Title: Extraction of Protein Dynamics Information Hidden in Cryogenic Electron Microscopy Maps Using Deep Learning

Author: Shigeyuki Matsumoto[1,†], Shoichi Ishida[2,†], Mitsugu Araki[3], Takayuki Kato[4], Kei Terayama[1,3,5,*] , Yasushi Okuno[1,3,*]

Author information

†These authors contributed equally to this study

Affiliations

[1] Medical Sciences Innovation Hub Program, RIKEN Cluster for Science, Technology and Innovation Hub, 1-7-22 Suehiro-cho, Tsurumi-ku, 230-0045 Kanagawa, Japan.

[2] Graduate School of Pharmaceutical Sciences, Kyoto University, 46-29 Yoshidashimo-Adachicho, Sakyo-ku, 606-8501 Kyoto, Japan.

[3] Graduate School of Medicine, Kyoto University, 53 Shogoin-Kawaharacho, Sakyo-ku, 606-8507 Kyoto, Japan.

[4] Institute for Protein Research, Osaka University, 3-2 Yamadaoka, Suita, 565-0871 Osaka, Japan.

[5] RIKEN Center for Advanced Intelligence Project, 1-4-1 Nihonbashi, Tokyo 103-0027, Japan.

Corresponding authors

*Correspondence should be addressed to Kei Terayama (kei.terayama@riken.jp) and Yasushi Okuno (okuno.yasushi.4c@kyoto-u.ac.jp).

Abstract

Technical breakthroughs in cryogenic electron microscopy (cryo-EM)-based single-particle analysis have enabled the structures of numerous proteins to be solved at atomic or near-atomic resolutions, including extremely large macromolecules whose structures could not be solved by conventional techniques. Determining the dynamics properties of these macromolecules, based on their solved structures, can further improve our understanding of their functional mechanisms. However, such analysis is often hampered by the large molecular size and complex structural assembly, making both experimental and computational approaches to determine dynamics properties challenging. Here, we

report a deep learning-based approach, DEFMap, to extract the dynamics information "hidden" in cryo-EM density maps. By relying only on cryo-EM maps, DEFMap successfully provided dynamics information equivalent to that determined from molecular dynamics (MD) simulations and experimental approaches at the atomic and residue levels. Additionally, DEFMap could detect dynamics changes associated with molecular recognition and the accompanying allosteric conformational stabilizations, which trigger various biological events such as signal transduction and enzyme catalysis. This approach will provide new insights into the functional mechanisms of biological molecules, accelerating modern molecular biology researches. Furthermore, this advanced strategy combining experimental data, deep learning approaches, and MD simulations would open a new multidisciplinary science area

Main

The functions of a protein under physiological conditions are associated with its three-dimensional (3D) structure and dynamic behavior in solution. High-resolution investigations of the structural and dynamics properties of proteins are, therefore, essential for elucidating the molecular mechanisms underlying their biological functions, such as the regulation of cellular signaling mediated by protein–protein interactions and metabolic reactions catalyzed by enzymes[1,2]. Various techniques have been developed to determine protein structures, such as X-ray crystallography, nuclear magnetic resonance (NMR), and cryogenic electron microscopy (cryo-EM) single-particle analysis (SPA)[3-6]. Dynamics information has also been quantitatively obtained through several experimental and computational approaches, such as NMR, hydrogen-deuterium exchange (HDX) mass spectrometry (MS)[7], and molecular dynamics (MD) simulations[8].

Recent breakthroughs in cryo-EM SPA[3,9] have enabled the determination of the structures of numerous biological molecules at atomic or near-atomic resolution[10,11], including extremely large and complex macromolecules[12-15] that cannot be solved using conventional techniques. However, investigating the dynamics of such molecules is technically challenging due to the large molecular sizes and complex structural assemblies involved.

To address these challenges, we developed a deep learning-based approach, named as "dynamics extraction from cryo-EM map" (DEFMap), to obtain dynamics information of proteins using a cryo-EM map alone (Fig. 1a and Supplementary Figure 1). The 3D cryo-EM maps solved by SPA are reconstructed from a large number of two-dimensional (2D) images of molecular particles identified in a micrograph[3-5]. Specimens used for cryo-

EM SPA are prepared by rapidly freezing a solution in which the proteins adopt variable conformations. Thus, it has been generally known that the local map intensities in reconstructed 3D cryo-EM maps tend to correlate with dynamics properties, i.e., weaker intensities correspond to more flexible regions of the macromolecule. Therefore, their dynamics properties could be "hidden" in the reconstructed cryo-EM maps. However, no strategies for quantitatively extracting the hidden dynamics information directly from 3D cryo-EM maps alone have been reported.

To quantitatively extract the hidden dynamics information at the atomic or residue level only from density data in the cryo-EM map, DEFMap is constructed by using a deep learning method and MD simulation (Fig. 1a). The deep learning method is designed to learn the relationship between local density data and dynamics information. Although the quantitative dynamics information is required to train the neural network model, it is not realistic to obtain enough dynamics information from existing experimental methods. To overcome this limitation, we performed MD simulations using cryo-EM map-derived atomic structures in the Protein Data Bank (PDB)[16], and calculated the root-mean-square fluctuation (RMSF) representing atomic fluctuations as dynamics information (see the stage of training data-construction in Fig. 1a). MD simulations have been widely used to elucidate the detailed dynamic behavior of biological molecules[8,17,18], as MD techniques and computing performance have advanced. Next, we trained the 3D-convolutional neural network (3D-CNN)[19-21] model to learn the relationship between cryo-EM density data and MD-derived RMSF values, to capture 3D patterns of the cryo-EM density data reflecting protein dynamics (see the training stage in Fig. 1a). 3D-CNN has been widely used to detect or classify patterns in 3D objects in various fields[22-24], and has exhibited remarkable performance in the processing of 3D cryo-EM maps[25-28]. Finally, using the trained 3D-CNN model, DEFMap can directly and quantitatively extract "hidden" dynamics information as RMSF values only from 3D cryo-EM density data of a new target protein (see the prediction stage in Fig. 1a). Note that at the prediction stage, DEFMap does not require any MD calculations. For visualization, DEFMap maps residue-specific RMSF values averaged over each residue after normalization (termed $\text{Dynamics}_{\text{DEFMap}}$) onto the corresponding atomic models.

