Structural biology has evolved greatly due to the advances introduced in fields like electron microscopy. This image-capturing technique, combined with improved algorithms and current data processing software, allows the recovery of different conformational states of a macromolecule, opening new possibilities for the study of its flexibility and dynamic events. However, the ensemble analysis of these different conformations, and in particular their placement into a common variable space in which the differences and similarities can be easily recognized, is not an easy matter. To simplify the analysis of continuous heterogeneity data, this work proposes a new automatic algorithm that relies on a mathematical basis defined over the sphere to estimate the deformation fields describing conformational transitions among different structures. Thanks to the approximation of these deformation fields, it is possible to describe the forces acting on the molecules due to the presence of different motions. It is also possible to represent and compare several structures in a low-dimensional mapping, which summarizes the structural characteristics of different states. All these analyses are integrated into a common framework, providing the user with the ability to combine them seamlessly. In addition, this new approach is a significant step forward compared with principal component analysis and normal mode analysis of cryo-electron microscopy maps, avoiding the need to select components or modes and producing localized analysis.

1. Introduction

The application in electron microscopy of techniques such as cryo-electron microscopy (cryo-EM), single-particle analysis (SPA) (Carroni & Saibil, 2016) or electron cryo-tomography (Schur, 2019) has proven to be a versatile tool to trace high-resolution structures. In particular, cryo-EM SPA has proven to be especially good at providing not only one structure, but a series of them, with most methods aiming to resolve stable states that are referred to as classes. In this way, we get a first approximation to the conformational landscape of the macromolecule, albeit restricted to these stable states.

However, the limited number of classes that can be extracted from a 3D classification is usually not enough to unveil fully the dynamics of a given macromolecule. The complete characterization of a conformational landscape can only be achieved through the analysis of multiple transient and stable states needed to describe the molecular flexibility in a more accurate manner. The knowledge of these transient and stable states leads to a better description of how structural
changes might affect molecular function or interaction affinity, among other properties of interest.

The formulation we introduce here is oriented towards modelling continuous flexibility (Sorzano et al., 2019), which can be used to characterize the motions undergone by a molecule when exploring different states. We have already addressed this problem in our previous work on continuous heterogeneity using normal mode analysis (NMA) (Sanchez Sorzano et al., 2016). However, this process relied on manual selection of the modes describing the structural changes reflected by two cryo-EM maps, thus making the analysis of molecular flexibility more complex for the user. The new algorithm that we propose in this work tries to address this problem by simplifying the analysis for the user.

In our new methodology, there is no longer a normal modes space where some choices have to be made. Instead, a totally new approach is presented here, based on an expansion on a 3D basis that does not require user intervention at all. We have also improved the analysis of pairwise comparisons by introducing a multidimensional scaling algorithm that automatically combines the outputs from two different metrics. Finally, the new algorithm also allows the analysis of local strains and rotations, as done by us earlier (Sorzano et al., 2016), with the advantage of having all the analyses integrated into a single mathematical framework. We provide a more in-depth comparison with alternative methods in Section 2.1.

The paper makes the following major contributions.

(i) The development of an automatic algorithm to analyse continuous heterogeneity of macromolecules through cryo-EM maps.

(ii) Representation of the strain and rotation components defining a transition between two different conformational states.

(iii) Representation of a series of conformations in a structure mapping and consensus of different mappings defined by different comparison metrics.

(iv) A methodology to compare cryo-EM maps with simulated data.

(v) The application of deformation fields to atomic structures to predict different conformations given by a series of cryo-EM maps.

2. Methods

2.1. Determining structural deformations

In order to detect the movements defining a conformational transition between two states of the same macromolecule, we need to determine the displacements that each region of the molecule will undergo between the two states. The key development in this work is the successful expression of the maps in terms of a mathematical basis on which the displacements are calculated. Although full details are provided in Appendices A and B, here it suffices to say that we use a generalized form of Zernike polynomials to expand functions on a ball (as the macromolecule we are interested in is defined inside a spherical volume). This is not the only possible choice of basis functions [for example, it would have been possible to use the Laguerre polynomials described by Provencher & Voguel (2010) or the prolate spheroidal functions (Greengard & Serkh, 2018)], and we do not expect superiority of any of these possible bases as long as all of them are bases of functions defined within a sphere. Additionally, we find that Zernike polynomials have some appealing mathematical properties especially well suited to our problem. Indeed, these Zernike polynomials allow for the expansion of functions on a sphere which do not vanish at the boundaries (so that the more external parts of the macromolecule can move). Moreover, the basis is closed under rotations. In Appendix B we further explore its properties and its relationship to spherical harmonics.

Considering a pair of electron-density maps representing two conformational states of a macromolecule, it is possible to pose the displacement-finding problem as

\[
\min_{g_l} \int |V_1(r) - V_2[r + g_l(r)]| \, dr,
\]

where \( V_1 \) and \( V_2 \) represent two conformations of a given molecule. Here it is important to note that we are measuring the distance between the target and the distorted volumes in terms of the \( L_1 \) norm. Although it would also have been possible to use the \( L_2 \) norm, we have chosen this definition as it is more robust to outliers (i.e. it is more robust to those cases where the maps do not match completely or have missing regions). The displacement to be applied to the coordinates of \( V_1 \) is defined by the deformation field \( g_l(r) \) parameterized through the expansion in Zernike polynomials \( Z_{l,n,m}(r) \) (see Appendix A),

\[
g_l(r) = \sum_{l=0}^{L} \sum_{n=0}^{N} \sum_{m=-l}^{l} \left( a_{l,n,m}^x, a_{l,n,m}^y, a_{l,n,m}^z \right) Z_{l,n,m}(r),
\]

where \( N \) and \( L \) represent the maximum allowed degrees for the Zernike polynomials and the corresponding spherical harmonics, respectively.

The amount of displacement at every point is controlled by the deformation coefficients \( a_{l,n,m} \). Our objective is to find the deformation coefficients that minimize the goal function in equation (1). This is achieved through a Powell’s conjugate direction method starting from an initial guess of \( a_{l,n,m} = 0 \) for all indices \( l, n, m \) and directions \( x, y, z \) (that is, no deformation). This initialization of the minimization method assumes that the identity/equilibrium solution \( (a_{l,n,m} = 0) \) is close enough to the real solution defining the structural transition represented by the cryo-EM maps. Since in most of the cases this assumption is fulfilled, this initial guess allows the minimization method to find the set of coefficients that appropriately describes the motion between the two maps. However, it is important to note that there are many local minima where the minimization process could be trapped. In this respect, and although in our experience the initialization conditions proposed in this work provide results close enough to the ideal solution, there could be cases in which other ways to initialize the algorithm could be more beneficial in terms of minima search.
The deformation field estimated above can be submitted to the local strain and rotation analysis described by Sorzano et al. (2016). This analysis reveals the nature (stretching, compression or rotation) of the local forces acting on \( V_1 \) to transform it into \( V_2 \) as well as their local intensity.

