nature communications

Article

https://doi.org/10.1038/s41467-025-59135-0

Real-space heterogeneous reconstruction, refinement, and disentanglement of CryoEM conformational states with HetSIREN

Received: 26 September 2024

Accepted: 11 April 2025

Published online: 22 April 2025

Check for updates

David Herreros[1], Carlos Perez Mata[1,2], Chari Noddings[3], Deli Irene[4], James Krieger[1], David A. Agard[5,6], Ming-Daw Tsai[4], Carlos Oscar Sanchez Sorzano[1,7] & Jose Maria Carazo[1,7]

Single-particle analysis by Cryo-electron microscopy (CryoEM) provides direct access to the conformations of macromolecules. Traditional methods assume discrete conformations, while newer algorithms estimate conformational landscapes representing the different structural states a biomolecule explores. This work presents HetSIREN, a deep learning-based method that can fully reconstruct or refine a CryoEM volume in real space based on the structural information summarized in a conformational latent space. HetSIREN is defined as an accurate space-based method that allows spatially focused analysis and the introduction of sinusoidal hypernetworks with proven high analytics capacities. Continuing with innovations, HetSIREN can also refine the images' pose while conditioning the network with additional constraints to yield cleaner high-quality volumes, as well as addressing one of the most confusing issues in heterogeneity analysis, as it is the fact that structural heterogeneity estimations are entangled with pose estimation (and to a lesser extent with CTF estimation) thanks to its decoupling architecture.

Cryo-electron microscopy (CryoEM) Single Particle Analysis (SPA)[1] ability to capture individual images of biological samples brings to light the challenging capacity to identify several conformational and/or compositional states from the acquired image dataset. Classically, compositional heterogeneity and conformational heterogeneity/flexibility have been addressed through rounds of 3D classification[2] under the assumption that macromolecules adopt a discrete set of states. Discrete classification has been applied successfully and is at the heart of the so-called "Resolution Revolution"[3]. However, the discrete approach introduces a series of limitations that arise from the assumptions on which it is based. Removing this discretization constraint is methodologically a very challenging task. However, the pay-offs are clear in obtaining richer conformational landscapes than is currently done, providing improved algorithmic stability and objectivity, removing assumptions not supported by the data, and streamlining the analysis process without trial error tests and decisions on the quality and number of classes.

Algorithms for identifying continuous heterogeneity from particle images were first introduced in 2014[4,5]. More recently, CryoDRGN[6] introduced the concept of heterogeneous reconstruction, applying advanced neural network techniques to address the approximation of the conformational continuum through the decodification of per-particle structural states. Similarly, other approaches have been proposed to tackle the heterogeneous reconstruction problem, such as heterogeneous reconstruction with Gaussian Mixtures[7] or heterogeneous reconstruction with a priori information on conformational latent space[8], among others.

[1]Centro Nacional de Biotecnologia-CSIC, C/ Darwin, 3, Cantoblanco Madrid, Spain. [2]PKF Attest innCome, Orense 81, Madrid, Spain. [3]Altos Labs, Redwood City, CA, USA. [4]Institute of Biological Chemistry, Academia Sinica, Taipei, Taiwan. [5]Department of Biochemistry & Biophysics, University of California, San Francisco, CA, USA. [6]Chan Zuckerberg Imaging Institute, Redwood City, CA, USA. [7]These authors jointly supervised this work: C.O.S. Sorzano and J.M. Carazo. e-mail: dherreros@cnb.csic.es

In addition to the reconstruction of heterogeneous states, other methods have focused on estimating molecular motions/flexibility using deformation fields. Some methods rely on a neural network to decode the deformation field directly from a latent space representation[9,10], while other approaches have proposed to expand the field on a different basis and then use a reduced set of parameters to estimate the complete field[11,12].

This work proposes a different approximation to the heterogeneous reconstruction problem, moving the reconstruction process ultimately to real space. This approach makes it possible to fine-tune the neural network architecture to improve the quality of decoded volumes, reduce noise overfitting, and perform focused/exclusion heterogeneous reconstructions. We achieve these goals by combining three critical innovations. The first one is introducing (in CryoEM) SIREN activation functions in the network architecture. Indeed, compared to other popular approaches to approximate functions with a decoder architecture, such as ReLU with positional encoding, SIREN activations have been shown to preserve much better the quality and high-frequency features of the original signals fed to the network[13]. The second factor is the effective decoupling of pose and CTF effects from the estimation of conformational landscapes, a key issue considering how intertwined these processes are, as noted in[14]. This second goal is achieved by introducing constraints in latent space relating multiple projections of the same structure from different directions. Finally, we introduce a set of regularizers, including L_1 and Total Variation minimization, that helps to obtain high-resolution maps from individual coordinates of conformational space.

