Biophysics and Physicobiology

Review Article (Invited)

Quantitative analysis of protein dynamics using a deep learning technique combined with experimental cryo-EM density data and MD simulations

Shigeyuki Matsumoto[1], Shoichi Ishida[2], Kei Terayama[2,3], Yasuhshi Okuno[1,3]

[1] *Graduate School of Medicine, Kyoto University, Kyoto 606-8507, Japan*

[2] *Graduate School of Medical Life Science, Yokohama City University, Yokohama, Kanagawa 230-0045, Japan*

[3] *RIKEN Center for Computational Science, Kobe, Hyogo 650-0047, Japan*

Protein functions associated with biological activity are precisely regulated by both tertiary structure and dynamic behavior. Thus, elucidating the high-resolution structures and quantitative information on in-solution dynamics is essential for understanding the molecular mechanisms. The main experimental approaches for determining tertiary structures include nuclear magnetic resonance (NMR), X-ray crystallography, and cryogenic electron microscopy (cryo-EM). Among these procedures, recent remarkable advances in the hardware and analytical techniques of cryo-EM have increasingly determined novel atomic structures of macromolecules, especially those with large molecular weights and complex assemblies. In addition to these experimental approaches, deep learning techniques, such as AlphaFold 2, accurately predict structures from amino acid sequences, accelerating structural biology research. Meanwhile, the quantitative analyses of the protein dynamics are conducted using experimental approaches, such as NMR and hydrogen-deuterium mass spectrometry, and computational approaches, such as molecular dynamics (MD) simulations. Although these procedures can quantitatively explore dynamic behavior at high resolution, the fundamental difficulties, such as signal crowding and high computational cost, greatly hinder their application to large and complex biological macromolecules. In recent years, machine learning techniques, especially deep learning techniques, have been actively applied to structural data to identify features that are difficult for humans to recognize from big data. Here, we review our approach to accurately estimate dynamic properties associated with local fluctuations from three-dimensional cryo-EM density data using a deep learning technique combined with MD simulations.

◄ *Significance* ►

The experimentally derived structural data of the macromolecules reflect the conformational states found in the samples. Three-dimensional cryo-EM density data implicitly contain dynamics information of the target molecule as it is reconstructed from numerous particle images representing variable conformations attributed to the in-solution dynamics properties. This indicates the potential of cryo-EM data for quantitative investigation of dynamics. A deep learning technique, three-dimensional convolutional neural network, combined with molecular dynamics simulations formulates the relationship between the density data and the quantitative dynamics information, allowing the extraction of the dynamics properties only from the cryo-EM density maps.

Corresponding authors: Yasushi Okuno, Graduate School of Medicine, Kyoto University, 53 Kawahara-cho,shogoin Sakyo-ku, Kyoto 606-8507, Japan. ORCID iD: https://orcid.org/0000-0003-3596-4208, e-mail: okuno.yasushi.4c@kyoto-u.ac.jp; Shigeyuki Matsumoto, Graduate School of Medicine, Kyoto University, 53 Kawahara-cho,shogoin Sakyo-ku, Kyoto 606-8507, Japan. ORCID iD: https://orcid.org/0000-0001-9329-6362, e-mail: matsumoto.shigeyuki.4z@kyoto-u.ac.jp

Introduction

Protein function is precisely regulated by its three-dimensional (3D) structure and dynamic properties. Thus, it is important to elucidate both features at the atomic level to understand the molecular mechanisms of protein functions. High-resolution 3D structures are experimentally determined by X-ray crystallography, cryogenic electron microscopy (cryo-EM), and nuclear magnetic resonance (NMR). Meanwhile, quantitative analysis of in-solution dynamic behavior requires other experimental or computational approaches, such as NMR, hydrogen-deuterium exchange mass spectrometry (HDX-MS) [1], and molecular dynamics (MD) simulations [2]. Their complementary use uncovers detailed structural features associated with protein function (Fig. 1). Among the approaches for experimentally investigating structural properties and with recent developments in hardware and analysis techniques, cryo-EM has increasingly uncovered novel 3D structures at atomic or near-atomic resolution, particularly those of large and complex macromolecules, accelerating the development of structural biology [3–6]. Nevertheless, quantitative analysis of the dynamics behavior of such large and complex molecules through conventional experimental and computational approaches is substantially difficult owing to fundamental limitations, such as significant signal crowding and extremely high computational cost. Therefore, the quantitative analysis of the dynamics behavior of large and complex macromolecules targeted by cryo-EM is an important issue in current structural biology research.
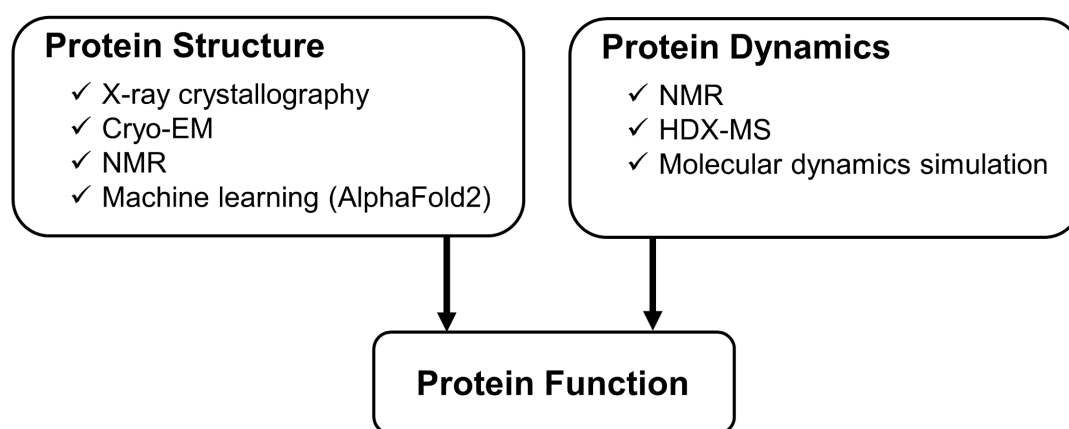

Figure 1 Procedures for investigating protein functions. Protein function is investigated from both aspects of the tertiary structure and the dynamics.

The structural data obtained by the experimental and computational approaches are collected and maintained in databases (Table 1). Protein Data Bank (PDB) contains structural models of biological macromolecules [7]. Electron Microscopy Public Image Archive (EMPIAR) [8] and Electron Microscopy Data Bank (EMDB) [9] contain cryo-EM image data and 3D density maps reconstructed from them, respectively. Biological Magnetic Resonance Bank (BMRB) is a database of NMR data for biological macromolecules [10]. Biological Structure Model Archive (BSM-Arc) is a recently developed database that publishes structural information obtained by computational methods, such as MD simulations and homology modeling methods [11]. These databases are publicly available and facilitate research in structural biology.