At first, to evaluate the performance of DEFMap, we collected 25 cryo-EM maps from the Electron Microscopy Data Bank (EMDB)[29] (Supplementary Table 1) and used them to train the DEFMap 3D-CNN model (see Methods section for the details of the training dataset, MD simulation and training). Within the dataset, the map intensities tended to correlate, to some extent, with the residue-specific RMSF values obtained from MD (termed $\text{Dynamics}_{\text{MD}}$) as shown in Fig. 1b and Supplementary Table 2. However,

this correlation was insufficient to directly estimate the dynamics properties from the cryo-EM maps, as the average correlation coefficient *r* was only 0.464 ± 0.164. Using the same dataset, we performed leave-one-out cross-validation[30] to accurately evaluate the performance of DEFMap within the dataset. The results showed an improved correlation (r = 0.665 ± 0.124) between the $\text{Dynamics}_{\text{DEFMap}}$ and $\text{Dynamics}_{\text{MD}}$ outputs (Fig. 1c and Supplementary Table 2). This finding indicates that DEFMap can efficiently extract dynamics features from 3D cryo-EM density data.

We subsequently trained the 3D-CNN model of DEFMap using the full training dataset (25 cryo-EM maps) and then tested DEFMap on other three additional cryo-EM maps (EMDB/PDB entry; EMD-4241/6FE8[31], EMD-7113/6BLY[32], and EMD-20308/6PCV[33]) to further evaluate its potential for dynamics analysis (Supplementary Table 3). They were selected to exhibit distinct structural aspects in terms of secondary structure contents (α-helix/β-strand/others, see legend of Figure 2a). Particularly, experimental dynamics of EMD-20308/6PCV has been reported and is compared with $\text{Dynamics}_{\text{DEFMap}}$ later. The performance of DEFMap in extracting dynamics features from cryo-EM maps is illustrated in Figure 2. The dynamics values calculated by the trained DEFMap model showed good correlation with those derived from MD simulations at both the atomic (r = 0.704, 0.726 and 0.673, respectively) and residue (r = 0.727, 0.748 and 0.711, respectively) levels (Supplementary Fig. 2). The results showed that DEFMap could extract accurate dynamics information from the cryo-EM map data alone. Moreover, the mappings of $\text{Dynamics}_{\text{DEFMap}}$ onto the corresponding atomic models demonstrated that DEFMap captured conformational aspects such as rigidity in the protein interior and flexibility of the solvent-exposed secondary structure elements, with accuracy similar to that of MD simulations (Fig. 2a–c, upper panels). The $\text{Dynamics}_{\text{DEFMap}}$ data agreed well with the $\text{Dynamics}_{\text{MD}}$ data throughout the protein (Fig. 2a–c, bottom panels). However, in some regions, the DEFMap-based calculations failed to extract accurate dynamics properties. We hypothesize that a low local map resolution hinders accurate extraction of dynamics data. In fact, lower overall map resolutions (particularly those >8 Å) resulted in inferior performances (Fig. 2d and Supplementary Fig. 3). This tendency reflects the loss of detailed structural information in low-resolution maps. The apparent resolution dependence of the DEFMap model suggests that its performance will be enhanced alongside the continuous improvements in resolution associated with the development of advanced equipment, such as cold field emission guns[34,35], for cryo-EM data acquisition.

Confirming the consistency of the DEFMap predictions with experimentally derived dynamics properties is important for assessing the potential of the present method. Under the appropriate conditions, the dynamics of large proteins (such as those targeted by cryo-

EM) can also be experimentally determined by HDX-MS[7]; using this approach, the dynamics information at the peptide fragment level is obtained by monitoring the effects of deuterium incorporation into protein amide groups (Supplementary Figure 4). We compared the $\text{Dynamics}_{\text{DEFMap}}$ of EMD-20308 with the publicly available HDX-MS data[33]. The average $\text{Dynamics}_{\text{DEFMap}}$ data for each fragment detected in the HDX-MS experiments were well-correlated with the corresponding HDX rates ($\text{Dynamics}_{\text{HDX-MS}}$, r = 0.743) and $\text{Dynamics}_{\text{MD}}$ data (r = 0.791), confirming that DEFMap accurately captured the local dynamics and can provide insight equivalent to that derived from experimental approaches (Fig. 3).

Finally, we explored the potential of DEFMap for identifying molecular binding sites and investigating allosteric effects triggered by ligand binding. Ligand binding is a fundamental biological event and is often accompanied by the suppression of dynamics at the recognition interface. We therefore monitored dynamics changes associated with ligand-induced perturbations in the cryo-EM density maps. Particularly, we used DEFMap to detect ligand-induced dynamics changes for three pairs of apo/holo proteins not included in the training dataset (apo, holo: EMD-20080, EMD-20081[36]; EMD-9616, EMD-9622[37]; EMD-3957, EMD-3956[38], Supplementary Table 4). We found good agreement between $\text{Dynamics}_{\text{DEFMap}}$ and $\text{Dynamics}_{\text{MD}}$ profiles for the above pairs (Supplementary Figure 5). Moreover, the DEFMap-derived dynamics of the residues located near the binding partners were significantly suppressed by ligand binding (Fig. 4a, b and Supplementary Table 5), demonstrating that DEFMap could detect the conformational stabilization at the binding interface. Among the residues located at the interfaces (Supplementary Table 5), we observed significant dynamics suppression in the regions extensively discussed in the literature[36-38]. This finding suggests that DEFMap can identify the key interactions involved in complex formation using density data (Fig. 4b). We found additional dynamics suppression for regions distant from the ligand-binding site in the EMD-20080/EMD-20081 pair [Arabidopsis defective in meristem silencing 3 (DMS3)-RNA-directed DNA methylation 1 (RDM1) complex with defective RNA-directed DNA methylation 1 (DRD1) peptide][36] (Fig. 4c). In the apo state, DMS3 dimers establish contacts with a RDM1 dimer through their coiled-coil arms (middle panel in Fig. 4c). In the holo conformation, the binding partner, DRD1 peptide, is recognized by the RDM1 dimer, the coiled-coil arms of a DMS3 dimer (dimer 1 in Fig. 4c), and a hinge domain of an opposite DMS3 dimer (dimer 2 in Fig. 4c). DEFMap-based dynamics analysis showed that the binding of the DRD1 peptide induced suppression of dynamics at the RDM1-DMS3 interface and at the hinge domains in DMS3 dimer 1 (Fig. 4c), indicating that ligand binding allosterically stabilizes these regions. Interestingly, the