In our deformation model, it is possible to divide the movements that a molecule may undergo into ‘low’- and ‘high’-frequency movements, depending on how localized these movements are, e.g. a transition from an open to a closed state can be considered a low-frequency movement, while the rotation of a specific \( \alpha \)-helix might be a high-frequency movement. Parameters \( L \) and \( N \) specify the maximum degree of the polynomials used in the description of molecular flexibility. In this way, we may control the maximum frequency of the movements that could be analysed by the basis. Obviously, analysing larger \( L \) and \( N \) will result in a longer computational time, because more \( a_{n,m} \) coefficients will need to be determined and there is a higher risk of overfitting. However, the larger the values given to the parameters \( L \) and \( N \), the higher the frequencies the algorithm will be able to analyse (although in general, global motions dominate the conformational change; Bahar et al., 2010).

Although in many cases analysing global motions is enough to describe in a precise manner the structural changes a macromolecule may undergo, it can be the case that the motions of interest are focused on a very localized area of the molecule. In that case, being able to go to higher degrees on the basis will allow the algorithm to study those motions specifically, without modifying the areas that should remain still. Another possibility is direct restriction of the structural analysis to any specific region in the macromolecule by centring a sphere on that area and selecting an appropriate radius. In this way, it will not be necessary to reach very high degrees in the basis (thus reducing the computational complexity). However, by imposing these kinds of restrictions the algorithm might include artefacts in the surface of the sphere as the molecular regions outside of it will remain untouched. Depending on the molecule and motions to be analysed, the researcher can decide which analysis will be more appropriate for a specific case.

It is important to mention that only in a very few cases did we need to increase the degree of the basis to analyse a localized motion that we were interested in, or have to play with the regularization parameter to get a better approximation of the deformation fields, since the default values were good enough for most of the experiments we have performed so far.

To reduce the possibility of overfitting as much as possible, we regularize the cost function by adding two penalty terms, the first of the regularization penalizes excessive deformation and the second penalizes changes in the mass of \( V_1 \) due to the deformation. Regularization terms \( \lambda_1 \) and \( \lambda_2 \) are usually given low values to prevent large deviations from the ideal solution. Nevertheless, both can be set by the user to any value they consider appropriate for their specific analysis. The guideline for their selection should be that the three terms in the goal function should have values of the same order of magnitude. In our implementation, we report the three contributions helping the user to choose these multipliers.

### 2.2. Relationship to other continuous deformation models

Probably the two most widely continuous deformation models used by the structural biology community in mapping the conformational space of biomolecules (or in analysing cryo-EM images) are principal component analysis (PCA) (Tagare et al., 2015) and normal mode analysis (NMA) (Cui & Bahar, 2006). The three models (PCA, NMA and 3D Zernike) claim to be bases for continuous movements. However, as will be clarified below, they define bases of different mathematical entities.

PCA considers a volume of size \( N^3 \) voxels as a vector in \( \mathbb{R}^{N^3} \). Due to the continuous heterogeneity and the uncertainty in the 3D reconstruction process, the reconstructed map can be considered as the mean of a set of other vectors (maps) whose projections are acquired by the microscope. If we consider the covariance matrix associated with that set of maps (a matrix of size \( N^3 \times N^3 \)), then the principal components form a basis (if the covariance matrix is not degenerate) in which the set of maps can be linearly expressed. The PCA approach approximates the deformed volume by a linear combination of volumes (the principal directions),

\[
V_2(r) \simeq V_1(r) + \sum_n \alpha_n V_n(r),
\]

(4)

where \( V_n \) are the eigenvolumes of the PCA decomposition. The undeformed model is then obtained by subtracting the appropriate amount of each of the eigenvolumes,

\[
V_1(r) \simeq V_{2-1}(r) = V_2(r) - \sum_n \alpha_n V_n(r).
\]

(5)

Due to the low-frequency nature of the PCA principal directions (Sorzano & Carazo, 2021), the undeformed volume is necessarily of low resolution.

In our model, we assume that any deformed volume \( V_2 \) can be undeformed by applying \( g_z \).

\[
V_1(r) \simeq V_{2-1}(r) = V_2[r + g_z(r)].
\]

(6)

Zernike polynomials provide a basis for \( g_z(r) \), not the volumes. Our model revolves around the location of the voxel (which implies a nonlinear relationship between \( V_1 \) and \( V_2 \)), providing an intrinsically better handling of the local characteristics of the map, while in PCA there is a linear model at the level of the volumes themselves (not their internal coordinates).

Our approach has another potential advantage over the PCA model: it can easily be applied to atomic structures fitted...
into $V_1$. For any given atom in the atomic structure at a position $r_i$, that is defined in the same coordinate system as $V_1$, we simply have to move it to the location $r_i + g_i(r_i)$.

In NMA, volumes are approximated by a set of $P$ pseudoatoms with weights $c_p$ and basis function $b(r)$ located at the locations $r_p$ (Jonić & Sorzano, 2016).

$$V_2(r) = \sum_p c_p b(r - r_p).$$

(7)

NMA is based on a second-order Taylor approximation of the energy landscape of the macromolecule, starting with a description of the interactions between the pseudoatoms. This is typically treated using an elastic network model where pseudoatoms within a distance criterion are connected by harmonic springs (Bahar et al., 2010). The associated Hessian is of size $3P \times 3P$ and the normal modes are its eigenvectors (sorted by increasing eigenvalue) and a basis of the R$^{3P}$ space. Let us call $u_k \in \mathbb{R}^{3P}$ the $k$th normal mode, and $u_{k, p} \in \mathbb{R}^3$ the part of the normal mode corresponding to the $p$th pseudoatom. To deform $V_2$ to make it similar to $V_1$ we consider the first $K$ normal modes with different weights $\alpha_k$.

$$V_{2-1}(r) = \sum_{p=1}^P c_p b\left[r - \left(r_p + \sum_k \alpha_k u_{k, p}\right)\right].$$

(8)

Similar to our method, NMA acts by displacing the location of the pseudoatoms (our model acts by displacing the location at which we must interpolate $V_2$). However, an advantage of our new method with respect to NMA is that the NMA deformation is only known at the location of the pseudoatoms, while our new method is fully defined within the sphere containing the macromolecule. In this way, the NMA would be a discretized version of the underlying continuous deformation field, while 3D Zernike polynomials would be an estimate of that continuous field.

Summarizing, each of the methods described so far (PCA, NMA and 3D Zernike polynomials) has a basis in different mathematical entities (vectors in $\mathbb{R}^N$, $\mathbb{R}^{3P}$ or the set of square integrable functions defined within the sphere of a given radius). 3D Zernike polynomials have the advantage that they are defined for every point in the macromolecule (as opposed to NMA) and the undeformed volumes do not lose resolution (as opposed to PCA).