Table 1 Databases of structural data associated with biological macromolecules

Database	Contents	URL	Ref
PDB	Experimentally determined 3D structures and computed models.	https://www.rcsb.org/	[7]
EMDB	Cryo-EM density maps and tomograms.	https://www.ebi.ac.uk/emdb/	[8]
EMPIAR	Raw images in cryo-EM investigations.	https://www.ebi.ac.uk/empiar/	[9]
BMRB	NMR data in investigations of biological macromolecules and metabolites.	https://bmrb.io/	[10]
BSM-Arc	Structural data obtained by computational works.	https://bsma.pdbj.org/	[11]
AlphaFold DB	Atomic models predicted by AI system	https://alphafold.ebi.ac.uk/	[12,13]

Structural studies using machine learning (ML) techniques, particularly deep learning (DL) techniques, have been made possible with the availability of the accumulated structural data, and many structural studies combined with these techniques have been reported in recent years. Among these studies, AlphaFold2 [12,13] is one of the most impressive. By applying DL techniques to the structural data in PDB and the primary sequence information, AlphaFold2 has achieved highly accurate predictions of 3D protein structures. DL techniques have also been intensively applied to cryo-EM data, which have increasingly accumulated due to recent technical breakthroughs. cryoDRGN [14] with an image encoder–volume–decoder architecture reconstructs heterogeneous cryo-EM maps from single-particle images. Topaz-Denoise [15] is a noise reduction method that uses an ML framework, Noise2Noise [16], and has been shown to improve the SNR of raw images by approximately 100 times. Emap2sec [17] successfully estimated secondary structure information from intermediate-resolution 3D cryo-EM maps using a 3D-convolutional neural network (CNN) [18–20], which shows high performance in object detection and classification in 3D images [21–23]. These studies and other DL techniques [24] indicate that ML approaches for finding features hidden in big structural data have become a powerful tool in structural biology research.

Here, we introduce our recently developed DL-based approach, Dynamics Extraction From cryo-EM Map (DEFMap), for predicting dynamic information only from cryo-EM 3D density maps [25]. This paper is an extended version of a Japanese review [26].

Dynamic Properties Hidden in Cryo-EM

In cryo-EM single-particle analysis (SPA) [3,27,28], a 3D density map is reconstructed using a large number of single-particle images of biological macromolecules found in the micrograph (Fig. 2). Because the specimen is prepared by rapidly freezing protein solutions, single-particle images represent various conformational states found in the solution. Therefore, the reconstructed 3D density maps are reflected by the dynamics behavior, in other words, the dynamics information is hidden in the density maps; that is, while the density map intensities of the rigid regions (e.g., protein interior forming the hydrophobic core) are strong, the intensities of flexible regions (e.g., loop regions exposed on the molecular surface) tend to be weak because the various conformational states are averaged (Fig. 2). It is generally recognized that such a relationship exists between map intensities and protein dynamics. However, it is difficult to quantitatively estimate the dynamics properties only from the map intensities as these intensities are affected by several factors other than dynamics, such as local denaturation during sample preparation and preferred particle orientation. In fact, the correlation between the dynamics properties determined by MD simulations and the raw map intensities at the corresponding regions was relatively poor (Fig. 3A, left panels). Thus, other methods, such as HDX-MS and MD simulations, have been additionally applied for the quantitative analysis of protein dynamics.
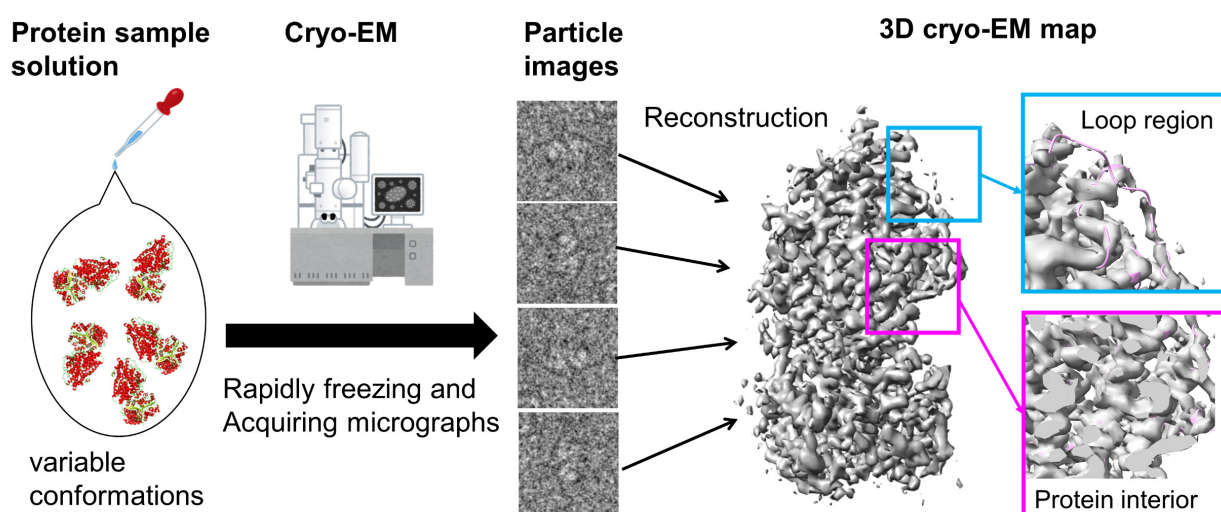