static models constructed from the cryo-EM maps showed no substantial differences between the apo and holo proteins in these regions (Supplementary Figure 6). This finding further highlights the potential of DEFMap to provide dynamics information that cannot be determined from static tertiary structures.

Cryo-EM SPA opened a new era in molecular biology by enabling researchers to solve 3D structures of macromolecules and supramolecules at high resolution. In this study, we showed that a deep neural network-based approach, DEFMap, has successfully extracted "hidden" dynamics information only from static 3D cryo-EM density maps. The DEFMap approach is not restricted by the molecular sizes and the complexity of the systems because DEFMap infers dynamics from local density data, as shown in the above results. We expect that complementary use of DEFMap with conventional cryo-EM SPA will accelerate the elucidation of the molecular mechanisms underlying biological events. Furthermore, DEFMap will help researchers readily access the dynamics properties of biological molecules, as this open-source tool does not require additional expensive and/or time-consuming experiments or in-depth expertise. We believe that in the future, DEFMap will accelerate data-driven structural investigations aiming to understand protein functions and to develop strategies for molecular targeted therapy against various diseases. This study bridges experimental data, deep learning approaches, and MD simulations, allowing for accurate extraction of hidden information, which humans could not utilize, from the data. It is expected that this advanced strategy would open a new multidisciplinary science area which integrate experimental science, simulation science, and data science.

Figure legends

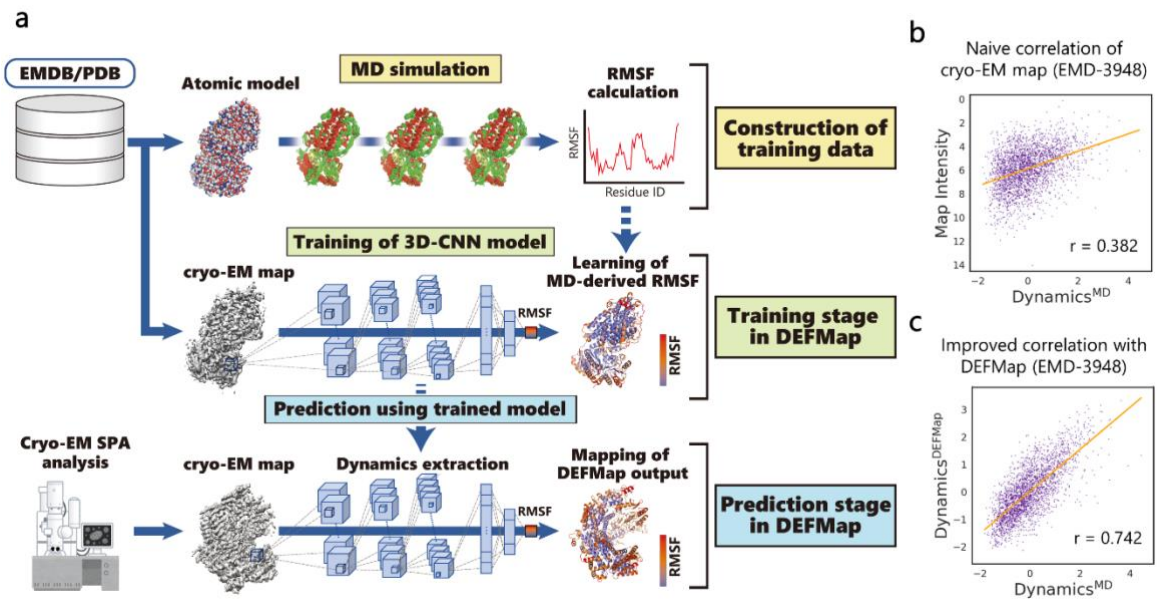

Fig. 1: DEFMap-based extraction of dynamics features from cryo-EM maps. a, Whole workflow of DEFMap. Model training is carried out using EMDB/PDB-derived macromolecules which are conveniently handled in all-atom MD simulations. In the stage of training data-construction, the dynamics properties are derived from RMSF values (for heavy atoms) calculated from MD simulations. In the training stage, a 3D-CNN model in DEFMap (Supplementary Figure 1) learns the relationship between dynamics values and density data at the corresponding positions. In the prediction stage, for other cryo-EM maps not included in the training dataset, the trained model predicts dynamics values based on input density data. In this study, 25 macromolecules were used to validate and train the DEFMap model, and other 9 macromolecules were used for dynamics predictions using the trained model and further structural analyses. **b,** A correlation plot between the cryo-EM map intensities and $\text{Dynamics}_{\text{MD}}$ for EMD-3948 (entry 1 in Supplementary Table 2). $\text{Dynamics}_{\text{MD}}$ were calculated from RMSF values derived from MD simulations. **c,** A correlation plot between $\text{Dynamics}_{\text{DEFMap}}$ and $\text{Dynamics}_{\text{MD}}$ for EMD-3984 (entry 1 in Supplementary Table 2). *r* denotes the correlation coefficient.