The practical result of HetSIREN's capabilities is that conformational landscapes are now substantially better at presenting structurally relevant heterogeneity information, and their exploration can be done at high resolution. This fact can significantly affect many biological systems when rounds of 3D classification lead to reduced data sets and low resolution.

Our major contributions are:

- We propose a encoder decoupling architecture to disentangle the pose and CTF estimation explicitly from the structural information in the structural landscapes, directly tackling one of the primary sources of error heterogeneity algorithms face. This approach generates more understandable, accurate, and interpretable landscapes than standard network architectures.
- Application of meta-sinusoidal layers and hypernetworks to decode high-resolution 3D conformational states with enhanced local resolution and structural features compared to standard reconstruction methods.
- Efficient reconstruction of complete 3D volumes in real space, including the possibility to add structural priors to improve the representation of the electron density maps. To that end, real space constraints are added to explicitly mitigate the noise and negative values in the decoded volumes. In addition, constraints in real space to enhance the continuity and sharpness of the protein signal against the noise are included in the network.
- Possibility to include reconstruction masks to focus on or exclude unwanted structures during the heterogeneous reconstruction process.
- We propose a robust multiresolution training scheme to simplify training on high-resolution data where noise becomes a solid limiting factor.
- We apply these methods to identify multiple conformations of the SARS-CoV-2 Spike protein from single datasets and demonstrate their variation with temperature.

Results

This section analyzes a simulated dataset, followed by a classical public data set commonly used when presenting heterogeneous reconstruction methods, ending with the presentation of collaborative work on challenging specimens.

All the datasets were analyzed with Scipion 3.8.0 software package. Inside Scipion, CryoSPARC 4.5.1, Relion 4.0, and Xmipp 3.24.12.0 packages were also used to process the data.

Simulated adenylate kinase landscape and landscape disentanglement

To accurately evade the effect of decoupling pose and CTF from the estimated HetSIREN landscape, we propose a simple and conceptual experiment based on the simulation of an open-to-closed trajectory of the adenylate kinase protein (PDB entry 4AKE) using Normal Mode Analysis with HEMNMA[15]. The simulated trajectory is recovered from the excitation of two modes, leading to a ground truth landscape with a straight-line shape. The trajectory was then sampled to generate a set of 500 projections with uniformly distributed poses and variable CTF information.

The simulated projections were imported into Scipion[16] to train two different HetSIREN networks: a network without pose and CTF decoupling and a network with a pose and CTF decoupling encoder. The resulting landscapes are provided in Fig. 1a, b.

As seen in Fig. 1a, the standard autoencoder architecture does not recover the ground truth landscape. However, it effectively captures the simulated motion along the first principal component of the landscape. In general, this is the type of effect expected to arise on standard heterogeneity algorithms due to unwanted factors that compromise the quality of the latent spaces and significantly limit the interpretability of the landscape.

Figure 1b shows the landscape obtained from the decoupling architecture of HetSIREN. In this case, the new landscape successfully captures a structure resembling the ground truth landscape, mainly arising from a more prominent structural component. Therefore, the combination of the decoupling encoder and the decoder in HetSIREN dramatically aids in the interpretation and understanding of the molecular transition captured in a given dataset.

Conformational landscape of EMPIAR 10028 dataset

To allow the direct evaluation of HetSIREN compared to other popular heterogeneity tools, we have performed a heterogeneity analysis of EMPIAR-10028[17], which has become one of the de facto standard datasets in the field to address the performance of heterogeneity methods.

EMPIAR-10028 entry corresponds to a CryoEM acquisition of the *P. falciparum* 80S ribosome bound to emetine. The raw data from the database was further processed with Scipion[16], resulting in about 50,000 particles. The workflow within Scipion included several consensus and cleaning steps, trying to reduce unwanted images to a minimal representation and increasing the stability of the angular assignment, shifts, and Contrast Transfer Function (CTF) estimations, which are inputs to the algorithm. It should be noted that most heterogeneity algorithms (but not HetSIREN) treat these inputs as fixed variables, increasing the need to work only with well-curated datasets to avoid misleading conclusions during the heterogeneity analysis.