Figure 2 Overall workflow to reconstruct 3D cryo-EM maps. In the single particle analysis, 3D cryo-EM maps are reconstructed from a vast number of particle images representing biological macromolecules. The specimens are prepared by rabidly freezing the protein sample solution.
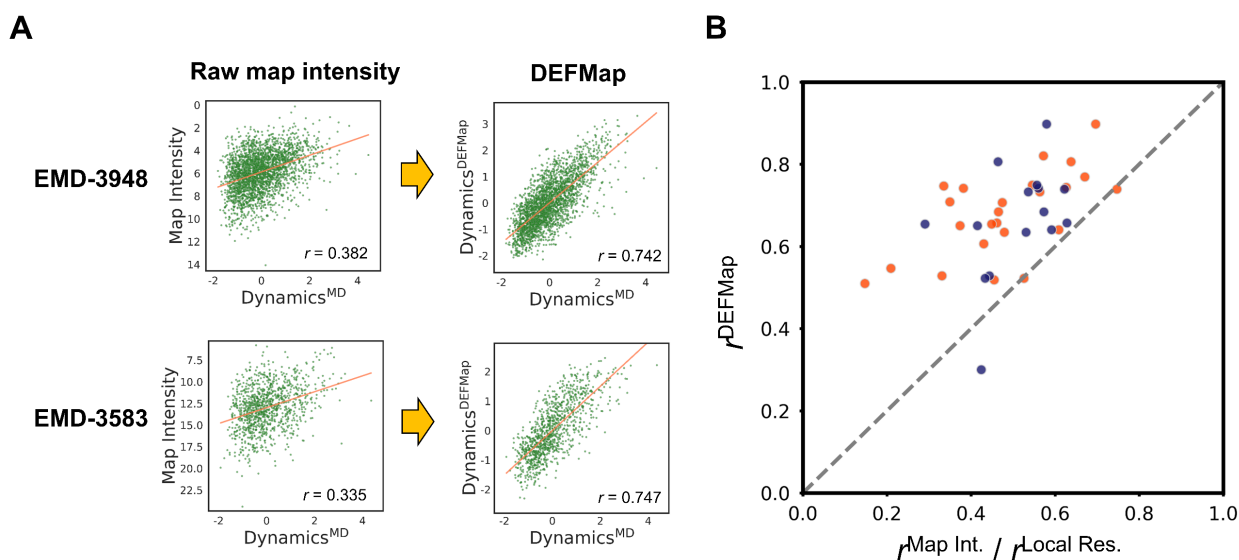
A

B

Figure 3 Correlation of MD-derived dynamics (Dynamics^{MD}) with raw map intensities, the derived local resolution estimates, and the values predicted by DEFMap (Dynamics^{DEFMap}). (A) Improvement of the correlations of raw map intensities with Dynamics^{MD} using DEFMap. Correlation plots for raw map intensities (left panels) and Dynamics^{DEFMap} (right panels) with Dynamics^{MD} are shown along with their corresponding regression lines (orange). Each point represents the residue-specific values which are calculated by averaging the values over each residue. r denotes the correlation coefficients. (B) Comprehensive comparison of the correlation coefficients for Dynamics^{DEFMap} with those for raw map intensities and the derived local resolution estimates. The correlation coefficients are calculated against Dynamics^{MD}. Each point represents the individual cryo-EM maps used in the evaluations and the relationships with raw map intensities and the local resolution estimates are colored by orange and navy, respectively. Regarding local resolution estimates, 10 out of 25 datasets were excluded from the plots because they exhibited inverse correlations. The $y = x$ line is represented by a black dashed line.

Developing a 3D CNN Model to Extract Dynamics Properties from Cryo-EM Maps

DEFMap formulates the relationship between the density data of the cryo-EM map and the corresponding protein dynamics properties with a supervised learning framework utilizing a 3D CNN. To achieve this, many datasets are required for the model training. The 3D cryo-EM maps used as explanatory variables are available in EMDB. However, the database containing the corresponding dynamics information used as objective variables is not available. In this study, dynamics data were generated using MD simulations with the program GROMACS [29]. The initial structures for the simulations were prepared utilizing the experimentally determined atomic coordinates deposited in PDB [7]. The disordered regions containing less than 7 residues were modeled and other non-natural termini were capped with acetyl or formyl groups. As target proteins used for the training, we selected 25 proteins based on the following criteria: (1) proteins with relatively small molecular weights and soluble nature for convenience of the MD simulations; and (2) proteins whose 3D density maps were determined with a resolution better than 4.5 Å.

Based on the idea of predicting the local dynamics properties in DEFMap, the local density data centered on the position of the existing heavy atoms in the corresponding atomic model were extracted from the overall 3D maps as subvoxels with grid lengths of 15 Å. As a preprocessing to efficiently train the model, a 5 Å low-pass filter and unification of the grid width (1.5 Å/grid) were applied to the downloaded cryo-EM maps (Fig. 4). After performing data augmentation by rotating the subvoxels by 90 ° in the xy, xz, and yz planes, 4,249,300 input datasets were prepared. The logarithm of the root-mean square fluctuation (RMSF), which represents the atomic fluctuations from the averaged positions in the MD trajectories with a length of 30 nsec, was used as the dynamics information (Fig. 4). The 3D CNN model trained using the prepared datasets quantitatively predicts the local dynamics properties in a regression manner only from the 3D cryo-EM map.

The performance of the constructed model was evaluated by a leave-one-out cross-validation method, in which one of the 25 proteins was used as a test dataset and the remaining 24 were used as training datasets. The correlations between the predicted values and the dynamics were evidently improved compared with those calculated from the raw map

intensities (Fig. 3A). The mean (± variance) of the correlation coefficient r obtained in the 25-fold cross validation was 0.665 (± 0.124), whereas that calculated from the raw map intensities and the local resolution estimates, which are conventionally used as indices of the dynamics in cryo-EM analyses [30,31], was 0.459 (± 0.179) and 0.510 (± 0.091), respectively (Fig. 3B). This indicates that DEFMap successfully extracted the patterns associated with the dynamics properties from the cryo-EM density data. While the present DL-based method should provide similar information as local resolution estimates, DEFMap was found to capture the dynamics-associated features better than local resolution estimates on the current datasets. This advantage may be attributed to the supervised learning framework, which enables the model to learn a large amount of density data derived from multiple proteins, i.e., big data.
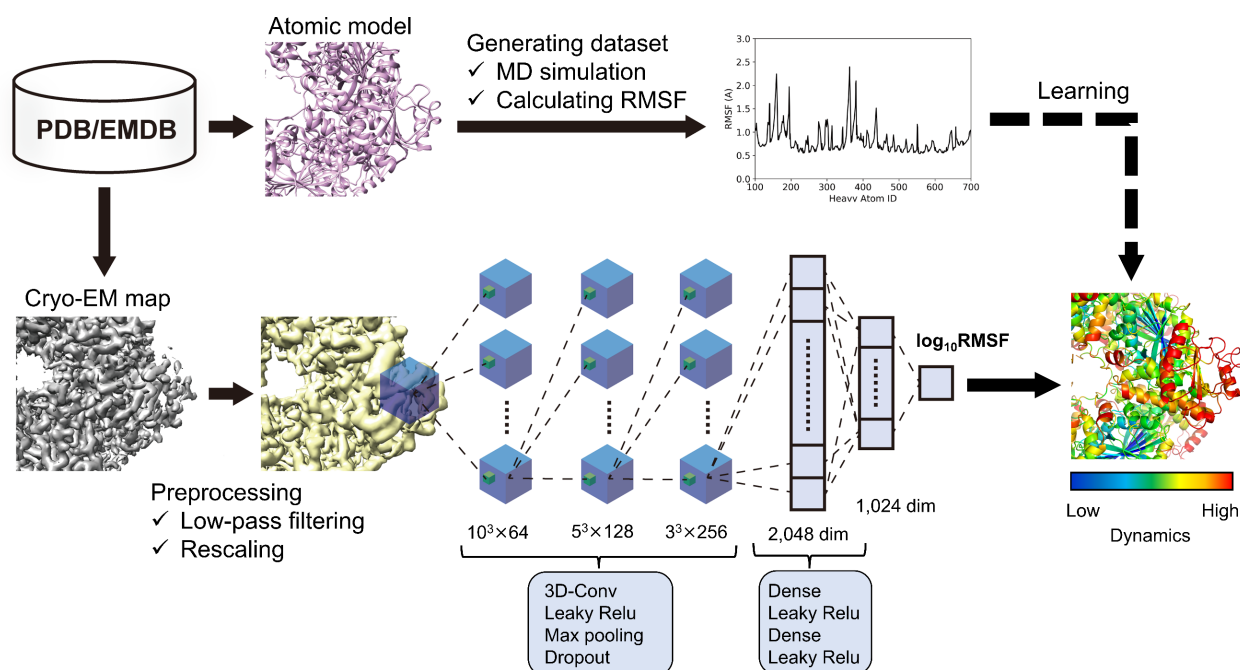