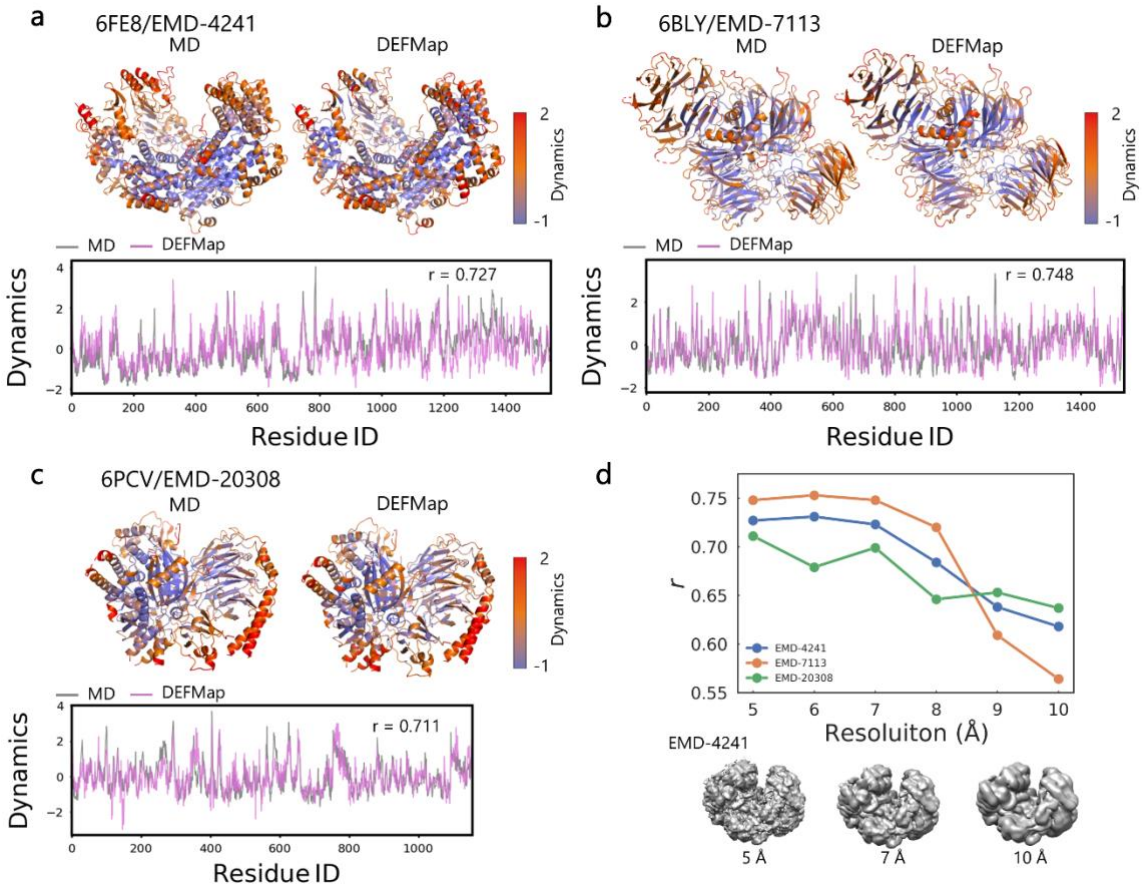

Fig. 2: DEFMap performance for three proteins not included in the training dataset. a-c, $\text{Dynamics}_{\text{MD}}$ and $\text{Dynamics}_{\text{DEFMap}}$ outputs for EMD-4241/6FE8 (**a**), EMD-7113/6BLY (**b**), and EMD-20308/6PCV (**c**). The cryo-EM maps were preprocessed by a 5 Å low-pass filter using the program EMAN2[39]. The proportions of secondary structure contents (α-helix/β-strand/others) are 0.56/0.06/0.38 (**a**), 0.05/0.43/0.52 (**b**), and 0.30/0.27/0.43 (**c**). The residue-specific values were calculated by averaging the normalized atomic values over each residue. The resulting values were mapped onto 3D atomic models (upper panels) using different colors as indicated in the color bars. The 3D $\text{Dynamics}_{\text{MD}}$ and $\text{Dynamics}_{\text{DEFMap}}$ maps are shown on the left- and right-hand sides, respectively, of each panel. The bottom panels show the $\text{Dynamics}_{\text{MD}}$ (gray) and $\text{Dynamics}_{\text{DEFMap}}$ (pink) profiles as a function of the residue IDs, numbered according to their order in the corresponding PDB file; *r* denotes the correlation coefficient between $\text{Dynamics}_{\text{MD}}$ and $\text{Dynamics}_{\text{DEFMap}}$. **d,** Dependence of DEFMap performance on map resolution. The plot shows the correlation coefficients between $\text{Dynamics}_{\text{MD}}$ and $\text{Dynamics}_{\text{DEFMap}}$ obtained for different map resolutions. The maps of the training and other three datasets were preprocessed to identical resolutions with low-pass filters (e.g.,

the extraction of the Dynamics$_{DEFMap}$ data from the 10.0 Å low-pass filtered test map was carried out using the model trained by the 10.0 Å maps). The low-pass filtered cryo-EM maps of EMD-4241 at 5, 7, and 10 Å are shown at the bottom of the panel.