The particles resulting from the previous analysis were fed to the HetSIREN network during the training phase, followed by the analysis of the latent space encoded from the experimental images. To this end, the latent space mentioned here was explored with the help of interactive tools integrated within the Scipion Flexibility Hub[18]. A landscape visualization and exploration example is provided in Fig. 2. Two landscapes are presented in Fig. 2 in the form of 3D UMAPs[19] obtained from the original 10-dimensional space encoded by the network; the one on the left is without pose and CTF decoupling, while the one on the right implements decoupling. Although a ground truth does not exist for this data set, the landscape after pose and CTF decoupling can be segmented much more easily, a fact that we interpret as an

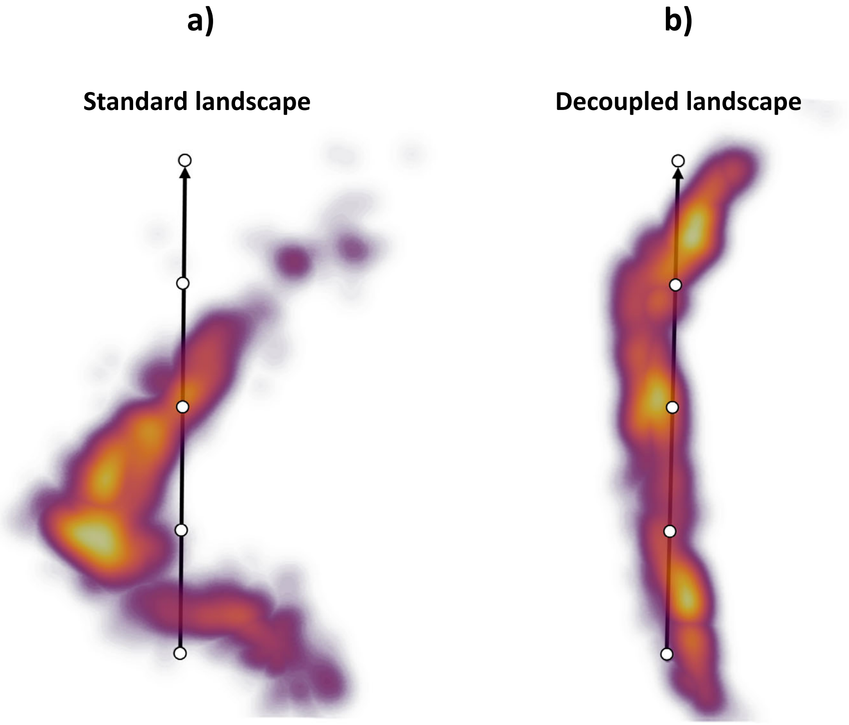


Fig. 1 | Comparison of HetSIREN standard and pose-CTF decoupled landscapes for the adenylate kinase protein's open-to-close simulated transition. Ideally, the landscape should approximate the ground-truth trajectory defined as a straight line arising from the excitation of two protein modes. (**a**) shows the landscape obtained with a standard architecture, which suffers from a strong deviation from the ground truth due to the pose and CTF coupling. (**b**) shows the pose and CTF decoupled landscape obtained with HetSIREN. Decoupling the pose and CTF information makes the structural information more prominent, allowing the latent space to approximate the ground-truth conformational landscape well, which should be just a straight line.

enhancement of structural information over positional and CTF "noise". Furthermore, the decoupled landscape was then explored (Fig. 2b) by visualizing a set of maps corresponding to the cluster representatives (centroids) obtained from the 10-dimensional space using the K-Means algorithm, which was later decoded with the network to recover the electron density maps at those points. An initial landscape inspection through the maps shows a non-negligible degree of compositional heterogeneity affecting the ribosome, mainly focused on the 40S ribosomal subunit. Furthermore, it was possible to identify a low-populated state (as shown on Map 12) characterized by a complete lack of the 40S subunit, which is usually not detected due to its low representation (close to 600 particles - around 1.3% of the data). As shown in Fig. 2b, Map 12, HetSIREN successfully identified this evasive state and decoded a map with a resolution similar to those obtained from other more populated landscape regions.

In addition to the compositional heterogeneity analysis of the sample, HetSIREN also allows the identification of the continuous changes captured by the images and the interplay of continuous and compositional changes resulting from their combined influence on the structural characteristics of the complex. In Fig. 3a, b, we provide an example of four decoded maps showing continuous conformational states with and without an extra compositional component, respectively; in each case, the change of conformation is shown by superimposing two maps in two different colors (blue and yellow). The structural change presented in Fig. 3a corresponds to a rotation from left to right of the 40S subunit, one of the main structural changes captured in this dataset. In contrast, the proposed structures in Fig. 3b represent a compositional variation of one of the RNAs found in the ribosomal structure and a significant motion of another RNA generally undetected due to its low resolvability. These examples demonstrate the capacity of HetSIREN to analyze various types of heterogeneity in the same landscape and to understand the interplay of the different structural modifications that a biomolecule may undergo.