Figure 4 Schematic diagram of dataset preparation and learning in DEFMap. The dynamics information, the logarithm of RMSF, are generated by MD simulations, of which initial structures were modeled from the atomic coordinates downloaded from PDB (the upper workflow). The preprocessed cryo-EM density data obtained from EMDB are used as inputs of the neural network composed of three 3D convolutional layers with Leaky ReLU activation, max pooling and dropout and two dense layers (the lower workflow). Different filter sizes (64, 128, and 256) are applied to the three 3D convolutional layers.

Performance of DEFMap Against External Datasets

Because DEFMap learns a large scale of the local features found in cryo-EM maps, the constructed model is expected to show high generalization performance for external data not used in the training. To confirm the performance against the external datasets, we predicted the dynamics for three newly selected cryo-EM data (EMD-4241/6FE8 [32], EMD-7113/6BLY [33], and EMD-20308/6PCV [34]). The predicted values for all cases agreed well with the MD-derived dynamics values with correlation coefficients r of 0.727, 0.748, and 0.711, respectively, indicating that DEFMap can make accurate predictions in external datasets (Fig. 5A, left panels). Mapping the predicted values onto the 3D structures showed that DEFMap could successfully capture the general structural features, such as the rigidity of the protein interiors and the flexibilities of the regions exposed to the bulk solvent, as well as the MD simulations (Fig. 5A, right panels). It should be noted that the prediction performances gradually declined as the overall resolution of the density maps worsened, and DEFMap was applicable to maps with a resolution of up to approximately 6–7 Å (Fig. 5B). This can be explained by the loss of detailed local structural information as the resolution decreases. Because the local resolutions of cryo-EM maps are known to vary widely across the molecule, the prediction results for regions with extremely low local resolution should be carefully interpreted.

It is important to verify the predicted values with the experimentally determined dynamics properties because DEFMap learns computationally calculated dynamics information. It is therefore favorable, the dynamic data determined by HDX-MS are publicly available for one of the external datasets used in the evaluation: EMD-20308/6PCV [34]. We compared

the predicted and MD-derived dynamics values with the experimentally determined values and found that both correlated well with the experimental data with correlation coefficients r of 0.743 and 0.791, respectively (Fig. 5C). These evaluation results using external datasets emphasize that predictions using DEFMap can provide insights equivalent to those obtained by experimental approaches from the 3D cryo-EM maps that the user determined themselves. Thus, we further explored the impact of DEFMap in structural biology research, and the findings are introduced in the following section.
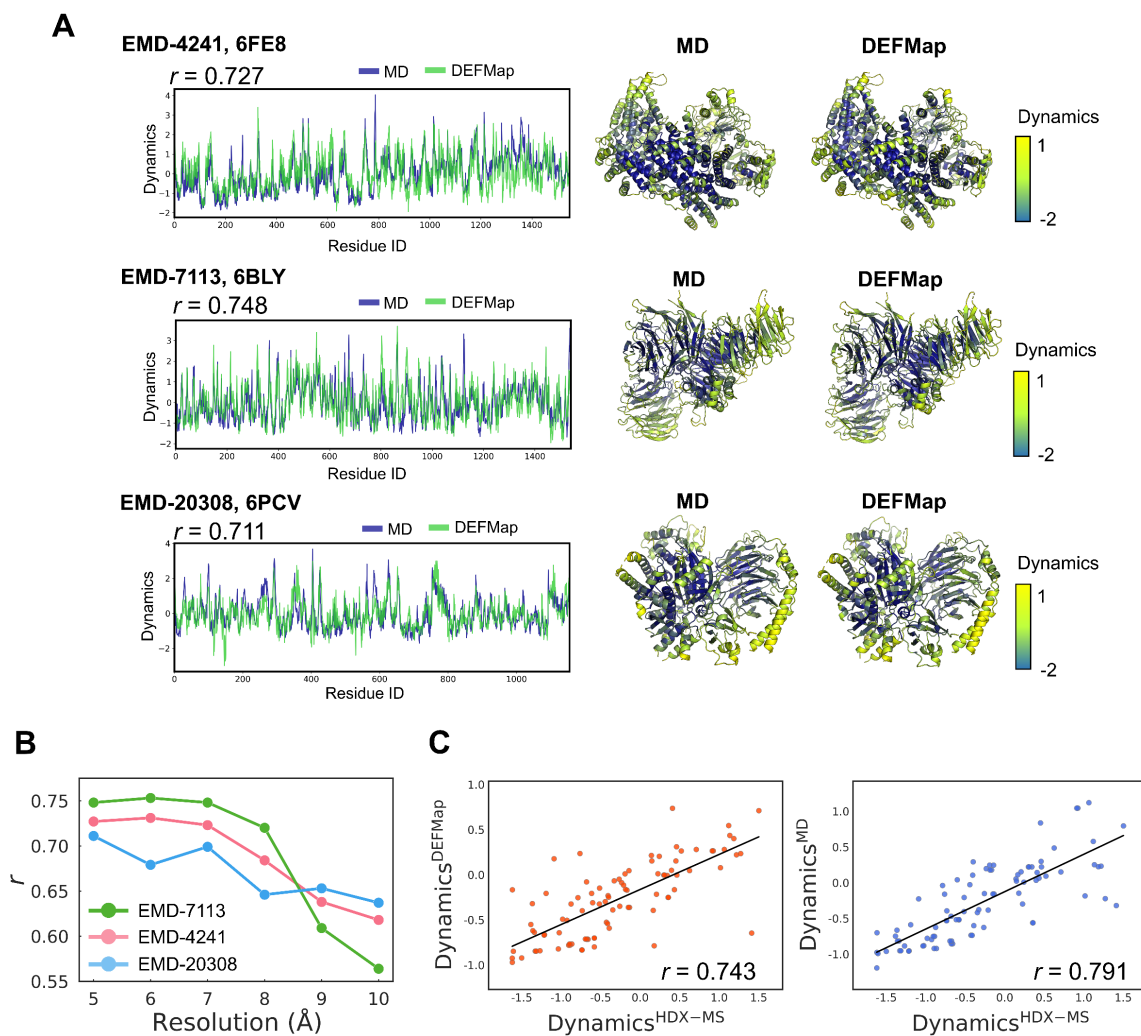