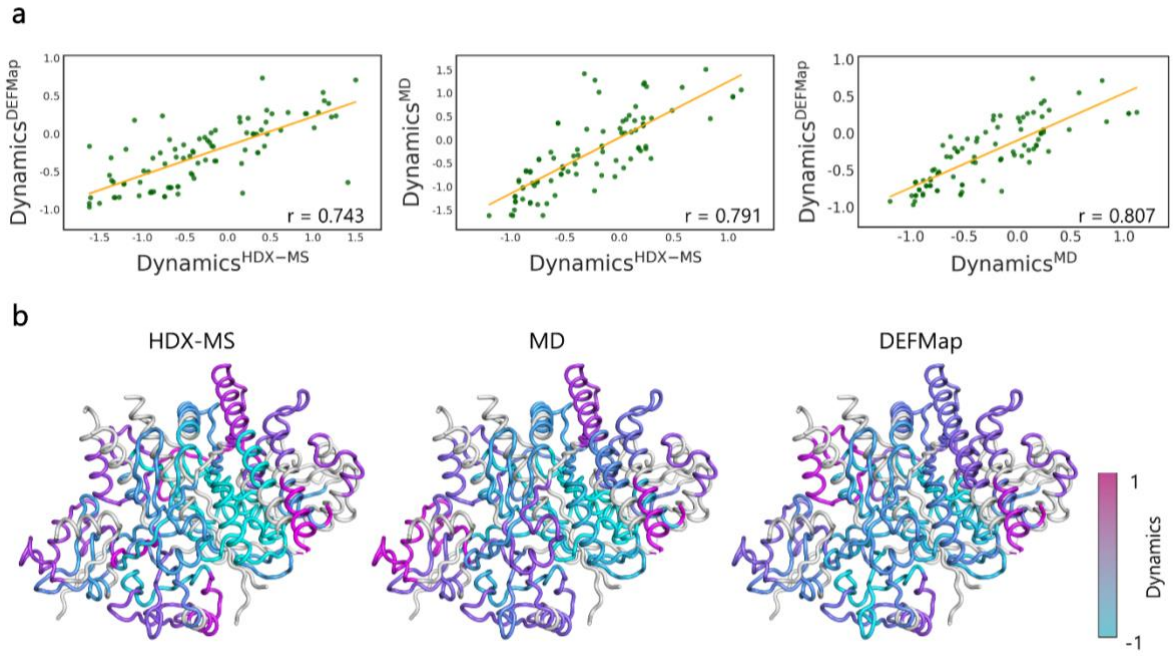

Fig. 3: The relationship between DEFMap-determined and experimentally derived dynamics properties. a, A correlation plot of DEFMap-derived and HDX-MS-derived dynamics. The HDX exchange rates at 10^4 s for Rac exchanger 1 complexed with G protein beta gamma subunits[33] were normalized within the detected fragments and used as the experimental dynamics data. The residue-specific dynamics obtained from MD and DEFMap were converted to fragment-specific values by averaging them over each fragment in the atomic model. The HDX-MS *vs*. DEFMap (r = 0.743), HDX-MS *vs*. MD (r = 0.807), and MD *vs*. HDX-MS (r = 0.791) correlation plots are shown in the left, middle, and right panels, respectively, along with their corresponding regression lines (orange). **b,** The spatial distribution of the fragment-specific dynamics in the atomic models. The representative fragments-specific dynamics are mapped onto the atomic model using different colors as indicated in the color bar (HDX-MS, left; MD, middle; DEFMap, right). The fragments not detected in the HDX-MS experiments are colored gray.