One of the characteristics included in HetSIREN is the possibility of focusing the conformational landscape on a region of interest rather than considering the whole complex during the training phase. This functionality allows us to identify the relevant motions of those regions more easily, which is especially important for small areas since they may have weak relevance in the overall conformational landscape or exclude certain regions from the variability analysis, such as membranes or nanodiscs. In the case currently analyzed, we decided to perform a focused heterogeneity analysis of the ribosomal L1 stalk region, a substantially small area compared to the ribosome but that exhibits a high degree of flexibility. Due to size differences, the contribution of L1 to the general landscape is not predominant, limiting the interpretability of the motions that the L1 stalk undergoes.

To focus on the L1 stalk, we trained a new HetSIREN network with the same particles presented before but providing a spherical mask that covers only the L1 stalk region. HetSIREN can only modify the L1 stalk region thanks to the previous mask, effectively focusing the landscape on this region and excluding the contributions of everything outside the mask. The L1 stalk landscape approximated by HetSIREN and reduced with UMAP is presented in Supplementary fig. 1a. The landscape shows two main motion directions, which can be isolated with PCA as shown in Supplementary fig. 1b. The two main motions correspond to a non-negligible lateral and vertical translation of the L1 stalk, which is more easily identified thanks to the focusing capabilities of HetSIREN.

In addition, in Fig. 4a, b, we provide a comparison of the local resolution computed with DeepRes[20] between a map decoded by HetSIREN and the primary reconstruction obtained from the initial image processing workflow performed in Scipion (that is, the map obtained from the 3D refinement carried out with the complete

