Centre for Structural Systems Biology, Hamburg D-22607, Germany, [4]Universitätsklinikum Hamburg Eppendorf (UKE), Hamburg 20246, Germany and [5]Institute of Structural and Molecular Biology, Department of Biological Sciences, Birkbeck College, University of London, London WC1E 7HX, UK

Received December 06, 2022; Revised August 10, 2023; Editorial Decision August 11, 2023; Accepted August 21, 2023

ABSTRACT

Molecular structures are often fitted into cryo-EM maps by flexible fitting. When this requires large conformational changes, identifying rigid bodies can help optimize the model-map fit. Tools for identifying rigid bodies in protein structures exist, however an equivalent for nucleic acid structures is lacking. With the increase in cryo-EM maps containing RNA and progress in RNA structure prediction, there is a need for such tools. We previously developed RIBFIND, a program for clustering protein secondary structures into rigid bodies. In RIBFIND2, this approach is extended to nucleic acid structures. RIBFIND2 can identify biologically relevant rigid bodies in important groups of complex RNA structures, capturing a wide range of dynamics, including large rigid-body movements. The usefulness of RIBFIND2-assigned rigid bodies in cryo-EM model refinement was demonstrated on three examples, with two conformations each: Group II Intron complexed IEP, Internal Ribosome Entry Site and the Processome, using cryo-EM maps at 2.7–5 Å resolution. A hierarchical refinement approach, performed on progressively smaller sets of RIBFIND2 rigid bodies, was clearly shown to have an advantage over classical all-atom refinement. RIBFIND2 is available via a web server with structure visualization and as a standalone tool.

GRAPHICAL ABSTRACT



INTRODUCTION

Cryo-electron microscopy (cryo-EM) is the method of choice for elucidating structures of large macromolecular assemblies at high (better than ∼4 Å) to medium resolutions (∼4–10 Å). Already ∼20% of cryo-EM structures in the Electron Microscopy Data Bank (EMDB) (1) contain RNA components. A large portion of the genome encodes for non-coding RNA (ncRNA) (2) and the Nucleic Acid Knowledge Base (NAKB) (3), the successor to the Nucleic Acid Database (NDB) (4,5), currently holds 16473 structures (as of August 2023). In the last year, 56% of new entries were derived from cryo-EM experiments. In total, 22% of all structures in the NAKB are from cryo-EM techniques at various resolutions, some of which prohibit clear determination of the atomic positions. These could be combined with RNA structure prediction and refinement algorithms, which are continuously improving (6,7). These changes in the field could lead to more insights into biological processes and experiments, such as CAS9-CRISPR gRNA generation and ribonucleoprotein assemblies.

*To whom correspondence should be addressed. Tel: +49 40 8998 87660; Email: maya.topf@cssb-hamburg.de
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

To derive an atomic model of an assembly, usually, atomic structures of assembly components are fitted into the cryo-EM map and then further refined within the map. Especially, but not exclusively, at medium resolutions, the latter process (also called 'flexible fitting') can be assisted and sped up by using rigid bodies (RBs) linked by flexible linkers in the fitted components. It can also improve the accuracy of the final refined model (8). At present, methods to identify RBs are mostly designed for protein structures (8).

The RIBFIND algorithm (9) was originally designed to detect RBs in protein structures via the clustering of secondary structural elements (SSEs), primarily to aid the fitting of structures into cryo-EM maps. RIBFIND was made available both as a web server and a standalone program (6). Here, we have developed a new algorithm, RIBFIND2, which identifies RBs in ncRNA structures by clustering SSEs assigned using the RNAView program (10). We have also optimized the original RIBFIND algorithm parameters for clustering protein structures. The algorithm was tested on structures containing proteins and RNA in different conformations.

We have implemented RIBFIND2 in a web server (Figure 1) with no login requirements—https://ribfind.topf-group.com/, which also supports a molecular JavaScript viewer—NGL viewer (11)—as it is more interactive, faster and scalable than our previous Java-based viewer (JMol, http://jmol.sourceforge.net/). We also provide the software as a standalone package which can be downloaded from a link provided in the web server.

MATERIALS AND METHODS

Secondary structure determination for RNA

The secondary structure of RNA molecules was determined with the RNAView (10) program, which is also used by the NDB to assign secondary structures to nucleic acids. RNAView calculates base-pairing interactions in a molecule based on distance and angle restraints. From those base pairs, it divides the molecule into double-stranded helical segments and single-stranded loop segments. Although the secondary structure classification of RNA is significantly more complex than these two categories, for the purposes of clustering this binary division contains enough information. Because single-stranded segments are an important part of RNA tertiary structure and are involved in intramolecular interactions, they were treated as secondary structure elements (SSEs) in their own right, rather than merely as connecting elements (as loops generally are in proteins). The RNAView secondary structure predictions (which are in XML format) were used for further calculations.