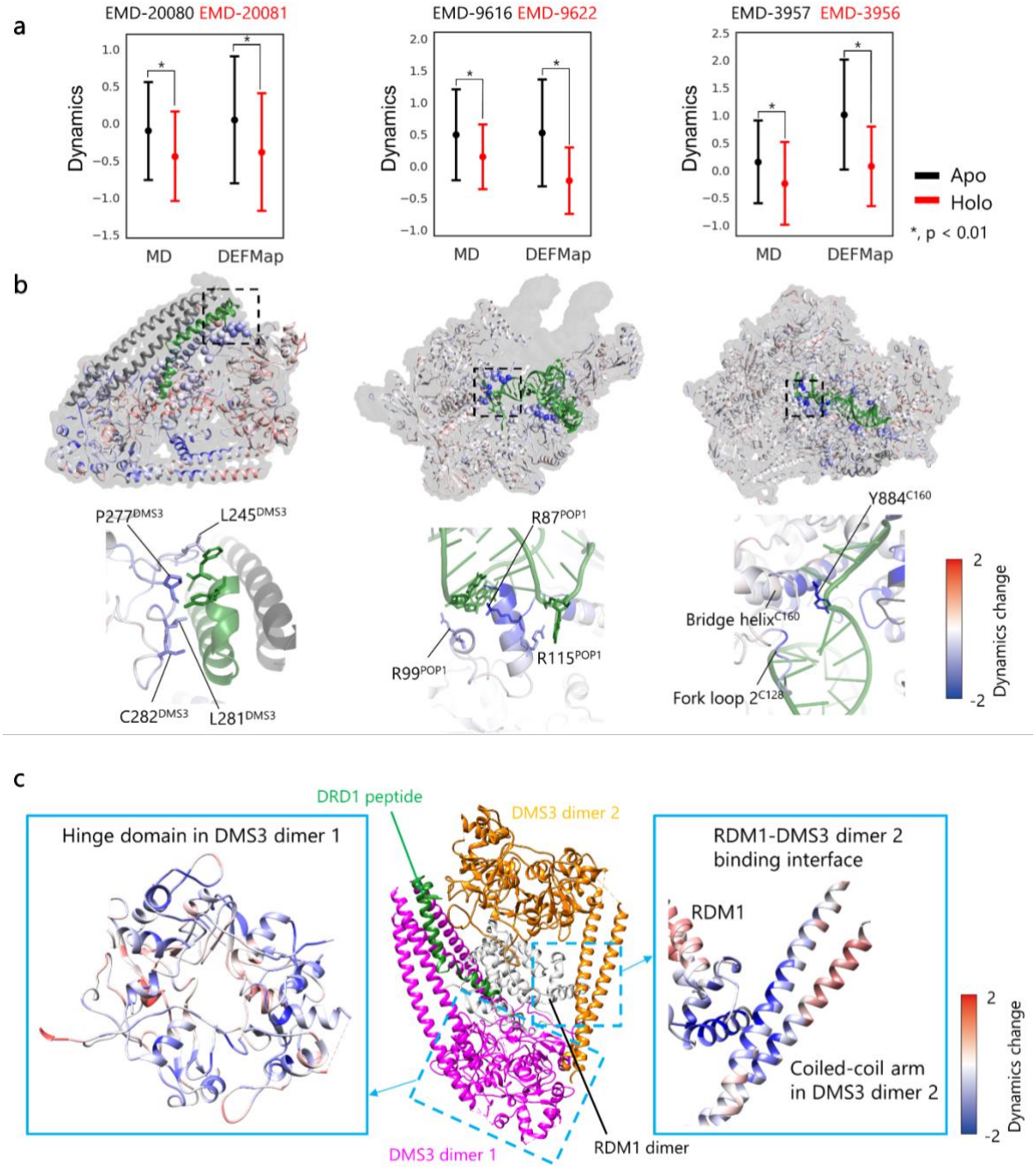

Fig. 4: DEFMap-based detection of dynamics changes induced by ligand binding. a, Ligand-induced dynamics suppression at the binding interfaces. The dynamics at the binding interfaces in apo (black) and holo (red) proteins were calculated by averaging the $\text{Dynamics}_{\text{MD}}$ and $\text{Dynamics}_{\text{DEFMap}}$ data for residues located within 5 Å of the binding partners, and the error bars indicate standard deviations (* symbols denote statistical significance at $p < 0.01$). **b,** The spatial distributions of ligand-induced dynamics changes derived from DEFMap. The dynamics changes were calculated by subtracting the

$Dynamics_{DEFMap}$ of the apo state from those of the holo state and mapped onto the holo atomic models using different colors as indicated in the color bar (lower values denote regions with suppressed dynamics). The binding partners and disordered regions in the apo proteins are colored green and dark gray, respectively. In the overall view (upper panels), the Cα atoms of the residues located within 5 Å of the binding partners are represented as spheres, and the cryo-EM maps are colored gray. The lower panels show enlarged views of the regions indicated by dashed rectangles in the upper panels. The relevant residues extensively discussed in the literature are represented as sticks, with the subunit names indicated as superscripts. **c,** DEFMap detection of ligand-induced allosteric change. Enlarged views of the regions discussed in the main text are shown in the left- and right-hand side panels, respectively. The color bar represents the ligand-induced dynamics changes.

Methods

Molecular dynamics simulations

All-atom MD simulations for macromolecules in the present datasets (34 molecules) are carried out to obtain RMSF values that used for training of a deep learning model in DEFMap or evaluation of the performance. The selected macromolecules included proteins or the complex with DNA/RNA which are conveniently handled in the simulations. The initial coordinates used in the simulations were obtained from the PDB (https://www.rcsb.org) and processed using the Structure Preparation module of the Molecular Operating Environment (MOE) software (Chemical Computing Group, Montreal, Canada). Briefly, loops were modeled for disordered regions containing less than seven residues, and other non-natural N-termini and C-termini were capped with acetyl and formyl groups, respectively. The addition of hydrogen atoms and the generation of topology files were carried out using the pdb2gmx module in the GROMACS package[40]. All MD simulations were carried out with periodic boundary conditions (PBC), using GROMACS[40] on an NVIDIA GeForce GTX 1080 GPU. The shape of the periodic cell was octahedron. The Amber ff99SB-ILDN force field was used for proteins, nucleotides, and ions[41], while the TIP3P potential was used for water[42]. Water molecules were placed around the substrate model with an encompassing distance of 10 Å, and counterions (NaCl) were added to neutralize the system. Electrostatic interactions were calculated using the particle mesh Ewald (PME) method[43] with a cutoff radius of 10 Å, and a nonbonded cutoff of 10 Å was applied for van der Waals interactions. The P-LINCS algorithm was used to constrain all bond lengths to their equilibrium values[44]. After energy minimization of the fully solvated models, the resulting systems were equilibrated for 100 ps under constant number of molecules, volume, and temperature (NVT) conditions, followed by a 100 ps run under constant number of molecules, pressure, and temperature (NPT) conditions, with the heavy atoms of the macromolecules held in fixed positions. The temperature was maintained at 298 K by velocity rescaling with a stochastic term[45], while Parrinello-Rahman pressure coupling[46] was used to keep the pressure at 1 bar, with the temperature and pressure time constants set to 0.1 and 2 ps, respectively. Subsequently, production runs of 20 ns were carried out under NPT conditions without positional restraints. After PBC corrections, the generated trajectories were aligned using overall Cα atoms, followed by the calculation of the RMSF values (Å) for heavy atoms using the rmsf module of GROMACS. The logarithmic RMSFs were then used to represent dynamics properties.

Data preparation for DEFMap

Twenty-five cryo-EM density maps were selected and downloaded from the EMDB/PDB (Supplementary Table 1, average overall resolution = 3.62 ± 0.46) to train the 3D-CNN model. Their resolutions are relatively high compared to the average overall resolution of cryo-EM maps deposited in 2019 (5.6 Å). To evaluate the DEFMap potential for dynamics analysis, the additional nine cryo-EM density maps were selected and downloaded from the EMDB/PDB (Supplementary Tables 3 and 4). The maps were rescaled to 1.5 Å/pixel and low-pass filtered using EMAN2.3[39]. Subsequently, the intensities were normalized within each map, and any negative values were discarded. Each grid point in the maps was associated with the MD-derived logarithmic RMSF of the nearest atom in the voxelized coordinates. The resulting maps were sub-voxelized to generate the input density data, with 10^3 voxels distributed over 15^3 $Å^3$. The training data were augmented by x-, y-, and z-axis rotations of 90°, 180°, and 270°. The preprocessed maps with an overall resolution of 5 Å were mainly used in this study unless otherwise stated. Voxelization of the atomic models was carried out by high-throughput molecular dynamics (HTMD)[47]. All preprocessing procedures were performed using Python.