Figure 5 DEFMap-predicted results for cryo-EM maps not included in the training datasets. (A) Comparison of the dynamics values derived from MD simulations and DEFMap prediction for three kinds of the external datasets (EMD-4241, EMD-7113, and EMD-20308). The dynamics profiles of as a function of residue IDs and the mapping onto the 3D atomic models with different colors as indicated in the color bar are shown in left and right panels, respectively. The residue IDs in the profiles of the dynamics are numbered in accordance with their order in the corresponding PDB file; r denotes the correlation coefficient. The atomic models are derived from PDB [PDB ID: 6fe8, 6bly, and 6pcv]. (B) DEFMap performances on variable map resolution. The cryo-EM maps used for training dataset are low-pass-filtered to the target overall resolutions, and the resulting maps are used for the model training. (C) Comparison of the computed values with the experimentally derived dynamics data. The predicted values with DEFMap, those derived from MD simulations and the experimentally determined values are denoted as Dynamics^{DEFMap}, Dynamics^{MD}, and Dynamics^{HDX-MS}. The experimental data is derived from Ref. [34]. Dynamics^{DEFMap} and Dynamics^{MD} are converted to fragment-specific values by averaging the values over residues in each fragment detected by the experiments. r denotes a correlation coefficient.

Impact of DEFMap on Structural Biology Research

Biological phenomena are supported by numerous molecular interactions. When a ligand interacts with a biological macromolecule, the dynamics of the binding sites are generally suppressed by conformational stabilization. We attempted to detect ligand-induced modulation of the dynamics properties at the binding sites using DEFMap. For this purpose, three macromolecules, for which cryo-EM maps of both the unbound (apo form) and bound (holo form) states have been determined, were selected (apo, holo: EMD-20080, EMD-20081 [35]; EMD-9616, EMD-9622 [36]; EMD-3957, EMD-3956 [37]), and their dynamic properties were analyzed using DEFMap and MD simulations. As expected, significant suppression of the dynamics at the ligand-binding site was detected in both DEFMap and MD simulations (Fig. 6A).

DEFMap predicts the overall dynamic properties of a molecule. Mapping the difference in dynamics between the apo and holo forms on the 3D map visualizes the overall ligand-induced dynamics changes. Interestingly, for a protein associated with DNA methylation (Fig. 6B), *Arabidopsis* defective in meristem silencing 3 (DMS3)-RNA-directed DNA methylation 1 (RDM1) complex as an apo form and its complex with the ligand-defective RNA-directed DNA methylation 1 (DRD1) peptide as a holo form (apo, holo: EMD-20080, EMD-20081), additional suppression of the dynamics was observed in a region distant from the ligand-binding site, including the RDM1-DMS3 interaction interface and the hinge region of DMS3 involved in DRD1 peptide recognition (Fig. 6C). This suggests that DRD1 peptide binding stabilizes the RDM1-DMS3 complex formation and the conformation of the DMS3 hinge region. It should be noted that between the experimentally constructed atomic models of the apo and holo forms, no significant structural differences were found in these regions showing conformational stabilization (Fig. 6C). This observation indicates that we cannot find out the dynamic changes only from experimental data, emphasizing the usefulness of DEFMap in structural biology research.

Other advantages of dynamics analysis with DEFMap are that it is not limited by molecular size and does not require an atomic model. Conventional experimental approaches for extremely large molecules, such as viral particles, are significantly hampered by signal crowding. In addition, MD simulations of such molecules require high computational costs and high-resolution atomic models to capture reliable behavior. DEFMap can easily provide their dynamics properties if cryo-EM maps are available [38–41] (Fig. 6D).

Limitation in Prediction Using DEFMap

Prediction using a supervised learning technique for data not included in the training dataset generally performs poorly. DEFMap achieves high generalization performance because it learns a vast amount of local density data. Nevertheless, the prediction performances were relatively poor for the maps with extremely low resolutions. Furthermore, for unusual density data, such as those found in the transmembrane regions and those derived from post-translational modifications, the prediction performance is not expected because their data were not included in the current training datasets in consideration of the convenience of performing MD simulations. Expanding the training datasets of the flexible domain data and the unusual density data is a simple solution to these limitations, although it requires high computational costs. Development of an environment to use supercomputers, including the state-of-the-art Fugaku, will strongly support for this solution.