Elastic deformations have also become popular for the alignment of frames within a movie (Abrishami et al., 2015; Tegunov & Cramer, 2019; Zheng et al., 2017). Although they have not been explicitly used to deform volumes, one could envision that they could be easily extended to three dimensions. This would certainly be a possible approach and we earlier used cubic splines for this purpose (Sorzano et al., 2016). However, the basis used in this paper, which is defined exclusively within a sphere, is more appropriate for the task at hand (describing a function whose support is fully contained within that sphere) than for a more generic set of functions that constitute a basis of functions defined within a cube. This ‘greater appropriateness’ translates into requiring fewer coefficients to express the same deformation field to the same level of accuracy.

2.3. Distances between a set of maps

In most practical cases, the number of states that can be reconstructed by cryo-EM SPA is larger than two, which naturally implies the generalization of the case presented above to a number of pairwise operations capturing the different structural relationships among the set of maps under consideration. This information is summarized in a graph known as a structure map (Sanchez Sorzano et al., 2016) or conformational landscape (Zhang et al., 2021), which represents each conformation as a point in conformational space. The closer two points are in the structure map, the more similar they are.

By estimating the Zernike polynomial deformation for all possible pair combinations in a set of $N$ cryo-EM maps, a distance matrix can be computed in which we measure how far two cryo-EM maps are from each other. The deformation field between the two cryo-EM maps $g_i(r)$ provides a mechanism for calculating such a distance. For instance, we may define the distance between two cryo-EM maps $V_1$ and $V_2$ as the sum of the lengths of the deformations at each point,

$$d_i(V_1, V_2) = \int \|g_i(r)\|^2 \, dr.$$ (9)

Besides equation (9), there are additional sensible ways of defining the distance between two cryo-EM maps. One of them consists of measuring the correlation between $V_1$ and $V_2$ once $V_2$ is undeformed to resemble $V_1$.

$$d_i(V_1, V_2) = 1 - \rho \left\{V_2[r + g_i(r)], V_1(r)\right\},$$ (10)

where $\rho$ is Pearson’s correlation coefficient.

By comparing all cryo-EM maps, we would construct a matrix of the distances of all versus all maps.

It is worth mentioning here that, in order to get accurate comparison measurements, it is desirable to have a set of cryo-EM maps with similar characteristics. In particular, it is important to filter the maps in the set so that all their resolutions match the lowest value present in the data set. In this way, the structure mappings and distance matrices will not be affected by resolution changes, leading to a more meaningful projection of the different maps in the low-dimensional space resulting from the application of this method.

2.4. Embedding of conformations using multiple multidimensional scaling

Once we have the above-mentioned distance matrix, we may use multidimensional scaling (MDS) (Härdle & Simar, 2012) to find points in a low-dimensional space of dimension $p$ (typically $p = 2$ or $p = 3$ for ease of representation) such that the distances between points in the low-dimensional space represent in some form the distances between the cryo-EM maps in the full dimensionality space [e.g. equation (9)]. For a detailed description of MDS, see Härdle & Simar (2012). If we have N cryo-EM maps to compare, let us refer to the matrix...
collecting all the points in the low-dimensional space as $X_1$ [$X_1 \in \mathbb{H}(N, p)$, that is, the set of cryo-EM maps of size $N \times p$]. The subscript 1 indicates that we used $d_1$ to perform the low-dimensional mapping.

If instead of equation (9) we use equation (10), then this would give us another MDS representation $X_2$. While the distance $d_1$ concentrates on the amount of deformation required to transform $V_1$ into $V_2$, $d_2$ describes the distance between $V_1$ and $V_2$ after applying the inverse deformation to $V_2$.

We could similarly conceive other strategies to measure the distance between any pair of cryo-EM maps $V_1$ and $V_2$. None of them should necessarily be better than the others, since each one addresses the problem from a different perspective. In this regard, it is impossible to favour any one of the different metrics without a specific task to accomplish. However, it is still sensible to combine the different mappings induced by each one of the distances as a way of producing a single summary of all their information. For the task of producing such a summary, we propose a combination of the embeddings that minimizes the entropy of the result, understanding that the entropy is reduced when more order is found.

At this point and following the aforementioned idea, we may want to combine all those low-dimensional mappings into a single set of points to summarize the relative distances derived from each distance definition. For doing so we have found useful the following procedure that we call multiple multidimensional scaling:

(i) We take one of the mappings as reference, for instance, $X_1$.

(ii) We look for the affine transformation $T_i$ that minimizes the Frobenius norm $\|A\|_F = \left(\sum_{i,j} |a_{ij}|^2\right)^{1/2}$ between each $X_i$ transformed mapping and the reference mapping (since the MDS mappings of different distances, performed in an independent way, normally result in mappings of different scales, central locations, rotations and mirrors).

$$\arg \min_{T_i} \|X_i - T_i(X)\|_F.$$  

For convenience of notation, let us define $T_i(X_i) = X_i$

(iii) The consensus mapping is constructed as the convex combination of all transformed mappings (the determination of the specific $\alpha_i$ coefficients for the combination will be addressed in the following step).

$$X_\alpha = \sum_i \alpha_i T_i(X_i),$$

with the constraints $\alpha_i \geq 0$ and $\sum \alpha_i = 1$ (with these constraints $X_\alpha$ is said to be a convex combination of the input matrices). Note that the $j$th row of the matrix $X_\alpha$ (referred to as $x_{\alpha,j}$) indicates the position of the $j$th cryo-EM map in the low-dimensional space (whose dimension is $p$). For each one of the consensus candidates we associate the probability density function

$$p_\alpha(x) = \sum_{j=1}^N \frac{1}{N} G_\alpha(x - x_{\alpha,j}),$$

where $G_\alpha$ is a $p$-multivariate spherical Gaussian whose covariance matrix is $\sigma^2 I$ [in our experiments, we chose $\sigma = \max|\text{range}(X_1), \ldots, \text{range}(X_n)|/20$, where $\text{range}(X_1)$ is the difference between the maximum and minimum values of any of the components of the mapped vectors].

(iv) Since the best combination of coefficients $\alpha_i$ is not known beforehand, each possible convex combination has to be analysed. The criterion followed was to look for the convex combination that minimized the Shannon entropy of the probability density function defined above,

$$\arg \min_{\alpha} \left\{-\int p_\alpha(x) \log[p_\alpha(x)] dx\right\}.$$  

The rationale is that we are looking for the convex combination that brings maximum order to the low-dimensional mapping.

We observe that the procedure described above normally finds a good balance between the properties of the different low-dimensional mappings, resulting in well structured summaries.

3. Results

This algorithm has been implemented in Xmipp (de la Rosa-Trevín et al., 2013) and it is available through Scipion (de la Rosa-Trevín et al., 2016) under the protocols named volume deform - Zernike3D and struct map - Zernike3D.

We performed some tests with a pair of maps to compare these two implementations to analyse the performance improvement. The maps used for the tests had dimensions of 250 in $X$, $Y$ and $Z$, leading to averaged execution times of 1 h and 20 min (CPU) and 39.5 s (GPU). The tests were performed with an Intel i7-9750H and a Nvidia 2060 with Cuda 10.1, respectively.