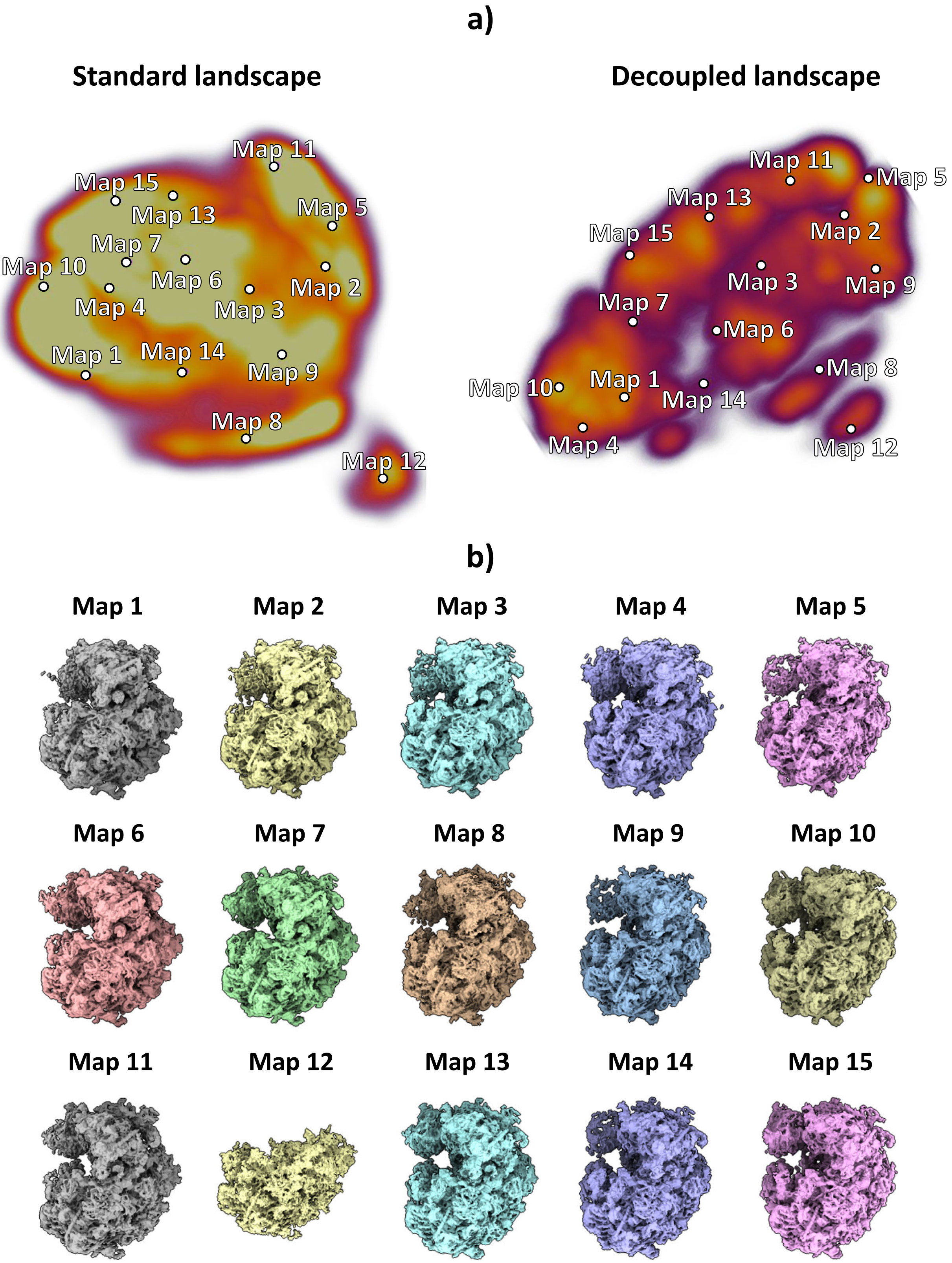


Fig. 2 | HetSIREN landscape and exploration for Empiar 10028 dataset. (**a**) shows the UMAP[19] representation of the landscape obtained from the latent space encoded by HetSIREN from the particle images after training. The landscape without pose and CTF decoupling is presented on the left, while decoupling is implemented on the right landscape. Each dot in the landscape corresponds to the centroid of the cluster representative obtained from a KMeans clustering of the decoupled HetSIREN latent space. (**b**) shows the decoded HetSIREN maps obtained from the decoupled latent space coefficients assigned to every representative, as shown in (**a**) (right). The maps provide a sensible exploration of the different conformational states identified by HetSIREN and a comparison of the structural features learned by the network.

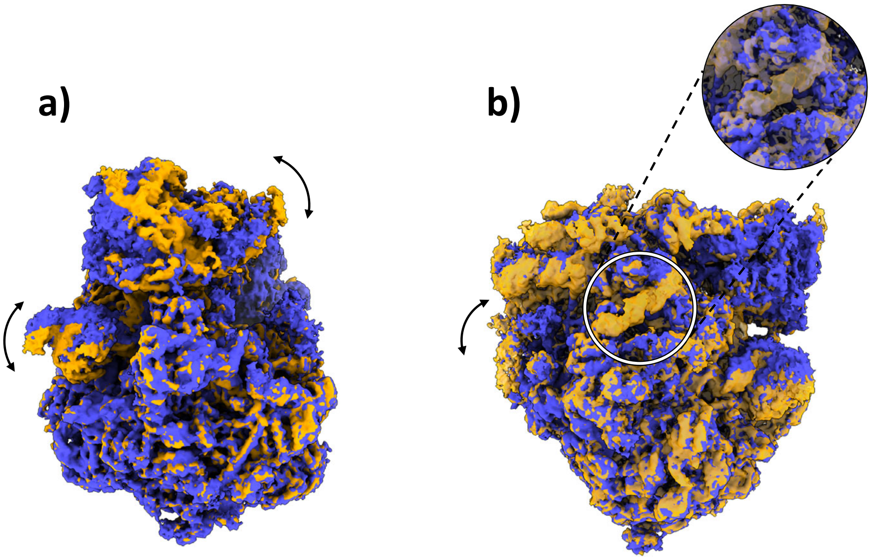


Fig. 3 | Example of some conformational changes captured in the decoupled HetSIREN landscape presented in Fig. 2a. (**a**) shows the main continuous conformational change captured in the dataset, corresponding to a coordinated rotation of the 40S subunit and the R1 stalk of the ribosome. (**b**) shows a compositional variation in one of the ribosomal RNAs (better shown in the magnified image), as well as a large continuous motion of the RNA to the left of the panel, usually not detected due to its low resolvability.

dataset of 50k particles posteriorly used to train HetSIREN). The comparison shows that the HetSIREN decoder does not sacrifice resolution in the decoding process, significantly increasing the local resolution of the map.

Conformational landscape of the GR:Hsp90:FKBP51 complex

The GR:Hsp90:FKBP51 complex represents a critical molecular assembly regulating the glucocorticoid receptor (GR), a key player in numerous physiological processes, including stress response, metabolism, and immune function. This complex involves the chaperone protein Hsp90 and the immunophilin FKBP51, which together influence the conformation and activity of GR[21]. Unlike its counterpart, FKBP52, which enhances GR activity, FKBP51 acts antagonistically, inhibiting GR's ability to bind ligands and translocate to the nucleus. This inhibition is crucial for maintaining the receptor's homeostasis and responsiveness to hormonal signals. The GR:Hsp90:FKBP51 complex's ability to modulate GR activity has significant implications, as GR dysregulation can lead to various health issues, including immune dysfunction and increased susceptibility to stress-related disorders. Understanding this complex provides insight into potential therapeutic targets for diseases influenced by glucocorticoid signaling.