Conclusion

Here, we demonstrated the potential of cryo-EM data for quantitatively analyzing protein dynamics by developing a framework, DEFMap, using an integrated approach of an experimental technique, an MD simulation, and a DL technique. This study advances cryo-EM-based analysis techniques in the field of structural biology. The DEFMap code and the trained models are available in the GitHub repository (https://github.com/clinfo/DEFMap). It can be used by preparing an environment with TensorFlow, Keras, HTMD [42], and EMAN2 [43], all of which are freely available for academic use. However, it may not be easy for some users to construct a computational environment. We have recently developed and released ColabDEFMap, which can be run on Google Colaboratory, a Python programming and execution environment provided by Google (Fig. 7, accessible from the GitHub repository of DEFMap). This will allow more researchers to easily obtain the prediction results from DEFMap without the need for complex environment construction or programming. We hope that the widespread use of DEFMap will accelerate research in structural biology.
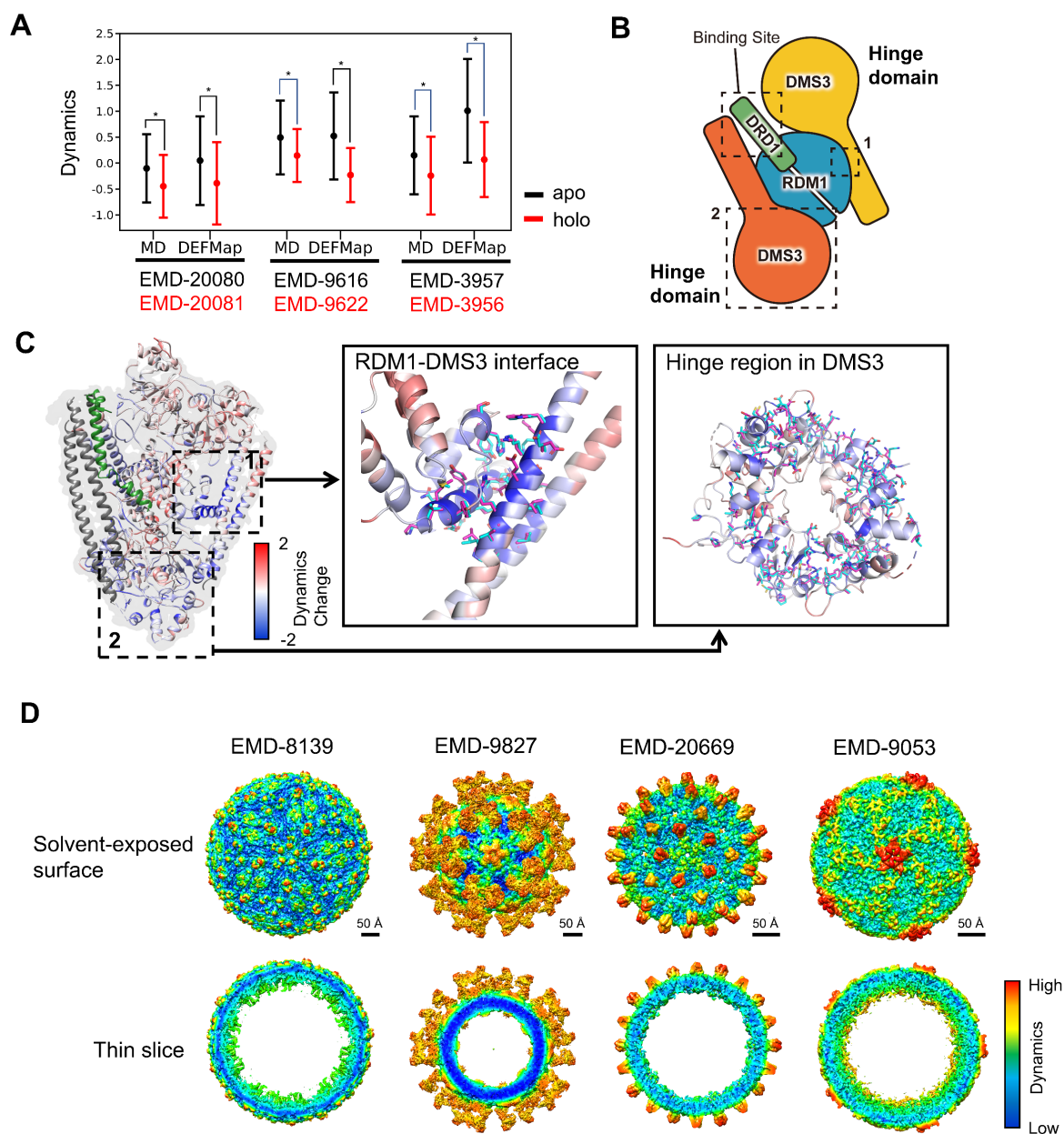
Figure 6 Case studies using DEFMap in structural biology research. (A) Comparisons of the dynamics for apo (red) and holo forms (black) at the ligand binding sites. The residues located at the ligand binding sites are identified by 5 Å cutoff from the ligands, and their averaged values are compared. The error bars indicate standard deviations (*$p <$ 0.01). (B) Schematic image of the protein assembly of RDM1-DMS3-DRD1 peptide complex. The regions indicated by black dashed rectangle with the labels, 1 and 2, corresponds to the expanded region in (C). (C) Mapping of the differences in the DEFMap-predicted dynamics of RDM1-DMS3 complex between apo and holo forms onto the atomic models. The mapped values are calculated by subtracting the values of the apo form from those of the holo forms. The resulting values are colored with different colors as indicated in the color bar, and lower values denote ligand-induced suppression of the dynamics. DRD1 peptide and disordered regions in apo form are colored by green and dark gray, respectively. The cryo-EM map is represented by light gray color. The expanded images of the regions indicated the black dashed rectangles in the overall image of the complex are shown. The atomic models are derived from PDB [PDB ID: 6ois and 6oit]. (D) Prediction of the dynamics for extremely large macromolecules using DEFMap. The predicted values are mapped onto the 3D cryo-EM maps with different colors as indicated in the color bar. The color range is defined by minimum and maximum values in the individual prediction. The scale bar represents 50 Å and is indicated
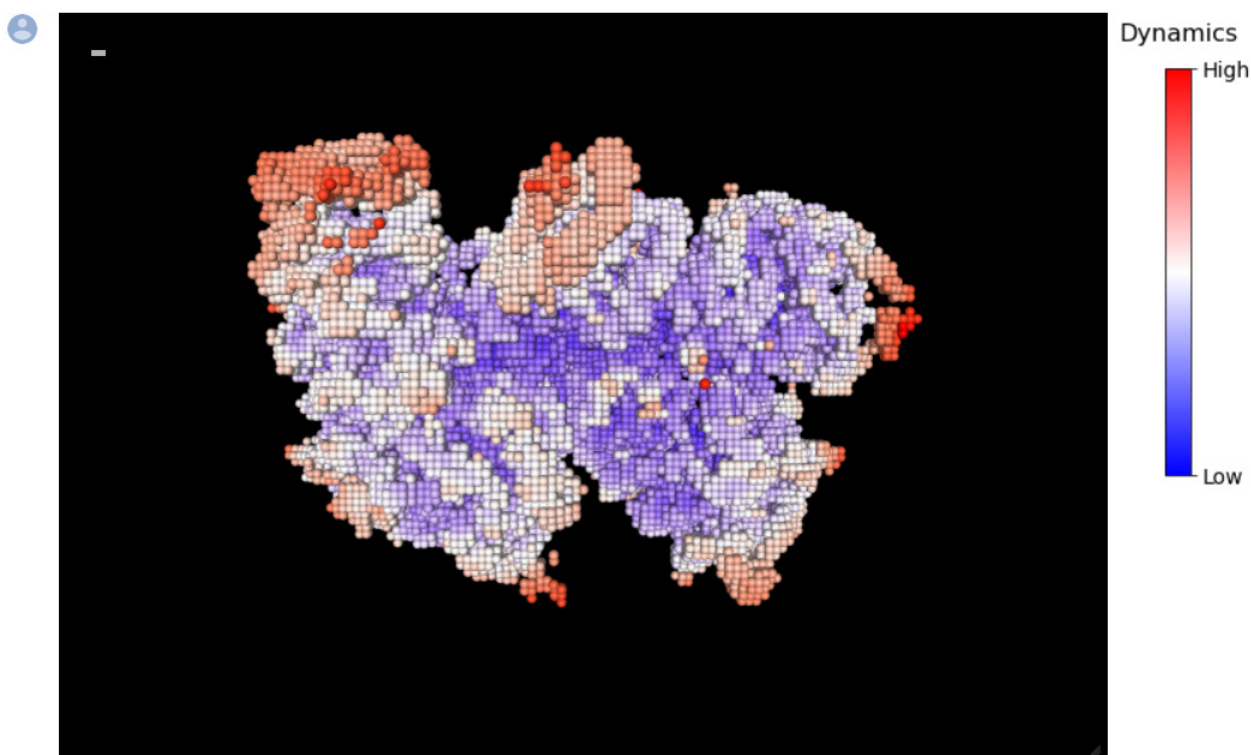
Figure 7 A screenshot of the visualization of the predicted results in ColabDEFMap.