3.1. Experiment 1: cryo-EM maps of the human mitochondrial ribosome

We first tested our approach using a small data set covering a range of conformational states of a human mitochondrial ribosome (Amunts et al., 2015), as previously described by Sanchez Sorzano et al. (2016). To check whether the structure map suggested two independent (pre-translocation and post-translocation) states following different conformational transitions as found in our previous study (Sanchez Sorzano et al., 2016), we applied the methodology described above with $N = 3$ and $L = 2$ (the maximum allowed degrees for the Zernike polynomials and the spherical harmonics, respectively). As expected, the structure map indeed suggests two independent (pre-translocation and post-translocation) states following their own conformational transitions, grouped as red and blue dots in Fig. 1 (the black line segments joining the dots are just provided to enhance visualization). We thus conclude that the new approach is capable of reproducing the results of previous supervised methods that perform similar analyses and accurately groups the seven
cryo-EM structures [indicated by their EMDB (Electron Microscopy Data Bank, https://www.ebi.ac.uk/emdb/) identification numbers] into two groups of conformers, each representative of a different functional state.

Additionally, the new approach also allows for the local decomposition of the deformation field into local strains and local rotations, as was done by Sorzano et al. (2016). The representation of these two components is shown in Fig. 2 for one of the pairs of ribosomes (EMDB entries 1720 and 1723). In addition, Video 1 in the supporting information shows the conformational changes described by these two maps. For this video, we coloured the ribosomes using the rotation component represented in Fig. 2 to simplify their comparison. According to this analysis, the rotations appear to be distributed through the whole structure of the ribosome, although the larger rotations (shown in red) are mostly found in the small subunit. Similarly, the strains are mainly localized in the small subunit and appear to be less distributed. This reveals that the basis is capable of deforming in a localized fashion, leading to a better description and identification of the different movements that define the transition between the two conformations. It is also possible to see that we are obtaining results comparable with those found by Sorzano et al. (2016), with the advantage of having all these analyses unified in the same framework, which implies an overall simplification leading to more complete studies.

3.2. Experiment 2: trajectory recovery of the CCT complex

Our next experiment is aimed at characterizing the ability of the method to recover the sequence of events present in a set of conformations defining a certain trajectory in conformational space. Such conformations can be created computationally by taking advantage of biophysical methods such as molecular dynamics simulations (MD) (Adcock & McCammon, 2006) and normal mode analysis (NMA) (Bahar et al., 2010), simulating the movements defining a transition between two conformations. In this case, we used a trajectory from a recent study (Zhang et al., 2021b), which was generated using a purely NMA-based approach called the adaptive anisotropic network model (adaptive ANM; Yang et al., 2010) implemented in ProDy (Zhang et al., 2021a). This gave us 30 different models along an open–closed transition of the mammalian chaperonin CCT complex between two atomic models derived from a previous cryo-EM study (Cong et al., 2012), taken from the Protein Data Bank (PDB) (wwPDB Consortium, 2019), as described by Zhang et al. (2021b). The starting structure with one ring open and one ring closed (PDB entry 4a0w) (Cong et al., 2012) corresponded to an ATP-bound state and the target structure with both rings in an intermediate conformation (PDB entry 4a13) (Cong et al., 2012) corresponded to the ADP-bound state, allowing us to explore the conformational changes triggered by ATP hydrolysis. In the adaptive ANM method, all steps are based on coarse-grained normal modes calculated using the anisotropic network model (ANM) (Attilgan et al., 2001; Doruker et al., 2000; Eyal et al., 2006), providing coordinate changes for Cα atoms only. At each step, normal modes were selected that had the highest directional overlap (correlation cosine) with the deformation vector between the current conformation and the target structure up until the sum of the squared overlaps exceeded a threshold of 0.4. The contribution of each mode to the deformation was chosen so as to take 20% of the maximum provided by the unnormalized dot products (a scaling factor of 0.2) so as to avoid unphysical deformations while maintaining efficiency. The normal modes were recalculated until the root-mean-square deviation (r.m.s.d.) from the target structure fell below 1 Å, resulting in a total of 30 steps. Each step recruited a larger number of modes and had a smaller total size as the required deformation became less...
cooperative and more local (see Video 2). We focus our discussion on the ring that goes from open to intermediate-closed for simplicity.

We then transformed these atomic structures into Coulomb potential maps using the electron atomic scattering factors (EASFs) as described in previous work (Sorzano et al., 2015). Fig. 3 shows the structure maps recovered after applying our methodology. We can see that the sequential order of the 30 intermediate conformers along the trajectory was successfully recovered by our approach. The direction, however, is arbitrary and in this case the start of the trajectory was numbered as conformer 30 and the end as conformer 1.

With this example, we additionally illustrate the distinct MDS mappings obtained when the distances \( d_1 \) [amount of deformation, Fig. 3 (top)] and \( d_2 \) [similarity after deformation, Fig. 3 (middle)] are used. Although the trajectory was successfully recovered by both distances, the correlation distance \( d_2 \) was slightly more accurate in this case. The reason is that most of the changes between the structures at the end of the transition (labelled 1 to 13 by the algorithm) are high-frequency movements (i.e. movements of loops or small \( \alpha \)-helices and \( \beta \)-sheets) that cannot be fully captured by the Zernike 3D basis with \( N = 3 \) and \( L = 2 \) (although larger \( N \) and \( L \) would allow one to express these small-detail movements, they would also increase the computational cost). Fortunately, the consensus mapping [Fig. 3 (bottom)] is able to identify the existence of high-frequency movements and gives more weight to the \( d_2 \) mapping (correlation distance) automatically, resulting in an almost exact recovery of the volume sequence along the trajectory.

At least in this case we can conclude that \( d_1 \) is very good for describing the low-frequency movements (e.g. C23–C30), while \( d_2 \) is very good for characterizing the high-frequency differences (ca C1–C13), and both perform well in the intermediate-frequency regime. Depending on whether our set of input maps are related by large or small movements, one distance or the other will be better suited to capturing the overall set of relationships. The consensus mapping will thus analyse both mappings and automatically determine the optimal weight that results in a low-dimensional mapping that can be readily interpreted.

3.3. Experiment 3: comparison of atomic models and cryo-EM maps from the rabbit ryanodine receptor RyR1

In the following example, we explored the possibility of matching (pseudo/simulated) cryo-EM maps derived from atomic models with experimental electron microscopy maps in the same low-dimensional space. For this purpose we selected five experimental cryo-EM maps deposited for the ryanodine receptor 1 (RyR1) from rabbit (EMDB entries 8379, 8385, 8390, 8395 and 8373) and their respective atomic models in the PDB (PDB entries 5tam, 5tau, 5taz, 5tb4 and 5t9n).