Clustering protein and RNA structures

The clustering algorithm is partially based on the original RIBFIND algorithm (neighborhood-based clustering) developed for defining RBs in proteins (8,9). The algorithm groups SSEs together into RBs based on the 'strength' of their interaction. For proteins, 'cutoff distance' (previously called 'contact distance') is defined as the distance between the average atomic position of side-chain atoms, except for glycine where the Cα is used. For RNA it is the average atomic position of nucleotide atoms excluding the phosphate groups. The strength of the interaction between an SSE (A) and a partner SSE (B) is defined in terms of the fraction of 'allowed' residues (see below) in A which are within the cutoff distance of the allowed residue in B. For proteins and RNA, the default cutoff distance is 6.5 Å (12) and 7.5 Å (13,14), respectively. These cutoff values can be changed to user-defined values. For RNA, this default was selected based on the analysis of base-base interactions in ellipsoidal shells (13).

The interaction strength is defined in terms of the fraction of residues within the cutoff distance of one another, where $|X|$ denotes the number of elements in the set X:

$$frac(A, B) = \frac{|cutoff(allowed(A), allowed(B))|}{|allowed(A)|} \quad (1)$$

Because $frac(A, B)$ does not necessarily equal $frac(B, A)$, the interaction for the pair is instead defined as the maximum of the two:

$$interaction(A, B) = max(frac(A, B), frac(B, A)) \quad (2)$$

The 'allowed' residues of an SSE enable finer control of interaction calculations. These are computed for each type of SSE. For β-sheets, only strands longer than three residues are allowed in interaction calculations (9). For unpaired RNA strands, a similar rule is applied. For α-helix to α-helix interactions, the ratio of the helix lengths in residues must be >0.4 (9).

Given the interaction function, a graph is constructed where nodes are SSEs and edges are the computed interactions. By choosing an *interaction threshold* (originally termed 'cluster cutoff') and removing edges from the graph that fall below this, the set of RBs (strongly connected components) changes. The algorithm, thus, produces unique sets of RBs and their respective interaction thresholds by iteratively removing edges in order of strength.

A 'unique' cluster number (UCN) for a given *interaction threshold* is defined as:

$$UCN = \frac{|SSEs \in RBs|}{|SSEs|} + |RBs| \quad (3)$$

where $|SSE \in Rigid Bodies|$ denotes the number of SSEs which are within RBs in the cluster of interest. We have previously demonstrated in detail the usefulness of the highest UCN in the refinement of three protein cases (9), where flexible fitting using clustered RBs resulted in a model that better fit the experimental map. The highest UCN has previously been chosen for refinement as it tends to have most of SSEs clustered into a large number of RBs. However, the highest UCN cluster may not always be the best for this purpose. We therefore compare it against a more costly 'hierarchical' approach in this paper.

Benchmark dataset for protein-nucleic acid complexes

The NDB (5) was searched for RNA structures with tertiary interactions to test the algorithm. A series of group IIC intron structures in different states of catalysis was first used to test the algorithm (PDB IDs: 3eog, 3eoh, 3bwp,

Figure 1. Snapshot of the RIBFIND2 server. (**A**) The input parameters required to submit the job to RIBFIND2 web server. (**B**) The results page with colored rigid bodies for both protein and RNA components in an input PDB ID (1mji). The user can turn on/off the protein, RNA or the cryo-EM map. The slider lets the user control the *interaction threshold* for both protein and RNA components.

Table 1. Dataset used to assess the performance of RIBFIND2 rigid bodies during TEMPy-REFF refinement

Type	Model	Map	Res.† (Å)	Description
Processome	7MQ9	23937	3.9	Cryo-EM structure of the human SSU processome, state pre-A1*
	7MQA	23938	2.7	Cryo-EM structure of the human SSU processome, state post-A1
Group II Intron complexed with intron-encoded protein (IEP)	7D0F	30532	5.0	Cryo-EM structure of a precatalytic group II intron RNP
	7D0G	30533	5.0	Cryo-EM structure of a precatalytic group II intron
IRES	7SYR	25538	3.6	Structure of the wt IRES eIF2-containing 48S initiation complex, closed conformation. Structure 12(wt)
	7SYQ	25537	3.8	Structure of the wt IRES and 40S ribosome ternary complex, open conformation. Structure 11(wt)

†Res. refers to the resolution of the cryo-EM map.

4ds6, 5j01, 5j02). Additionally, structures of the 80S ribosome in different states of Internal Ribosome Entry Site (IRES) translocation were then used, in which the small and large subunits were run separately and with protein chains removed (PDB IDs: 5juo, 5jus, 5jut, 5juu). The clustering of structures were viewed and analyzed using UCSF Chimera (15).

Application to cryo-EM refinement

We selected three cases of RNA structures, in two conformations each, to test the usefulness of RIBFIND2 in refining those structures in cryo-EM maps. These were: Group II Intron complexes with intron-encoded protein (IEP), Internal Ribosome Entry Site (IRES) and the Processome, with cryo-EM maps between 2.7 and 5 Å resolution.