Given the importance of the GR:Hsp90:FKBP51 complex, we analyzed the intrinsic conformational variability of the dataset presented in ref. 21 with HetSIREN. The dataset included 106884 particles with CTF and angular information already estimated as required by the method.

The images were used to train the network to generate an 8D conformational latent space, posteriorly reduced by PCA[22] to a 3D space for representation purposes. The resulting PCA landscape is provided in Fig. 5a. To gain more insight into the motions detected by HetSIREN, we sampled the leading principal component to generate a set of five different conformational states. The corresponding latent space coordinates were transformed into density maps using the HetSIREN decoder, as presented in Fig. 5b. The figure highlights one of the extreme conformations along the sampled PC 1 axis in a black contour to simplify the understanding of the conformational change (black contour corresponding to Map 5). The results show a significant movement of the GR and FKBP51 regions with respect to the HsP90 protein, resulting in a rotational translation of these two components.

In addition to estimated motions, the quality of decoded HetSIREN volumes was further analyzed and compared with the deposited map from[21] (EMD-29069 [https://www.ebi.ac.uk/emdb/EMD-29069]). Figure 6a shows a direct comparison of the published (left) and HetSIREN (right) maps colored according to their local resolution estimated with DeepRes[20]. The local resolution analysis shows a slight improvement in the resolution of HetSIREN, mainly present in the GR:FKBP51 region. The previous results highlight the ability of HetSIREN to learn and decode high-quality maps, translating into an improved interpretation of the map even in highly dynamic areas.

Temperature dependence on the conformational landscape of the SARS-CoV-2 Spike protein

The human respiratory coronavirus SARS-CoV-2 is responsible for causing COVID-19, an acute and often severe respiratory illness characterized by intense inflammatory responses and lung damage[23]. Although the virion contains several structural and non-structural proteins, much attention has been directed towards the S (Spike) protein. This glycoprotein forms a trimeric Spike that interacts with the host cell receptor angiotensin-converting enzyme 2 (ACE2) through a mechanism involving the receptor binding domain (RBD) in an equilibrium between RBD opening and closing[24]. Throughout the pandemic, changes in the conformational equilibrium of the RBD have been directly related to the evolution of the virus and the emergence of new variants[25]. These mutations influence the ability of the virus to bind to ACE2 and enter host cells, affecting transmission dynamics and disease severity.

Previous studies have elucidated the impact of the temperature of storage or incubation on the overall integrity and denaturation of the Spike protein[26,27]. It has been reported that the temperature of the protein, equilibrated before and during vitrification, can have a pronounced effect on protein conformation[28], an observation that we wanted to precisely quantify through continuous heterogeneity analysis. In this study, we worked with the Spike protein's beta variant (B.1.351), initially identified in South Africa in the summer of 2020. The cryo-EM datasets for the same sample vitrified at 4 °C and 37 °C according to[29] were acquired separately. Following the acquisition of these datasets and the initial steps of image processing in Scipion, we employed a continuous flexibility analysis using HetSIREN to further