First, we converted the atomic models into density maps using EASFs, as described in the previous section. Then, to make the cryo-EM maps and atomic models comparable, we also filtered all volumes in the analysis to a common resolution.
(specifically, to the lowest of the reported resolutions of the cryo-EM maps). Note that without applying this low-pass filter the minimization process of equation (1) might not reach a meaningful minimum. Finally, we applied the algorithm presented in this work to this combined data set.

Our results, shown in Fig. 4, report the main difficulties that appear when mixing simulated and experimental cryo-EM maps. While the structure map based on $d_1$ (the distance based on the amount of deformation) illustrates that many pairs are correctly placed together, the structure map based on $d_2$ (the distance based on the similarity after undeforming) discriminates between maps derived from atomic models and maps coming from cryo-EM experiments. However, the point of this example was to intermix maps from different origins, so discrimination by origin was to be minimized, requiring a further adjustment to our approach. To tackle this problem, we extended our methodology by analysing separately the sub-blocks of the distance matrix including only atomic or only cryo-EM maps [see Fig. 5 (top)]. We thus performed the MDS of each one of the sub-blocks independently, obtaining the low-dimensional mappings $X_{AA}$ and $X_{CC}$ (the subscript indicates whether it corresponds to atomic/computational or cryo-EM/experimental maps). These two low-dimensional mappings were the input into the consensus procedure described in Section 2.4. Focusing on the consensus, we can see that the information provided by the two mappings $X_{AA}$ and $X_{CC}$ is combined into two different trajectories corresponding to each dimension in the distance matrices (simulated and experimental cryo-EM maps) that show a similar distance relationship among their points, illustrating that both trajectories correspond to the same states of RyR1. Therefore, the counterpart of each other, and their relative distances/positions, are retained [Fig. 5 (bottom)].

3.4. Experiment 4: application of the deformation field to atomic models of the CCT complex

We described our deformation field $g_L$ as a function that deforms $V_1$ to let it become similar to $V_2$, that is, as we have done in previous cases, acting only on two cryo-EM maps. However, since the deformation field is defined in the coordinate system of $V_1$, it can also be applied to atomic models defined in the same coordinate system and not only to maps. In this way, we can also deform an atomic model defined for $V_1$ and use it as starting point for a model of $V_2$. Obviously, since the new atomic model defined in the coordinate system of $V_2$ has been constructed purely based on geometrical considerations, all the stereochemical constraints have to be further imposed.

An example of an atomic model deformed following the previous procedure is presented in Fig. 6. The example was taken from the same data set as used in Experiment 2, which shows an open–closed transition of the CCT complex. The figure illustrates how the deformation applied to the atomic model of the open conformation results in an approximation to the closed conformation. Naturally, we can now compare this deformed model representing the closed conformation with the one obtained directly from the experimental map of the closed conformation. The r.m.s.d. (computed with ChimeraX) between these two models was 5.29 Å, certainly high, but substantially reduced compared with that between the open–closed models without applying any deformation, which was 7.90 Å. This r.m.s.d. reduction suggests that the deformation applied is appropriately reproducing a conformational change in the right direction, from open towards the closed state.

However, the overall scores obtained for the two deformed structures still show a high value, as many stereochemical
features are not taken into account when computing the deformations. In order to improve the geometry of the deformed structures, we applied a real-space refinement [executing Phenix software (Liebschner et al., 2019) with the default parameters] to the predicted structures using their respective electron-density maps. After this refinement, the r.m.s.d. value measured before decreased further to 4.52 Å. As a conclusion, the combination of deformation and refinement of atomic structures enables us to achieve predictions of different structure conformations on the path between two end points, suitable for performing further studies, though there is clearly room for improvement. For example, refining in between smaller deformations could be of benefit, e.g. in hybrid simulations methods where local refinement/simulation complements global deformations (Krieger et al., 2020).

4. Conclusions
The development of automatic algorithms to study continuous flexibility presented in this work results in simplified yet precise procedures, avoiding the need for user interference with the software and increasing the reproducibility of the results. It is also a significant step forward with respect to approaches like PCA and NMA of cryo-EM maps, avoiding the need to select components or modes and producing localized analysis.

The way this new approach works is by defining a new 3D basis where all deformation occurs. It is conceptually similar to the Fourier transform. The movements defining a transition between two different conformational states are decomposed into different components (that can be regarded as low-, medium- and high-frequency movements). Those components will depend on the degree of the basis used in the calculations. The displacements needed along each different component to minimize the distances between two electron-density maps are stored in a series of deformation coefficients $a_{l,n,m}$, which can be further analysed to obtain the local strains and rotations undergone by the macromolecules during conformational transitions. The new approach thus unifies two of our previous developments (NMA and strain/rotation component extraction) for the analysis of continuous heterogeneity.
is useful for building distance matrices. These distance
can be used afterwards to recover structure mappings
that show the structural relationships existing among the
diverse conformational states. Different definitions of the
distance measures may focus on different aspects of the
comparison. For this reason, we have devised a new procedure
to combine several low-dimensional mappings into a single
consensus mapping based on a minimum entropy criterion
that tends to produce well ordered low-dimensional mappings
and outperforms the results obtained by individual distance
metrics.

The possibility of converting atomic models back to
electron densities opens the possibility of a combined analysis on
maps and models in the same conformational space. An
illustrative example has been provided in Experiment 4, where
cryo-EM maps, together with their respective structural
models between two end points, have been represented in the
same space as a set of experimental cryo-EM maps.

In the future, it may be interesting to explore alternative
basises for the deformation field and the distance between
volumes (like the Wasserstein distance).

## APPENDIX A

### 3D real-valued generalized Zernike polynomials

In this section we discuss the functions that we use as basis
functions of the deformations in the unit ball $B$. We use the
generalized Zernike polynomials defined on the 3D ball; in
Appendix B1, we briefly review the relation of this basis to the
better-known 2D form of Zernike polynomials.

In general, the expansion of any real valued function $g(r) \in
L^2(B)$ in this basis is defined by the formula

$$g(r) = \sum_{l=0}^{\infty} \sum_{n=0}^{\infty} \sum_{m=-l}^{l} a_{l,n,m} Z_{l,n,m}(r),$$

where $a_{l,n,m}$ are real-valued coefficients, and $Z_{l,n,m}(r)$ are the
3D real-valued (normalized) generalized Zernike polynomials
defined by the formula

$$Z_{l,n,m}(r) = R_{l,n}(r) y^l_{n}(\theta, \phi),$$

where $r$ is the radial component of the 3D coordinate $r$, $\theta$ and
$\phi$ are its polar and azimuthal angles, respectively, in spherical
coordinates, $n$ and $l$ are non-negative integers, and $m$ is an
integer such that $-l \leq m \leq l$. We refer to $l$ as the spherical
frequency. $y^l_{n}$ is the real-valued spherical harmonic defined by
the formula

$$y^l_{n}(\theta, \phi) = (-1)^m \frac{2l+1}{4\pi} \frac{(l-|m|)!}{(l+|m|)!} \left( \frac{d^{l+|m|}}{dx^{l+|m|}} (x^2 - 1)^l \right),$$

where $P^m_l$ are the associated Legendre polynomials defined by
the formula

$$P^m_l(x) = \begin{cases} 
\begin{align*}
(1-x^2)^{m/2} d^{l+m}_{l+m}(x^2 - 1)^l & \quad \text{if } m > 0, \\
2^{l+|m|} \sin(m \phi) & \quad \text{if } m < 0.
\end{align*}
\end{cases}$$