We refined each atomic model into the cryo-EM map corresponding to the other conformation (Table 1). For the Group II intron models, both the RNA and protein chains (chains A and C respectively) were refined. Due to the large size of the processome and IRES models, we refined only two of the major RNA chains from these models, which corresponded to chains 'L1' and 'L2' and chains '2' and 'z', respectively. We compared two approaches of applying these restraints, the first based on the decomposition of RBs ('hierarchical') and the second based on choosing a single cluster with the highest UCN. In the hierarchical approach, RIBFIND2 clusters are selected in order of increasing interaction threshold, which leads to progressively smaller clusters and thus more and more flexibility. As a control, we performed an unrestrained refinement.

The refinement protocol included three steps (Supplementary Figure S1): (i) the model was first aligned to the target to produce a rough fit then locally optimized using the 'fitmap' tool in ChimeraX to produce the initial starting model; (ii) TEMPy-REFF (16) density-guided fitting was used in conjunction with progressively smaller RIBFIND2 RBs (hierarchical), the highest UCN set of RBs (UCN), or all-atom (unrestrained) and (iii) TEMPy-REFF all-atom Gaussian-mixture model refinement.

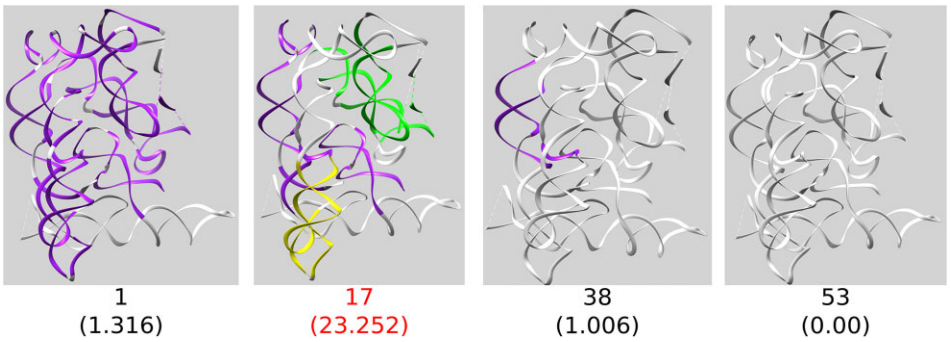


Figure 2. Clustering for group IIC intron (PDB ID: 3bwp) using RIBFIND2, labeled with *interaction threshold* and the UCN (below in parentheses). The *interaction threshold* for the highest UCN is highlighted in red.

Each set of RBs were refined using TEMPy-REFF until convergence, i.e. the variance in CCC score of the last five runs was $<10^{-9}$. The TEMPy-REFF density-guided force field was set to a strength of 20 in all experiments. The TEMPy-REFF GMM strength for the refinement step was 10^3.

RESULTS

Clustering RNA in group II introns

The algorithm was first tested on Group IIC introns. The SSE clustering using a 1% threshold to no clusters (at threshold ≥17%) is shown in snapshots for an intron structure in Figure 2 (PDB ID: 3bwp). The intron starts as one cluster (comprising all secondary structures) initially and then breaks off into smaller clusters, with the number of clusters peaking at the highest UCN. The algorithm was also run on several other group IIC intron structures in various states of catalysis; The *interaction threshold* and highest UCN for each of these states is shown in Figure 2. Structures with PDB IDs: 3bwp, 3eog, 3eoh and 4ds6 are truncated (lacking domain 6) and consequently linear introns, as the branch point adenosine resides in domain 6 (17) while structures 5j01 and 5j02 are branched chimeric introns and are derived from *Oceanobacillus iheyensis*. A similar clustering pattern is observed across the different catalytic states with three key clusters emerging highlighted in purple, yellow and green (and for those intron structures without broken chains additionally a cluster in blue). Exceptions to this are the chimeras (5j01 and 5j02), which were created by replacing part of the *O. iheyensis* sequence with intron AV.I.2 (17). In those cases, the yellow cluster does not break off and instead appears as part of larger purple clusters (Figure 3). 4ds6 also shows a red cluster at the top, which for all other states is non-clustered. This may reflect state-specific reduced flexibility in the pre-catalytic structure as well as for the chimeras.

Clustering RNA in the 80S eukaryotic ribosome

To test the algorithm on higher complexity RNA structures, a set of structures of the 80S ribosome bound to the Taura syndrome virus IRES were used (18). IRESs are RNA structures that carry out cap-independent translation of viral mRNA via interacting with the 40S subunit (18). The ensemble of structures illustrates translocation and rearrangements of the IRES, coupled with 40S intra and inter-subunit rearrangements and therefore represents a good example of biologically relevant rigid body RNA movements.

The small subunit has been well characterized in terms of its dynamics and domains in many ribosome structures. The canonical small subunit is composed of head, beak, body and platform domains (Figure 4), based on transitions between different states during translocation (19,20). Snapshots of the trajectory of clustering from 1% threshold to no clusters (at interaction threshold ≥61%) for a single 40S conformation (PDB ID: 5juo) are shown in Figure 4. As observed with intron structures, larger clusters are present at lower thresholds, which eventually separate into smaller domains, but still include most of the SSEs, which then gradually localize to subsections or peripheries excluding most SSEs as the *interaction threshold* is increased. At thresholds of 15–25% there is a good separation of head, beak, platform, body and IRES domains. Moreover, the IRES initially starts as one cluster that breaks into two clusters approximately corresponding to its two known domains (the 5′ region and PKI region) (18).