Construction of deep neural networks

The architecture of the neural network used in DEFMap included 3D convolutional blocks and dense blocks. The 3D convolutional blocks consisted of three, 3D convolutional layers with leaky ReLU activation, max pooling, and dropout. The kernel size for convolution, the maximum pooling size, and the dropout ratio were set to 4 × 4 × 4, 2 × 2, and 0.2, respectively. Different filter sizes (64, 128, and 256) were applied to the three 3D convolutional layers. The dense blocks consisted of two dense layers of 1,024 units with leaky ReLU activation and a dense layer of a unit with identity activation. The epochs, learning rate, and batch size hyperparameters were set to 50, 0.00005, and 128, respectively, and early stopping with a patience interval of three epochs was utilized to avoid overfitting. All models were trained using an NVIDIA Tesla V100 GPU with 16 GB of memory. The Keras[48] library with Tensorflow backend was employed in the calculations.

Postprocessing and visualization of output from the neural networks

Postprocessing of the atomic dynamics values calculated by DEFMap for the subsequent validation and analyses was carried out as follows: the output values were normalized and then averaged over each residue; the residue-specific values ($\text{Dynamics}_{\text{DEFMap}}$) were then assigned to the atomic models as temperature factors with HTMD[47] and visualized using the PyMOL[49] and UCSF Chimera[50] programs. All postprocessing procedures were performed using Python.

Code availability

The Python code used to implement DEFMap is available on GitHub at https://github.com/clinfo/DEFMap. DEFMap is licensed under MIT licence.

Data availability

All datasets used in this paper are publicly available via the PDB (https://www.rcsb.org/). Detailed descriptions of the datasets are provided in Supplementary Tables 1, 2, and 3.

References

1 Boehr, D. D., Nussinov, R. & Wright, P. E. The role of dynamic conformational ensembles in biomolecular recognition. *Nat Chem Biol* **5**, 789-796, doi:10.1038/nchembio.232 (2009).

2 Kohen, A. Role of Dynamics in Enzyme Catalysis: Substantial versus Semantic Controversies. *Accounts of Chemical Research* **48**, 466-473, doi:10.1021/ar500322s (2014).

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

4 Cheng, Y., Grigorieff, N., Penczek, P. A. & Walz, T. A primer to single-particle cryo-electron microscopy. *Cell* **161**, 438-449, doi:10.1016/j.cell.2015.03.050 (2015).

5 Murata, K. & Wolf, M. Cryo-electron microscopy for structural analysis of dynamic biological macromolecules. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1862**, 324-334, doi:10.1016/j.bbagen.2017.07.020 (2018).

6 Nitta, R., Imasaki, T. & Nitta, E. Recent progress in structural biology: lessons from our research history. *Microscopy (Oxf)*, doi:10.1093/jmicro/dfy022 (2018).

7 Masson, G. R. *et al.* Recommendations for performing, interpreting and reporting hydrogen deuterium exchange mass spectrometry (HDX-MS) experiments. *Nat Methods* **16**, 595-602, doi:10.1038/s41592-019-0459-y (2019).

8 Hollingsworth, S. A. & Dror, R. O. Molecular Dynamics Simulation for All. *Neuron* **99**, 1129-1143, doi:10.1016/j.neuron.2018.08.011 (2018).

9 Kuhlbrandt, W. Cryo-EM enters a new era. *Elife* **3**, e03678, doi:10.7554/eLife.03678 (2014).

10 Cheng, Y. Single-Particle Cryo-EM at Crystallographic Resolution. *Cell* **161**, 450-457, doi:10.1016/j.cell.2015.03.049 (2015).

11 Merk, A. *et al.* Breaking Cryo-EM Resolution Barriers to Facilitate Drug Discovery. *Cell* **165**, 1698-1707, doi:10.1016/j.cell.2016.05.040 (2016).

12 Gremer, L. *et al.* Fibril structure of amyloid-beta(1-42) by cryo-electron microscopy. *Science* **358**, 116-119, doi:10.1126/science.aao2825 (2017).

13 Kato, T., Makino, F., Miyata, T., Horváth, P. & Namba, K. Structure of the native supercoiled flagellar hook as a universal joint. *Nature Communications* **10**, doi:10.1038/s41467-019-13252-9 (2019).

14 Kujirai, T. *et al.* Structural basis of the nucleosome transition during RNA polymerase II passage. *Science* **362**, 595-598, doi:10.1126/science.aau9904 (2018).

15 Li, X. *et al.* A unified mechanism for intron and exon definition and back-splicing. *Nature* **573**, 375-380, doi:10.1038/s41586-019-1523-6 (2019).

16 Berman, H. M. The Protein Data Bank. *Nucleic Acids Research* **28**, 235-242, doi:10.1093/nar/28.1.235 (2000).

17 Hollingsworth, S. A. *et al.* Cryptic pocket formation underlies allosteric modulator selectivity at muscarinic GPCRs. *Nat Commun* **10**, 3289, doi:10.1038/s41467-019-11062-7 (2019).

18 Plattner, N., Doerr, S., De Fabritiis, G. & Noe, F. Complete protein-protein association kinetics in atomic detail revealed by molecular dynamics simulations and Markov modelling. *Nat Chem* **9**, 1005-1011, doi:10.1038/nchem.2785 (2017).

19 Ji, S., Xu, W., Yang, M. & Yu, K. 3D Convolutional neural networks for human action recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **35**, 221-231, doi:10.1109/TPAMI.2012.59 (2013).

20 Maturana, D. & Scherer, S. VoxNet: A 3D Convolutional Neural Network for real-time object recognition. in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* 922-928 (2015).

21 Zhirong, W. *et al.* 3D ShapeNets: A deep representation for volumetric shapes. in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 1912-1920 (2015).