The real and imaginary parts of the complex-valued
spherical harmonics are available in standard textbooks such as
that by Abramowitz & Stegun (1966). For completeness,
Table 1 shows these spherical harmonics in Cartesian coordinates.
Before defining the normalized generalized radial
Zernike polynomials denoted by $R_{l,n}$ above, we define the
(unnormalized) generalized radial Zernike polynomials $R_{l,n}^m$ as
follows:

### Table 1

| Degree $(l)$ | Order $(m)$ | $Z_l^m(r)$ |
|-------------|-------------|------------|
| 0           | 0           | $\frac{1}{2} \left( \frac{1}{\pi} \right)^{1/2}$ |
| 1           | 0           | $\left( \frac{3}{4\pi} \right)^{1/2} \frac{y}{r}$ |
| 2           | 0           | $\frac{1}{2} \left( \frac{15}{2\pi} \right)^{1/2} \frac{xy}{r^2}$ |
| 3           | 0           | $\frac{1}{4} \left( \frac{3\pi}{2\pi} \right)^{1/2} \frac{xyz}{r^3}$ |

Table 1 shows these spherical harmonics in Cartesian coordinates,
research papers

Table 2
Generalized and normalized radial Zernike polynomials.

| $R_{0,0}(r)$ | $R_{1,0}(r) = 3^{1/2}$ | $R_{2,0}(r) = (7^{1/2})^{2}$ | $R_{3,0}(r) = 3^{3}$ |
| $R_{1,1}(r) = (7^{1/2})^{2}$ | $R_{2,1}(r) = (11)^{1/2} \left( \frac{5r^2}{2} - \frac{3}{2} \right)$ | $R_{3,1}(r) = 3^{1/2} \left( \frac{15r^2}{2} - \frac{13}{2} \right)$ |
| $R_{2,1}(r) = (9^{1/2})^{2}$ | $R_{3,1}(r) = (13)^{1/2} \left( \frac{15r^2}{2} - \frac{13}{2} \right)$ |
| $R_{3,1}(r) = (11)^{1/2} \left( \frac{13r^2}{2} - \frac{11}{2} \right)$ | $R_{4,1}(r) = (15)^{1/2} \left( \frac{255r^2}{8} - \frac{255r^2}{4} + \frac{195r^2}{8} \right)$ |
| $R_{4,1}(r) = (21)^{1/2} \left( \frac{76475r^2}{16} - \frac{76475r^2}{4} + \frac{108675r^2}{16} \right)$ |

\[ R_{0,0}(r) = \left( -1 \right)^{\alpha} x^{\alpha} P_{n}^{(\alpha+\beta/2)}(1 - x^2) \]  
\[ R_{1,0}(r) = 2^{1/2} \left[ 2n + l + \frac{p}{2} + 1 \right]^{1/2} R_{n}^{\alpha}(x) \]  

where $P_{n}^{(\alpha, \beta)}$ are the Jacobi polynomials,

\[ P_{n}^{(\alpha, \beta)}(x) = \left( -1 \right)^{n} \frac{1}{2^{n} n!} \left( 1 - x \right)^{-\alpha} \left( 1 + x \right)^{-\beta} \times \frac{d^n}{dx^n} \left[ (1 - x)^{\alpha} (1 + x)^{\beta} (1 - x^2)^{n} \right] \]  

The parameter $p$ is associated with the choice of inner product and the dimensionality of the balls; in the case of a 3D ball, the natural choice of $p = 1$, which yields the basis in (16) which is orthonormal in the natural inner product on $L_2(B)$. We give in Table 2 the explicit list of radial functions $R_{n,\alpha}$ that we use. We recognize that our choice of basis functions for the expansion is certainly not the only possible choice. We used the Zernike polynomials (in the generalized form presented here) to obtain an expansion of functions in a ball which do not vanish at the boundaries. Our use of Zernike polynomials also yields a basis that is closed under rotations. The graphical representation of some components of the basis is shown in Fig. 7.

The definitions and properties of the standard associated Legendre polynomials and Jacobi polynomials are available, \textit{inter alia}, in the book by Abramowitz & Stegun (1966).

Finally, while the radial polynomials are orthogonal (with the appropriate norm), they are not orthonormal. This is easily corrected by replacing the radial polynomials with the normalized generalized radial Zernike polynomials, denoted by $\tilde{R}_{n,\alpha}$.
APPENDIX B
Properties and relationships of this basis

B1. Properties of the polynomials involved – Zernike polynomials

We note that slightly different definitions and normalization are used in different sources; the most commonly used form of Zernike polynomials is associated with 2D functions on the unit disc, whereas we are interested in 3D functions on the unit ball. The better-known traditional radial Zernike polynomials, denoted here by \( \bar{R}_m(x) \), are a special case of the generalized radial Zernike polynomials \( R_{n,l}^m(x) \),

\[
\bar{R}_m(x) = R_{m,0}^m(x) = R_{m,0}^m(\sqrt{x^2 + y^2})
\]

with \( \bar{R}_m(x) = 0 \) if \( l - m \) is odd or if \( m > l \). The definition of the 3D real-valued Zernike polynomial [equation (16)] is analogous to the definition of the traditional 2D Zernike polynomials \( Z_m \) on the unit disc,

\[
Z_m^*(r, \phi) = R_m^m(r) \left\{ \begin{array}{ll} \cos(m\phi) & \text{if } m \geq 0, \\ \sin(|m|\phi) & \text{if } m < 0. \end{array} \right.
\]

Unfortunately, the common notation for the 2D and 3D cases can be misleading: the parameter \( m \) here plays the role of the parameter \( l \) in the analogous 3D case; the parameter \( m \) in the 3D case is related to the existence of both sine and cosine for each \( m \) here, but does not otherwise have an immediate counterpart here.

Some properties of these generalized Zernike polynomials, including the higher-dimensional cases, are discussed in further detail by Slepian (1964), Serkh (2015), Greengard & Serkh (2018) and Lederman (2017). The radial Zernike polynomials in equation (19) are orthogonal with respect to the inner product,

\[
\langle f(x), g(x) \rangle = \int_0^1 x^{p+1} f(x) g(x) \, dx,
\]

so that \( \langle R_{n_1,l_1}^m(x), R_{n_2,l_2}^m(x) \rangle = 0 \) if \( n_1 \neq n_2 \). Note that they are not necessarily orthogonal for different \( l \), that is, \( \langle R_{n,l_1}^m(x), R_{n,l_2}^m(x) \rangle \) is not 0, in general. It follows that the Zernike polynomials \( Z_{l,n,m} \) (16) are orthogonal (across all different combinations of \( n, l \) and \( m \)) on the natural inner product on the unit ball.