Conflict of Interest

The authors declare no conflict of interests.

Author Contributions

S. M,. S. I., T. K. and Y. O. wrote the manuscript.

Data Availability

The models and the preprocessed input data are available in the Zenodo public repository with the DOI of https://doi.org/10.5281/zenodo.4317158.

References

[1] Masson, G. R., Burke, J. E., Ahn, N. G., Anand, G. S., Borchers, C., Brier, S., et al. Recommendations for performing, interpreting and reporting hydrogen deuterium exchange mass spectrometry (HDX-MS) experiments. Nat. Methods 16, 595–602 (2019). https://doi.org/10.1038/s41592-019-0459-y

[2] Hollingsworth, S. A., Dror, R. O. Molecular dynamics simulation for all. Neuron 99, 1129–1143 (2018). https://doi.org/10.1016/j.neuron.2018.08.011

[3]   Cheng, Y. Single-particle cryo-EM-How did it get here and where will it go. Science 361, 876–880 (2018). https://doi.org/10.1126/science.aat4346

[4]   Kühlbrandt, W. Cryo-EM enters a new era. eLife 3, e03678 (2014). https://doi.org/10.7554/eLife.03678

[5]   Cheng, Y. Single-particle Cryo-EM at crystallographic resolution. Cell 161, 450–457 (2015). https://doi.org/10.1016/j.cell.2015.03.049

[6]   Kato, T., Makino, F., Nakane, T., Terahara, N., Kaneko, T., Shimizu, Y., et al. CryoTEM with a cold field emission gun that moves structural biology into a new stage. Microsc. Microanal. 25, 998–999 (2019). https://doi.org/10.1017/s1431927619005725

[7]   Berman, H. M. The Protein Data Bank. Nucleic Acids Res. 28, 235–242 (2000). https://doi.org/10.1093/nar/28.1.235

[8]   Iudin, A., Korir, P. K., Salavert-Torres, J., Kleywegt, G. J., Patwardhan, A. EMPIAR: A public archive for raw electron microscopy image data. Nat. Methods 13, 387–388 (2016). https://doi.org/10.1038/nmeth.3806

[9]   Lawson, C. L., Baker, M. L., Best, C., Bi, C., Dougherty, M., Feng, P., et al. EMDataBank.org: Unified data resource for CryoEM. Nucleic Acids Res. 39, D456–D464 (2011). https://doi.org/10.1093/nar/gkq880

[10] Ulrich, E. L., Akutsu, H., Doreleijers, J. F., Harano, Y., Ioannidis, Y. E., Lin, J., et al. BioMagResBank. Nucleic Acids Res. 36, D402–D408 (2008). https://doi.org/10.1093/nar/gkm957

[11] Bekker, G. J., Kawabata, T., Kurisu, G. The biological structure model archive (BSM-Arc): An archive for in silico models and simulations. Biophys. Rev. 12, 371–375 (2020). https://doi.org/10.1007/s12551-020-00632-5

[12] Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589 (2021). https://doi.org/10.1038/s41586-021-03819-2

[13] Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., et al. AlphaFold protein structure database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res. 50, D439–D444 (2022). https://doi.org/10.1093/nar/gkab1061

[14] Zhong, E. D., Bepler, T., Berger, B., Davis, J. H. CryoDRGN: Reconstruction of heterogeneous cryo-EM structures using neural networks. Nat. Methods 18, 176–185 (2021). https://doi.org/10.1038/s41592-020-01049-4

[15] Bepler, T., Kelley, K., Noble, A. J., Berger, B. Topaz-Denoise: General deep denoising models for cryoEM and cryoET. Nat. Commun. 11, 5208 (2020). https://doi.org/10.1038/s41467-020-18952-1

[16] Lehtinen, J., Munkberg, J., Hasselgren, J., Laine, S., Karras, T., Aittala, M., et al. Noise2Noise: Learning image restoration without clean data. in 35th International Conference on Machine Learning, ICML vol. 7, pp. 4620–4631 (PMLR, 2018).