22 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, doi:10.1016/j.neuroimage.2017.04.041 (2018).

23 Kamnitsas, K. *et al.* Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Medical Image Analysis* **36**, 61-78, doi:10.1016/j.media.2016.10.004 (2017).

24 Qi, D. *et al.* Automatic Detection of Cerebral Microbleeds From MR Images via 3D Convolutional Neural Networks. *IEEE Trans Med Imaging* **35**, 1182-1195, doi:10.1109/TMI.2016.2528129 (2016).

25 Avramov, T. K. *et al.* Deep learning for validating and estimating resolution of cryo-electron microscopy density maps. *Molecules* **24**, 1181, doi:10.3390/molecules24061181 (2019).

26 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, doi:10.1038/s41592-019-0500-1 (2019).

27 Mostosi, P., Schindelin, H., Kollmannsberger, P. & Thorn, A. Automated interpretation of Cryo-EM density maps with convolutional neural networks. *bioRxiv*, 644476, doi:10.1101/644476 (2019).

28 Xu, K., Wang, Z., Shi, J., Li, H. & Zhang, Q. C. A2-Net: Molecular Structure Estimation from Cryo-EM Density Volumes. *Proceedings of the AAAI Conference on Artificial Intelligence* **33**, 1230-1237 (2019).

29 Lawson, C. L. *et al.* EMDataBank.org: unified data resource for CryoEM. *Nucleic Acids Research* **39**, D456-D464, doi:10.1093/nar/gkq880 (2010).

30 Goodfellow, I., Bengio, Y. & Courville, A. *Deep Learning*. (MIT Press, 2016).

31 Zhang, W., Lukoyanova, 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 Reports* **24**, 744-754, doi:10.1016/j.celrep.2018.06.068 (2018).

32 Sun, Y. *et al.* Molecular basis for the recognition of the human AAUAAA polyadenylation signal. *Proceedings of the National Academy of Sciences* **115**, E1419-E1428, doi:10.1073/pnas.1718723115 (2018).

33 Cash, J. N. *et al.* Cryo–electron microscopy structure and analysis of the P-Rex1–Gβγ signaling scaffold. *Science Advances* **5**, doi:10.1126/sciadv.aax8855 (2019).

34 Hamaguchi, T. *et al.* A new cryo-EM system for single particle analysis. *Journal of Structural Biology* **207**, 40-48, doi:10.1016/j.jsb.2019.04.011 (2019).

35 Kato, T. *et al.* CryoTEM with a Cold Field Emission Gun That Moves Structural Biology into a New Stage. *Microscopy and Microanalysis* **25**, 998-999, doi:10.1017/s1431927619005725 (2019).

36 Wongpalee, S. P. *et al.* CryoEM structures of Arabidopsis DDR complexes involved in RNA-directed DNA methylation. *Nature Communications* **10**, doi:10.1038/s41467-019-11759-9 (2019).

37 Lan, P. *et al.* Structural insight into precursor tRNA processing by yeast ribonuclease P. *Science* **362**, doi:10.1126/science.aat6678 (2018).

38 Abascal-Palacios, G., Ramsay, E. P., Beuron, F., Morris, E. & Vannini, A. Structural basis of RNA polymerase III transcription initiation. *Nature* **553**, 301-306, doi:10.1038/nature25441 (2018).

39 Tang, G. *et al.* EMAN2: An extensible image processing suite for electron microscopy. *Journal of Structural Biology* **157**, 38-46, doi: 10.1016/j.jsb.2006.05.009 (2007).

40 Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1-2**, 19-25, doi:10.1016/j.softx.2015.06.001 (2015).

41 Lindorff-Larsen, K. *et al.* Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins*, **78**, 1950-1958, doi:10.1002/prot.22711 (2010).

42 Jorgensen, W. L. & Thomas, L. L. Perspective on Free-Energy Perturbation Calculations for Chemical Equilibria. *J Chem Theory Comput* **4**, 869-876, doi:10.1021/ct800011m (2008).

43 Darden, T., York, D. & Pedersen, L. Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems. *J Chem Phys* **98**, 10089-10092, doi:10.1063/1.464397 (1993).

44 Hess, B. P-LINCS: A parallel linear constraint solver for molecular simulation. *J Chem Theory Comput* **4**, 116-122, doi:10.1021/ct700200b (2008).

45 Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *The Journal of Chemical Physics* **126**, doi:10.1063/1.2408420 (2007).

46 Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics* **52**, 7182-7190, doi:10.1063/1.328693 (1981).

47 Doerr, S., Harvey, M. J., Noe, F. & De Fabritiis, G. HTMD: High-Throughput Molecular Dynamics for Molecular Discovery. *J Chem Theory Comput* **12**, 1845-1852, doi:10.1021/acs.jctc.6b00049 (2016).

48 Chollet, F. Keras. https://keras.io (2015).

49 Schrdinger, L. L. C. The PyMOL Molecular Graphics System. (2015).

50 Pettersen, E. F. *et al.* UCSF Chimera—A visualization system for exploratory research and analysis. *Journal of Computational Chemistry* **25**, 1605-1612, doi:10.1002/jcc.20084 (2004).

Acknowledgements

This work was supported by MEXT as “Priority Issue on Post K computer (Building Innovative Drug Discovery Infrastructure Through Functional Control of Biomolecular Systems)” and JSPS KAKENHI Grant Number JP17K15106 to SM.

Author contributions

S.M., S.I., K.T., and Y.O. designed the study. S.M., and S.I. performed the pre-processing of cryo-EM maps, design the neural networks, and carried out model training and DEFMap analyses. S.M. and M.A. performed the MD calculations. S.M. and T.K. prepared the original cryo-EM map data. K.T. and Y.O. conceived the project. S.M. and S.I. wrote the manuscript. All authors discussed the research, edited the manuscript, and approved its final version.

Declaration of Interests

The authors declare no competing interests.