We further assessed the algorithm to generate functionally meaningful clusters using other conformations in addition to 5juo (PDB IDs: 5jut, 5juu, 5jup, 5jus). RIBFIND2 assignment with the highest UCN resulted in a similar clustering pattern into the classical domains, as well as of the IRES (which also adopts a different conformation in eac16h structure) (Supplementary Figure S2). As well as the canonical 40S domains, a set of 3–4 clusters (coloured orange, red, yellow and cyan in Figure 4) are consistently found in the lower half of the 40S subunit. The clustering of these RBs changes the least for the different states. Comparing the proportion of SSEs in clusters *vs.* the *interaction threshold* shows a drop around the threshold that corresponds to the highest UCN in all conformations (Supplementary Figure S3).

During the transition between the different conformations the head domain rotates by ∼40° (19,20), a phenomenon also reported for bacterial and mammalian systems (21–23) Thus, the similar clustering pattern observed along the trajectory of different states suggests the clusters identified by the algorithm represent biologically relevant RBs.

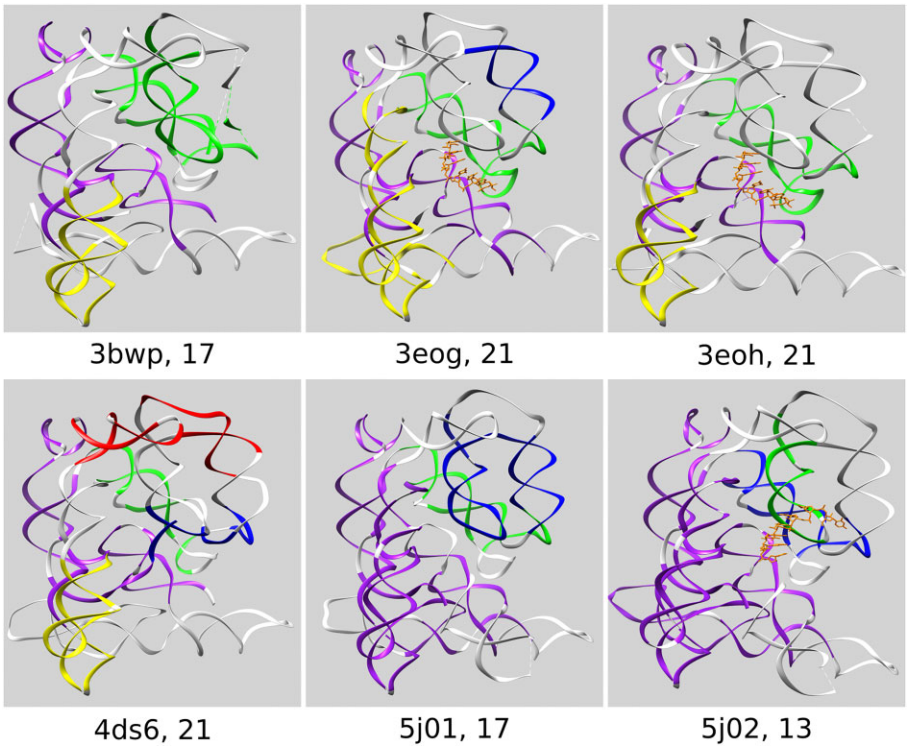


Figure 3. Clustering for group IIC introns in different catalytic states and conformations. For each structure, the clustering based on the highest UCN is indicated next to PDB ID.