B2. A remark on numerical evaluation

As is the case with many orthogonal polynomials, the direct computation using the explicit sum of monomials is generally unstable and not recommended in numerical computation. However, in this work, since we truncate the polynomials at low \( n \), the explicit form has been found experimentally to be sufficiently stable. For additional details on computation see Lederman (2017).

B3. Closure under rotations

We recall that we use basis functions defined in equation (15), which are composed of a radial component and an angular component. Furthermore, we truncate the expansion in equation (2) such that if \( Z_{l,n,m}(\mathbf{r}) \) is in the expansion, then \( Z_{l,n,m'}(\mathbf{r}) \) is also in the expansion for any \(-l \leq m, m' \leq m\). As is well known, the rotation of the frame of reference of spherical harmonics of a given spatial frequency \( l \) is a unitary operation and closed rotations. It follows that the linear combination of \( \sum_{n=0}^{N} \sum_{l=0}^{L} \sum_{m=-l}^{l} \alpha_{l,n,m} Z_{l,n,m}(\mathbf{r}) \) is closed under rotations. In other words, regardless of the frame of axis we choose for our spherical harmonics, we can represent the same functions using our choice of basis.

We recall that the deformation field \( g_{l}(\mathbf{r}) \) is a 3D vector defined in equation (2) at any point \( r \). If we rotate the axes \( x, y \) and \( z \) we must rotate the vector \( g_{l}(\mathbf{r}) \) to obtain a vector in the new coordinate system. As is well known,

\[
A g_{l}(A^{-1} \mathbf{r}) = \sum_{n=0}^{N} \sum_{l=0}^{L} \sum_{m=-l}^{l} A \alpha_{l,n,m} Z_{l,n,m}(A^{-1} \mathbf{r}),
\]

where \( A \) is the appropriate unitary rotation matrix.

It follows that the basis we have chosen is closed under rotations; regardless of the orientation of the frame of reference we choose (but depending on the position of the centre), we can represent the same fields. Furthermore, the transformation between frames of reference is unitary. We note that, since the volume is defined on the grid, the problem definition is not entirely closed under rotations, although the definition of the field is closed under rotations.
APPENDIX C
Complex-valued Zernike 3D basis

We may extend our basis function to a complex-valued Zernike 3D basis. This second basis uses spherical harmonics, which are well known basis functions for functions defined over the surface of a unit sphere. For doing so, we should define \( Z'_{l,n,m}(\mathbf{r}) \) [see equation (16)],

\[
Z'_{l,n,m}(\mathbf{r}) = R'_{l}(r) Y^{m}_{l}(\theta, \phi),
\]

where \( Y^{m}_{l}(\theta, \phi) \) are the standard spherical harmonics,

\[
Y^{m}_{l}(\theta, \phi) = \left[ \frac{2l + 1}{4\pi} \frac{(l - m)!}{(l + m)!} \right]^{1/2} P^{m}_{l}(\cos \theta) \exp(i m \phi).
\]

It is well known that the spherical harmonics are a complete orthonormal basis on the surface of the unit sphere such that

\[
\int_{\Omega} Y_{l_{1}m_{1}}^{*}(\mathbf{r}) Y_{l_{2}m_{2}}(\mathbf{r}) \, d\Omega = \begin{cases} 1 & \text{if } l_{1} = l_{2}, m_{1} = m_{2}, \\ 0 & \text{otherwise}, \end{cases}
\]

where \( \Omega \) is the surface of the unit sphere.

In this new basis the expansion would be expressed as

\[
g(\mathbf{r}) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \hat{\alpha}_{l,m} Z'_{l,n,m}(\mathbf{r}).
\]

The relationship between the expansion that uses the real-valued basis functions and the expansion that uses the complex-valued basis functions is

\[
\hat{\alpha}_{l,n,m} = \begin{cases} (-1)^{m+1} \frac{1}{\sqrt{2l + 1}} (\beta_{l,n,m} - \beta_{l,n,-m}) & \text{if } m < 0, \\ \beta_{l,n,0} & \text{if } m = 0, \\ (-1)^{m} \frac{1}{\sqrt{2l + 1}} (\beta_{l,n,m} + \beta_{l,n,-m}) & \text{if } m > 0,
\end{cases}
\]

and for the basis functions

\[
Z'_{l,n,m}(\mathbf{r}) = \begin{cases} \frac{1}{\sqrt{2l + 1}} [Z_{l,n,m}(\mathbf{r}) - (-1)^{m} Z_{l,n,-m}(\mathbf{r})] & \text{if } m < 0, \\ \frac{1}{\sqrt{2l + 1}} Z_{l,n,0}(\mathbf{r}) & \text{if } m = 0, \\ \frac{1}{\sqrt{2l + 1}} [Z_{l,n,m}(\mathbf{r}) + (-1)^{m} Z_{l,n,-m}(\mathbf{r})] & \text{if } m > 0.
\end{cases}
\]

Using the fact that the radial Zernike polynomials are a complete radial orthonormal basis, and the fact that spherical harmonics are a complete orthonormal basis on the surface of a sphere, one can show that the generalized Zernike polynomials are a complete orthonormal basis of function on the unit ball with respect to the natural \( L^{2} \) norm on the unit ball.

Suppose that \( \mathbf{r} \) is a 3D coordinate of a point in real space and \( \mathbf{R} \) is a 3D frequency. Then the Fourier transform of the complex-valued basis function would be

\[
(\mathcal{F}Z'_{l,n,m}) = \int_{\mathbb{R}^3} \exp(-i(\mathbf{r}, \mathbf{R})) Z'_{l,n,m}(\mathbf{r}) \, d\mathbf{r} = Y^{m}_{l}(\mathbf{R}) \frac{1}{R^{3/2}} \frac{1}{R} J_{2m+1}(\sqrt{\mathbf{R}^{2}+1})(\mathbf{R}),
\]

where \( \langle \mathbf{r}, \mathbf{R} \rangle \) is the usual inner product in \( \mathbb{R}^3 \), \( R \) is the modulus of \( \mathbf{R} \) and \( J_{\alpha}(x) \) is the Bessel function of the first kind and order \( \alpha \).

Funding information
The project that gave rise to these results received the support of a fellowship from ‘la Caixa’ Foundation (No. LCF/BQ/ DI18/11660021). This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement (No. 713673) and the European Regional Development Fund Project ‘CERIT Scientific Cloud’ (No. CZ.02.1.01/0.0/0/16_013/0001802). This work was also partially supported by the National Institutes of Health (P41GM103712 to IB), by a MoSs Seed Software Fellowship (to JK) and by the NIH/ NIGMS (No. 1R01GM136780-01 to RRL). Funding is also acknowledged from the Ministry of Science and Innovation through grants: SEV 2017-0712, PID2019-104757RB-I00/AEI18/10.1309/50110011033; the ‘Comunidad Autonoma de Madrid’ through grant: S2017/BMD-3817; and the European Union (EU) and Horizon 2020 through grant: HighResCells (ERC - 2018 - SyG, Proposal: 810057) and iNEXT-Discovery (Proposal: 871037).