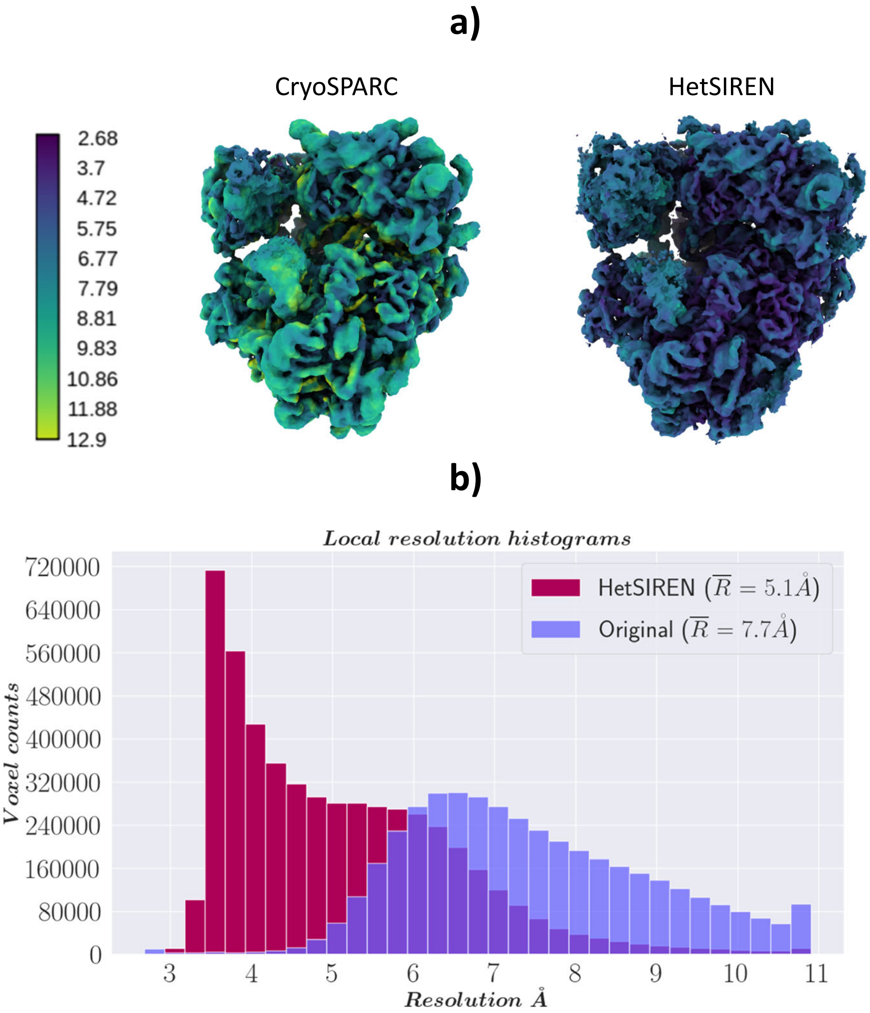


Fig. 4 | Resolution analysis of HetSIREN map compared against the map refined from the EMPIAR 10028 dataset with CryoSparc[38]. (**a**) shows first the CryoSparc refinement obtained from the 50k particle dataset processed inside Scipion (left) followed by the HetSIREN decoded map (right), both colored according to their local resolution estimated with DeepRes[20]. The comparison shows a significant improvement in the local resolution of the map decoded by HetSIREN. **b** Quantitatively compares the estimated local resolutions based on local resolutions histograms. Similarly to (**a**), the local resolution of HetSIREN shows a strong displacement of the voxel resolutions to the high-resolution domain, translating into an improvement of 2.7 Å in the mean resolution. Source data are provided as a Source Data file.

address the extreme conformational heterogeneity inherent in this sample and characterize subtle conformational changes between the two temperatures.

As previously described, images were used to train the HetSIREN's network to generate an 8D conformational latent space, which was subsequently reduced to a 3D space for representation and analysis purposes using UMAP[19]. To explore the entire conformational landscape and detect all potential conformations, we obtained 20 decoded HetSIREN volumes from a K-Means clustering of the original HetSIREN latent space for each sample.

At 4 °C, our analysis revealed the presence of 1 Up and 2 Up conformations (Fig. 7 and see below), which aligns with previous observations using conventional discrete classification protocols[27]. However, when the sample was at 37 °C, we observed a distinct conformational landscape, predominantly characterized by the 3 Down conformation, which was the only conformation identified by discrete methods (Fig. 8). Nevertheless, our studies benefited from an improved quantification capacity, thanks to the use of advanced analysis tools provided by HetSIREN. This allowed for the additional detection of a reduced contribution from the 1 Up state (Figs. 8, 9b). In addition to the noticeable differences observed in the RBD, we also identified additional dynamic patterns in the N-terminal domain (NTD), particularly pronounced in the sample at 37°C. As further evidence of the influence of temperature on the Spike protein, we observed that the 1 Up conformation at 37 °C exhibited a less open range of the opening configuration compared to its counterpart at 4 °C (Fig. 9). To quantitatively assess differences in the opening range of the RBD, we generated 29 atomic models that included the 3 Down and 1 Up conformations of the Spike protein at both 37 °C and 4 °C (Supplementary fig. 7). Using the ProDy software tools implemented in Scipion[30], we created an atomic structure ensemble that was subjected to PCA[22]. The RBD opening motions were accurately described by the first principal component (PC1) (Supplementary fig. 7a). For this analysis, we examined several loops in the RBD that exhibit high mobility, as indicated by the PCA results (Supplementary fig. 7b). To quantify the opening range of the RBD, we used centroids of two fixed regions within the core of the Spike protein (S2 domain) as reference points, specifically residues Val991 and Pro1140 from all three chains. We measured the angle between these constant regions and the RBD (Supplementary fig. 7c). The loop spanning residues Thr500-Gly502 provided the most precise description of the differences, showing a clear transition from the 3 Down (6.6° ± 0.3°) to the 1 Up conformation at 37 °C (21.6° ± 0.9°), followed by the 1 Up conformation at 4 °C (24.9° ± 0.7°) (Supplementary fig. 7c inset and Supplementary fig. 7d).