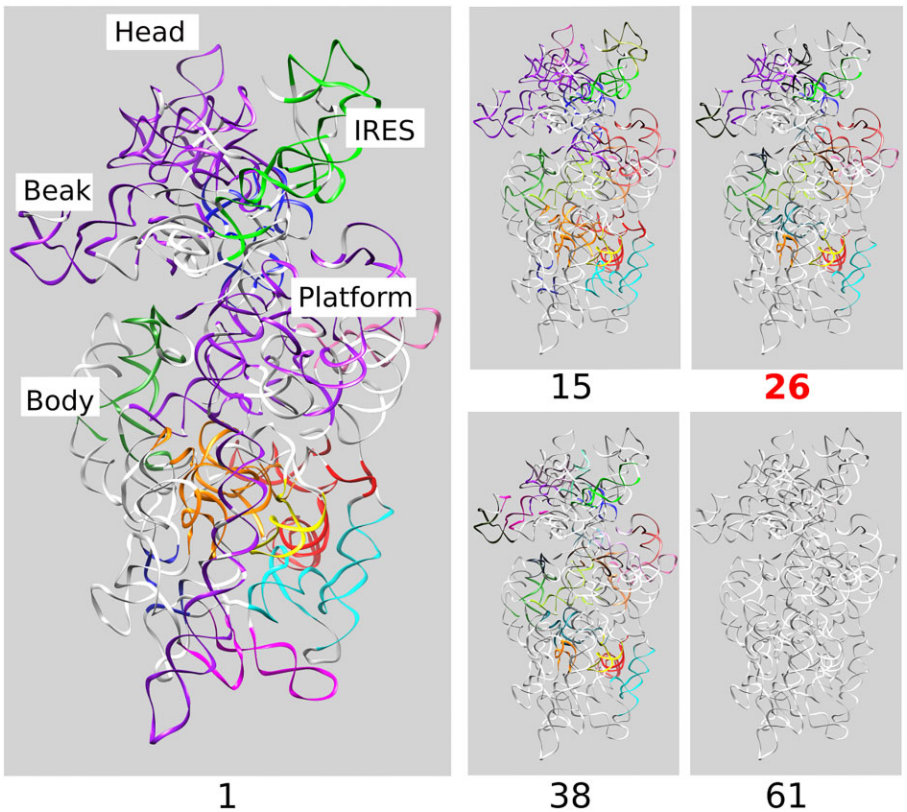


Figure 4. Clustering for small ribosomal subunit (PDB ID: 5juo). Canonical domains are marked in the first panel and clustering patterns are shown in ascending order labelled with corresponding interaction threshold. The interaction threshold of the clustering with the highest UCN is highlighted in red.

Table 2. Assessment of refined models using different refinement approaches. Best CCC and RMSD are highlighted in bold

Model	Target	Method	CCC	RMSD (Å)
7mq9	7mqa	Target	0.64	0.0
		Initial	0.49	15.1
		Hierarchical	**0.62**	**7.3**
		UCN (25.53)	0.6	8.5
		Unrestrained	0.57	12.6
7mqa	7mq9	Target	0.64	0.0
		Initial	0.47	14.9
		Hierarchical	**0.56**	**7.6**
		UCN (33.62)	0.53	11.24
		Unrestrained	0.54	12.3
7d0f	7d0g	Target	0.85	0.0
		Initial	0.69	6.2
		Hierarchical	**0.81**	3.2
		UCN (14.66)	**0.81**	3.2
		Unrestrained	**0.81**	**3.1**
7d0g	7d0f	Target	0.85	0.0
		Initial	0.69	5.9
		Hierarchical	**0.82**	3.1
		UCN (15.84)	**0.82**	3.1
		Unrestrained	0.81	**2.9**
7syr	7syq	Target	0.75	0.0
		Initial	0.70	6.9
		Hierarchical	**0.78**	2.8
		UCN (47.77)	**0.78**	**2.7**
		Unrestrained	**0.78**	2.9
7syq	7syr	Target	0.73	**0.0**
		Initial	0.68	6.9
		Hierarchical	**0.78**	**2.2**
		UCN (43.77)	0.77	2.4
		Unrestrained	0.76	3.3

Clustering in the large subunit

Compared to the small subunit the large subunit is less dynamic during translocation and more compact with the RNA not classically seen as separated into distinct domains (24,25). The clustering of a single large subunit (PDB ID: 5juo) from 1% threshold to no clusters (at interaction threshold ≥61%) is shown in snapshots in Supplementary Figure S3. The central core of the 60S subunit largely stays as a large cluster (orange, Supplementary Figure S2) with peripheral SSEs gradually breaking off into different clusters. Overall, the core of 60S is conserved between the different states, particularly at the lower end of the *interaction threshold* range shown. The surfaces break into small clusters for all states representing flexibility compared to the core globular domain.

Using clusters of RIBFIND2 for cryo-EM structure refinement

The examples from Table 1 were used to perform refinement in two ways: hierarchical and UCN-based (see Materials and Methods). The hierarchical approach combines the advantages of using both the highest UCN and unrestrained approaches: large cluster sizes, used at the start of refinement, enable large conformational changes during fitting and hence prevents the model from getting stuck in small pockets of density, whilst small cluster sizes facilitate the small adjustments required for accurate refinement once the model is placed in an approximately correct position.

In total, we performed refinements of six structures for the three cases (Materials and Methods and Table 1). We assessed the performance using a density-based metric, cross correlation (CCC), and a density-independent metric, root mean square deviation (RMSD). The latter was calculated over C4' RNA atoms (26,27) of the refined model from the target structure (Table 2). CCC scores for hierarchical refinements were generally higher or comparable to the UCN and unrestrained approaches. For the processome, we excluded residue ranges 1256–1516 and 1839–1860 from RMSD calculations. The former is in a low resolution part of the map, the latter is a small modelled fragment, which is disconnected from the rest of the model.

The combination of CCC and RMSD scores was best for the hierarchical approach, suggesting better fit was obtained whilst minimizing overfitting (lower RMSD values, Table 2). The geometry of the refined models were assessed using RNAValidate which is part of the PHENIX software (28) (Supplementary Table S1). There were no obvious differences between the restrained and unrestrained refinements. However, all refinements had a decrease in suite outliers and an increase in bond-angle outliers.