References
Abramowitz, M. & Stegun, I. A. (1966). Handbook of Mathematical Functions. New York: National Bureau of Standards.
Abridge, V., Vargas, J., Li, X., Cheng, Y., Marabini, R., Sorzano, C. S. & Carazo, J. M. (2015). J. Struct. Biol. 189, 163–176.
Adcock, S. A. & McCammon, J. A. (2006). Chem. Rev. 106, 1589–1615.
Amunts, A., Brown, A., Toots, J., Scheres, S. H. W. & Ramakrishnan, V. (2015). Science, 348, 95–98.
Atilgan, A. R., Durell, S. R., Jernigan, R. L., Demirel, M. C., Keskin, O. & Babar, I. (2001). Biophys. J. 80, 505–515.
Bahar, I., Lezon, T. R., Yang, L. W. & Eyal, E. (2010). Annu. Rev. Biophys. 39, 23–42.
Carroni, M. & Saibil, H. R. (2016). Methods, 95, 78–85.
Cong, Y., Schröder, G. F., Meyer, A. S., Jakana, J., Ma, B., Dougherty, M. T., Schmid, M. F., Reissmann, S., Levitt, M., Ludtke, S. L., Frydman, J. & Chiu, W. (2012). EMBO J. 31, 720–730.
Cui, Q. & Babar, I. (2006). Normal Mode Analysis: Theory and Applications to Biological and Chemical Systems. Boca Raton: CRC press.
Doruker, P., Atilgan, A. R. & Babar, I. (2000). Proteins, 40, 512–524.
Eyal, E., Yang, L. W. & Babar, I. (2006). Bioinformatics, 22, 2619–2627.
Greengard, P. & Serkh, K. (2018). arXiv:1811.02733.
Härde, W. K. & Simar, L. (2012). Multidimensional Scaling, pp. 397–412. Heidelberg: Springer.
Jonić, S. & Sanchez Sorzano, C. O. (2016). IEEE J. Sel. Top. Signal. Process. 10, 161–173.
Krieger, J. M., Doruker, P., Scott, A. L., Perahia, D. & Babar, I. (2020). Curr. Opin. Struct. Biol. 64, 34–41.
Lederman, R. R. (2017). arXiv:1710.02874.
Liebschner, D., Alonie, P. V., Baker, M. L., Bunkóczi, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L. W., Jain, S., McCoy, A. J., Moriarty, N. W., Oefner, R. D., Poon, B. K., Prisant, M. G., Read, R. J., Richardson, J. S., Sammito, M. D., Sobolev, O. V., Stockwell, D. H., Terwilliger, T. C., Urzhumtsev, A. G., Videau, L. L., Williams, C. J. & Adams, P. D. (2019). Acta Cryst. 65, 861–877.
Provencher, S. W. & Vogel, R. H. (1988). Ultramicroscopy, 25, 209–221.
Rosa-Trevín, J. M. de la, Otón, J., Marabini, R., Zaldívar, A., Vargas, J., Carazo, J. M. & Sorzano, C. O. S. (2013). J. Struct. Biol. 184, 321–328.
Rosa-Trevín, J. M. de la, Quintana, A., del Cano, L., Zaldívar, A., Fochi, I., Gutiérrez, J., Gómez-Blanco, J., Burguet-Castell, J., Cuenca-Alba, J., Abrishami, V., Vargas, J., Otón, J., Sharov, G., Vilas, J. L., Navas, J., Conesa, P., Kazemi, M., Marabini, R., Sorzano, C. O. S. & Carazo, J. M. (2016). *J. Struct. Biol.* **195**, 93–99.

Sanchez Sorzano, C. O., Alvarez-Cabrera, A. L., Kazemi, M., Carazo, J. M. & Jonić, S. (2016). *Biophys. J.* **110**, 1753–1765.

Schur, F. K. M. (2019). *Curr. Opin. Struct. Biol.* **58**, 1–9.

Slepian, D. (1964). *Bell Labs Technical J.*, **43**, 3009–3057.

Sorzano, C. O. S. & Carazo, J. M. (2021). *Acta Cryst.* **D77**, 835–839.

Sorzano, C. O. S., Jiménez, A., Mota, J., Vilas, J. L., Maluenda, D., Martínez, M., Ramírez-Aportela, E., Majtner, T., Segura, J., Sánchez-García, R., Rancel, Y., del Caño, L., Conesa, P., Melero, R., Jonić, S., Vargas, J., Cazals, F., Freyberg, Z., Krieger, J., Bahar, I., Marabini, R. & Carazo, J. M. (2019). *Acta Cryst.* **F75**, 99–32.

Sorzano, C. O. S., Martín-Ramos, A., Prieto, F., Melero, R., Martín- Benito, J., Jonić, S., Navas-Calvente, J., Vargas, J., Otón, J., Abrishami, V., de la Rosa-Trevín, J. M., Gómez-Blanco, J., Vilas, J. L., Marabini, R. & Carazo, J. M. (2016). *J. Struct. Biol.* **195**, 123–128.

Sorzano, C. O. S., Vargas, J., Otón, J., Abrishami, V., de la Rosa-Trevín, J. M., Riego, S., Fernández-Alderete, A., Martínez-Rey, C., Marabini, R. & Carazo, J. M. (2015). *AIMS Biophys.* **2**, 8–20.

Tagare, H. D., Kucukelbir, A., Sigworth, F. J., Wang, H. & Rao, M. (2015). *J. Struct. Biol.* **191**, 245–262.

Tegunov, D. & Cramer, P. (2019). *Nat. Methods*, **16**, 1146–1152.

wwPDB Consortium (2019). *Nucleic Acids Res.* **47**, D520–D528.

Yang, Z., Májek, P. & Bahar, I. (2010). *PLoS Comput. Biol.* **5**, e1000360.

Zhang, S., Krieger, J. M., Zhang, Y., Kaya, C., Kaynak, B., Mikulska-Ruminska, K., Doruker, P., Li, H. & Bahar, I. (2021a). *Bioinformatics*, doi: 10.1093/bioinformatics/btab187.

Zhang, Y., Krieger, J., Mikulska-Ruminska, K., Kaynak, B., Sorzano, C. O. S., Carazo, J. M., Xing, J. & Bahar, I. (2021b). *Prog. Biophys. Mol. Biol.* **160**, 104–120.

Zheng, S. Q., Palovcak, E., Armache, J. P., Verba, K. A., Cheng, Y. & Agard, D. A. (2017). *Nat. Methods*, **14**, 331–332.