[17] Maddhuri Venkata Subramaniya, S. R., Terashi, G., Kihara, D. Protein secondary structure detection in intermediate-resolution cryo-EM maps using deep learning. Nat. Methods 16, 911–917 (2019). https://doi.org/10.1038/s41592-019-0500-1

[18] Ji, S., Xu, W., Yang, M., Yu, K. 3D Convolutional neural networks for human action recognition. IEEE Trans. Pattern Anal. Mach. Intell. 35, 221–231 (2013). https://doi.org/10.1109/TPAMI.2012.59

[19] Maturana, D., Scherer, S. VoxNet: A 3D convolutional neural network for real-time object recognition. in Rep. U.S. vol. 2015, pp. 922–928 (2015). https://doi.org/10.1109/IROS.2015.7353481

[20] Wu, Z., Song, S., Khosla, A., Yu, F., Zhang, L., Tang, X., et al. 3D ShapeNets: A deep representation for volumetric shapes. In Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Vol. 07-12, pp. 1912–1920 (2015). https://doi.org/10.1109/CVPR.2015.7298801

[21] Chen, H., Dou, Q., Yu, L., Qin, J., Heng, P. A. VoxResNet: Deep voxelwise residual networks for brain segmentation from 3D MR images. Neuroimage 170, 446–455 (2018). https://doi.org/10.1016/j.neuroimage.2017.04.041

[22] Kamnitsas, K., Ledig, C., Newcombe, V. F. J., Simpson, J. P., Kane, A. D., Menon, D. K., et al. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Med. Image Anal. 36, 61–78 (2017). https://doi.org/10.1016/j.media.2016.10.004

[23] Dou, Q., Chen, H., Yu, L., Zhao, L., Qin, J., Wang, D., et al. Automatic detection of cerebral microbleeds from MR Images via 3D convolutional neural networks. IEEE Trans. Med. Imaging 35, 1182–1195 (2016). https://doi.org/10.1109/TMI.2016.2528129

[24] Chung, J. M., Durie, C. L., Lee, J. Artificial intelligence in Cryo-Electron microscopy. Life 12, 1267 (2022). https://doi.org/10.3390/life12081267

[25] Matsumoto, S., Ishida, S., Araki, M., Kato, T., Terayama, K., Okuno, Y. Extraction of protein dynamics information from cryo-EM maps using deep learning. Nat. Mach. Intell. 3, 153–160 (2021). https://doi.org/10.1038/s42256-020-00290-y

[26] Matsumoto, S., Terayama, K., Okuno, Y. Extraction of protein dynamics hidden in Cryo-EM maps using deep

learning. SEIBUTSU BUTSURI 62, 193–197 (2022). https://doi.org/10.2142/biophys.62.193

[27] Cheng, Y., Grigorieff, N., Penczek, P. A., Walz, T. A primer to single-particle cryo-electron microscopy. Cell 161, 438–449 (2015). https://doi.org/10.1016/j.cell.2015.03.050

[28] Murata, K., Wolf, M. Cryo-electron microscopy for structural analysis of dynamic biological macromolecules. Biochim. Biophys. Acta Gen. Subj. 1862, 324–334 (2018). https://doi.org/10.1016/j.bbagen.2017.07.020

[29] Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., et al. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX 1–2, 19–25 (2015). https://doi.org/10.1016/j.softx.2015.06.001

[30] Vilas, J. L., Gómez-Blanco, J., Conesa, P., Melero, R., Miguel de la Rosa-Trevín, J., Otón, J., et al. MonoRes: Automatic and accurate estimation of local resolution for electron microscopy maps. Structure 26, 337–344 (2018). https://doi.org/10.1016/j.str.2017.12.018

[31] Kucukelbir, A., Sigworth, F. J., Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. Nat. Methods 11, 63–65 (2014). https://doi.org/10.1038/nmeth.2727

[32] Zhang, W., Lukoynova, N., Miah, S., Lucas, J., Vaughan, C. K. Insights into centromere DNA bending revealed by the Cryo-EM structure of the core centromere binding factor 3 with Ndc10. Cell Rep. 24, 744–754 (2018). https://doi.org/10.1016/j.celrep.2018.06.068

[33] Sun, Y., Zhang, Y., Hamilton, K., Manley, J. L., Shi, Y., Walz, T., et al. Molecular basis for the recognition of the human AAUAAA polyadenylation signal. Proc. Natl. Acad. Sci. U.S.A. 115, E1419–E1428 (2018). https://doi.org/10.1073/pnas.1718723115

[34] Cash, J. N., Urata, S., Li, S., Ravala, S. K., Avramova, L. V., Shost, M. D., et al. Cryo-electron microscopy structure and analysis of the P-Rex1-Gβγ signaling scaffold. Sci. Adv. 5, eaax8855 (2019). https://doi.org/10.1126/sciadv.aax8855

[35] Wongpalee, S. P., Liu, S., Gallego-Bartolomé, J., Leitner, A., Aebersold, R., Liu, W., et al. CryoEM structures of arabidopsis DDR complexes involved in RNA-directed DNA methylation. Nat. Commun. 10, 3916 (2019). https://doi.org/10.1038/s41467-019-11759-9

[36] Lan, P., Tan, M., Zhang, Y., Niu, S., Chen, J., Shi, S., et al. Structural insight into precursor tRNA processing by yeast ribonuclease P. Science 362, eaat6678 (2018). https://doi.org/10.1126/science.aat6678

[37] Abascal-Palacios, G., Ramsay, E. P., Beuron, F., Morris, E., Vannini, A. Structural basis of RNA polymerase III transcription initiation. Nature 553, 301–306 (2018). https://doi.org/10.1038/nature25441

[38] Kostyuchenko, V. A., Lim, E. X. Y., Zhang, S., Fibriansah, G., Ng, T. S., Ooi, J. S. G., et al. Structure of the thermally stable Zika virus. Nature 533, 425–428 (2016). https://doi.org/10.1038/nature17994

[39] Cao, L., Liu, P., Yang, P., Gao, Q., Li, H., Sun, Y., et al. Structural basis for neutralization of hepatitis A virus informs a rational design of highly potent inhibitors. PLoS Biol. 17, e3000229 (2019). https://doi.org/10.1371/journal.pbio.3000229

[40] Wu, W., Watts, N. R., Cheng, N., Huang, R., Steven, A. C., Wingfield, P. T. Expression of quasi-equivalence and capsid dimorphism in the Hepadnaviridae. PLoS Comput. Biol. 16, e1007782 (2020). https://doi.org/10.1371/journal.pcbi.1007782

[41] Liu, Y., Sheng, J., van Vliet, A. L. W., Buda, G., van Kuppeveld, F. J. M., Rossmann, M. G. Molecular basis for the acid-initiated uncoating of human enterovirus D68. Proc. Natl. Acad. Sci. U.S.A. 115, E12209–E12217 (2018). https://doi.org/10.1073/pnas.1803347115

[42] Doerr, S., Harvey, M. J., Noé, F., De Fabritiis, G. HTMD: High-throughput molecular dynamics for molecular discovery. J. Chem. Theory Comput. 12, 1845–1852 (2016). https://doi.org/10.1021/acs.jctc.6b00049

[43] Tang, G., Peng, L., Baldwin, P. R., Mann, D. S., Jiang, W., Rees, I., et al. EMAN2: An extensible image processing suite for electron microscopy. J. Struct. Biol. 157, 38–46 (2007). https://doi.org/10.1016/j.jsb.2006.05.009