A comparison of the CCC score trajectories during refinement for the hierarchical, UCN and unrestrained approaches is presented in Figure 5A. A close-up of the IRES differences from the target model demonstrates the advantages of using RIBFIND2-defined RBs over the unrestrained refinement where no RBs are used (Figure 5B). Generally, both the hierarchical and UCN approaches enabled the flexible-fitting to converge on a conformation closer to the target structure (Figure 5, Tables 2 and S1). The local fit-to-map of the IRES model 7syq in map EMD-25538 was assessed using SMOC scores (Supplementary Figure S4A) which are part of the TEMPy (29). Compared to the hierarchical-based fitting (blue), UCN-based (orange) and unrestrained (green) refinements produced models with lower SMOC scores in the beak domain which is marked by a box. A close-up of the refined models in this region shows that the hierarchical model was closer to the target model (red) (Supplementary Figure S4B).

RIBFIND2 web server

To make the program user-friendly, RIBFIND2 has been implemented as a web server (https://ribfind.topf-group.com/). By default, the server accepts a single PDB file. However, the advanced form allows the previously described distance thresholds and interaction parameters to be adjusted from their defaults.

For proteins, the following parameters are user-definable:

1. The protein residue cutoff distance (default 6.5 Å).
2. The minimum ratio of lengths between helices for them to interact (default 0.4).

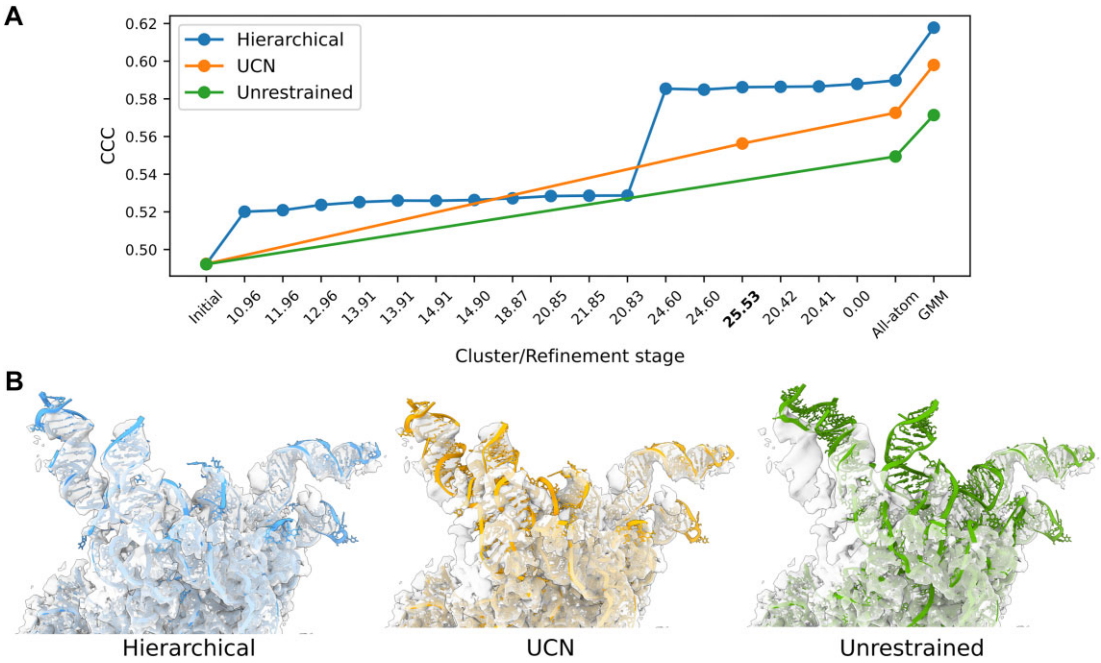


Figure 5. Refinement results for the processome model PDB ID: 7mq9 in cryo-EM map EMD-23938. (**A**) The CCC scores after each step for the hierarchical (blue), UCN (orange) and unrestrained (green) refinement procedures. Both restrained approaches yielded models with higher CCC scores. (**B**) Visualization of nucleotides 150–350 in the final model within the cryo-EM map (transparent grey). For the unrestrained model (green) and UCN model (orange) it is clear that some helices were unable to move to the correct density.

3. The minimum beta-strand length determines if a strand can be involved in interactions (default 4).
4. The cluster size, which determines the minimum number of clustered SSEs which can be called a RB.

For nucleic acids, the following parameters are user definable:

1. The nucleic acid cutoff distance (default 7.5 Å).
2. The minimum strand length of an RNA loop which can interact (default 4).
3. The minimum cluster size which is considered a RB (default 2).

The PDB file is read and validated internally before it is submitted for execution. After successful completion, the clusters are displayed on the web page with the results of the job. Failure to run the job caused by DSSP or RNAView processing their input are reported to the user so they may correct these issues.

Using the slider control on the result page, the user can view different sets of RB clusters generated for each *interaction threshold* and save the corresponding RB file in a text format (which can be used, e.g. by TEMPy-REFF or Flex-EM (30)). The run-time for some examples is listed in Supplementary Table S2.

NGL and visualization. It is useful to provide an efficient visualization of biomolecular structures, and their separation in structural elements or a group of structural elements. To this end, we have implemented a NGLview JavaScript molecular viewer, where each RB in the user uploaded PDB file is colored uniquely (8). All SSEs and loops that do not form part of any cluster are colored white. The clustering with the greatest UCN identified by the program is displayed by default. Directly embedded in the results page, it allows for quick and responsive visualization of very large structures (e.g. viral capsids). The viewer allows intuitive interaction using the mouse to quickly and easily change the display and consistent coloring of the structural blocks identified by RIBFIND2.

DISCUSSION

Rigid-body identification in biomolecular structures is a highly useful step to analyse structural models and to refine them against experimental data. Yet, few methods exist to do so in an automated fashion for nucleic acids. We have shown here that RIBFIND2 can be used to provide RBs that correspond to biologically relevant units in important RNA and protein/RNA structures, using group II introns and ribosome subunits as examples. We have also demonstrated that combining RBs identified by RIBFIND2 with a cryo-EM refinement method enhances the final quality of the model. This is particularly relevant for cryo-EM RNA structures, which are on average characterized by lower resolution compared to protein structures. Further, the optimal number of RBs is also dependent on the resolution of the map. We have previously shown using a simulated benchmark that the improvement in CCC drops as the resolution drops and at resolutions worse than 10 Å, it is hard to obtain an accurate refined model. At lower resolution, multiple structures tend to have similar fitting scores and hence it is more difficult to refine them. Previously, clus-

tering protein secondary structure elements into larger RBs was shown to improve the flexible fitting process (9).

Here we have demonstrated that RBs can also be used for the flexible fitting of RNA structures. To achieve this, we used a hierarchical approach (fitting each of the RIBFIND2 sets of RBs in order of 0–100%) which produced models which were closer to the target structure (lower RMSD) and were a better fit-to-map (higher CCC) compared to the model resulting from a standard unrestrained (all-atom) refinement or a refinement based on the highest UCN. Future work could include integrating the current method with deep-learning-based structure prediction methods due to its successful combination with cryo-EM model refinement. This could be done in an interactive manner, for example with molecular visualization and molecular dynamics tools, and could also aid in providing better model assessment and functional interpretation.

DATA AVAILABILITY

All data used in this study can be obtained from the PDB and EMDB respositories. The RIBFIND2 webserver is accessible at http://ribfind.topf-group.com. The refined models and associated RIBFIND clusters are deposited in the Zenodo repository, at https://dx.doi.org/10.5281/zenodo.8221020.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Jae Anne Bach Hardie for her contributions. We also thank Guendalina Marini and Aaron Sweeney for their helpful feedback on the manuscript.

FUNDING

Wellcome Trust grant [209250/Z/17/Z]; cooperation of Leibniz Institute of Virology (LIV) and Universitätsklinikum Hamburg Eppendorf (UKE) (as part of Leibniz ScienceCampus InterACt, funded by the BWFGB Hamburg and the Leibniz Association) and the Landesforschungsförderung Hamburg (HamburgX). Funding for open access charge: Leibniz-Institute for Virology.
Conflict of interest statement. None declared.

REFERENCES

1. Patwardhan,A. (2017) Trends in the Electron Microscopy Data Bank (EMDB). *Acta Crystallogr D Struct Biol*, **73**, 503–508.
2. ENCODE Project Consortium, Snyder,M.P., Gingeras,T.R., Moore,J.E., Weng,Z., Gerstein,M.B., Ren,B., Hardison,R.C., Stamatoyannopoulos,J.A., Graveley,B.R. *et al.* (2020) Perspectives on ENCODE. *Nature*, **583**, 693–698.
3. Berman,H.M., Lawson,C.L. and Schneider,B. (2022) Developing community resources for nucleic acid structures. *Life*, **12**, 540.
4. Coimbatore Narayanan,B., Westbrook,J., Ghosh,S., Petrov,A.I., Sweeney,B., Zirbel,C.L., Leontis,N.B. and Berman,H.M. (2014) The Nucleic Acid Database: new features and capabilities. *Nucleic Acids Res.*, **42**, D114–D22.
5. Berman,H.M., Olson,W.K., Beveridge,D.L., Westbrook,J., Gelbin,A., Demeny,T., Hsieh,S.H., Srinivasan,A.R. and Schneider,B. (1992) The nucleic acid database. A comprehensive relational database of three-dimensional structures of nucleic acids. *Biophys. J.*, **63**, 751–759.
6. Miao,Z., Adamiak,R.W., Antczak,M., Boniecki,M.J., Bujnicki,J., Chen,S.-J., Cheng,C.Y., Cheng,Y., Chou,F.-C., Das,R. *et al.* (2020) RNA-Puzzles Round IV: 3D structure predictions of four ribozymes and two aptamers. *RNA*, **26**, 982–995.
7. Kretsch,R.C., Andersen,E.S., Bujnicki,J.M., Chiu,W., Das,R., Luo,B., Masquida,B., McRae,E.K.S., Schroeder,G.M., Su,Z. *et al.* (2023) RNA target highlights in CASP15: Evaluation of predicted models by structure providers. *Proteins*, https://doi.org/10.1002/prot.26550.
8. Pandurangan,A.P. and Topf,M. (2012) RIBFIND: a web server for identifying rigid bodies in protein structures and to aid flexible fitting into cryo EM maps. *Bioinformatics*, **28**, 2391–2393.
9. Pandurangan,A.P. and Topf,M. (2012) Finding rigid bodies in protein structures: application to flexible fitting into cryoEM maps. *J. Struct. Biol.*, **177**, 520–531.
10. Yang,H., Jossinet,F., Leontis,N., Chen,L., Westbrook,J., Berman,H. and Westhof,E. (2003) Tools for the automatic identification and classification of RNA base pairs. *Nucleic Acids Res.*, **31**, 3450–3460.
11. Rose,A.S. and Hildebrand,P.W. (2015) NGL Viewer: a web application for molecular visualization. *Nucleic Acids Res.*, **43**, W576–W579.
12. Miyazawa,S. and Jernigan,R.L. (1996) Residue – residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. *J. Mol. Biol.*, **256**, 623–644.
13. Bottaro,S., Di Palma,F. and Bussi,G. (2014) The role of nucleobase interactions in RNA structure and dynamics. *Nucleic Acids Res.*, **42**, 13306–13314.
14. Bernauer,J., Huang,X., Sim,A.Y.L. and Levitt,M. (2011) Fully differentiable coarse-grained and all-atom knowledge-based potentials for RNA structure evaluation. *RNA*, **17**, 1066–1075.
15. Pettersen,E.F., Goddard,T.D., Huang,C.C., Couch,G.S., Greenblatt,D.M., Meng,E.C. and Ferrin,T.E. (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.*, **25**, 1605–1612.
16. Cragnolini,T., Kryshtafovych,A. and Topf,M. (2021) Cryo-EM targets in CASP14. *Proteins*, **89**1949–1958.
17. Costa,M., Walbott,H., Monachello,D., Westhof,E. and Michel,F. (2016) Crystal structures of a group II intron lariat primed for reverse splicing. *Science*, **354**, aaf9258.
18. Abeyrathne,P.D., Koh,C.S., Grant,T., Grigorieff,N. and Korostelev,A.A. (2016) Ensemble cryo-EM uncovers inchworm-like translocation of a viral IRES through the ribosome. *Elife*, **5**, e14874.
19. Rodnina,M.V., Fischer,N., Maracci,C. and Stark,H. (2017) Ribosome dynamics during decoding. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, **372**, 20160182.
20. Frank,J. (2017) The translation elongation cycle—capturing multiple states by cryo-electron microscopy. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, **372**, 20160180.
21. Fischer,N., Konevega,A.L., Wintermeyer,W., Rodnina,M.V. and Stark,H. (2010) Ribosome dynamics and tRNA movement by time-resolved electron cryomicroscopy. *Nature*, **466**, 329–333.
22. Agirrezabala,X., Lei,J., Brunelle,J.L., Ortiz-Meoz,R.F., Green,R. and Frank,J. (2008) Visualization of the hybrid state of tRNA binding promoted by spontaneous ratcheting of the ribosome. *Mol. Cell*, **32**, 190–197.
23. Budkevich,T., Giesebrecht,J., Altman,R.B., Munro,J.B., Mielke,T., Nierhaus,K.H., Blanchard,S.C. and Spahn,C.M.T. (2011) Structure and dynamics of the mammalian ribosomal pretranslocation complex. *Mol. Cell*, **44**, 214–224.
24. Agirrezabala,X., Liao,H.Y., Schreiner,E., Fu,J., Ortiz-Meoz,R.F., Schulten,K., Green,R. and Frank,J. (2012) Structural characterization of mRNA-tRNA translocation intermediates. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 6094–6099.
25. Ban,N., Nissen,P., Hansen,J., Moore,P.B. and Steitz,T.A. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science*, **289**, 905–920.

26. Das,R. and Baker,D. (2007) Automated *de novo* prediction of native-like RNA tertiary structures. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 14664–14669.
27. Yan,Y. and Huang,S.-Y. (2018) RRDB: a comprehensive and non-redundant benchmark for RNA-RNA docking and scoring. *Bioinformatics*, **34**, 453–458.
28. Adams,P.D., Afonine,P.V., Bunkóczi,G., Chen,V.B., Davis,I.W., Echols,N., Headd,J.J., Hung,L.-W., Kapral,G.J., Grosse-Kunstleve,R.W. *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 213–221.
29. Cragnolini,T., Sahota,H., Joseph,A.P., Sweeney,A., Malhotra,S., Vasishtan,D. and Topf,M. (2021) TEMPy2: a Python library with improved 3D electron microscopy density-fitting and validation workflows. *Acta Crystallogr. D Struct. Biol.*, **77**, 41–47.
30. Topf,M., Lasker,K., Webb,B., Wolfson,H., Chiu,W. and Sali,A. (2008) Protein structure fitting and refinement guided by cryo-EM density. *Structure*, **16**, 295–307.