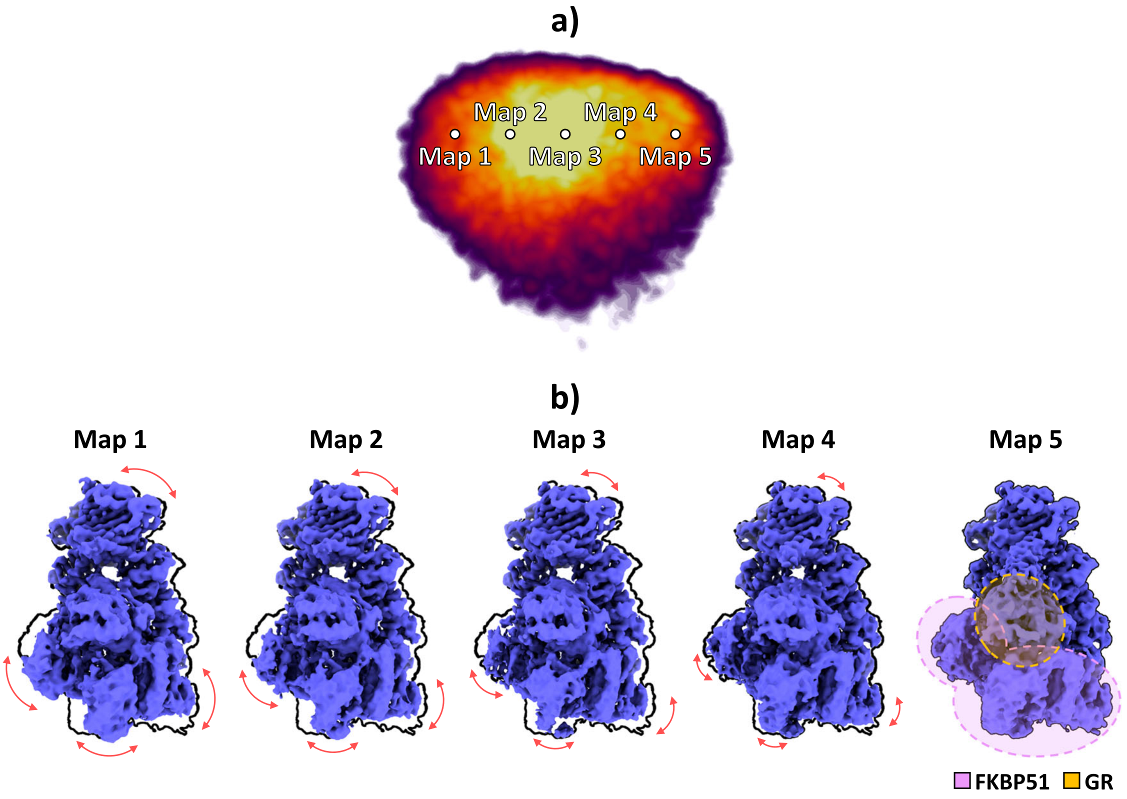


Fig. 5 | Conformational landscape analysis of the main motions detected by HetSIREN on the GR:Hsp90:FKBP51 complex. (**a**) shows the reduced PCA landscape obtained from the original 8D HetSIREN latent space learned by the network. Each dot in the landscape represents an even sampling along the main PC component. (**b**) displays the volumes decoded by the HetSIREN network from the sampled points shown in (**a**). The black outline shows the structural state represented by Map 5, which is provided to simplify the interpretation of the conformational change. The detected motion significantly translates the GR and FKBP51 components against the HsP90 protein.

Furthermore, we directly compared the local resolution computed by DeepRes[20] between the HetSIREN maps and the maps obtained using discrete image processing procedures, including initial model generation and subsequent refinement. This comparison revealed an increase in local resolution on the HetSIREN map, with particular emphasis on the RBD and NTD, which are typically the most mobile regions and exhibit lower resolvability (Fig. 10). To quantitatively assess the improved local resolution, we performed automatic RBD modeling with ModelAngelo[31], which showed an increase in the total number of modeled residues of 3.7 to 10.5% depending on the map (Supplementary Table 2).

By analyzing the volumes presented in Fig. 10, it is also possible to assess the reliability and precision of the method compared to discrete approaches. To that end, we compared the reconstruction obtained by CryoSparc using the 23k closest particles to the selected state as input against the map decoded by the network at the centroid of this subset in latent space. This comparison is a reasonable way to validate the method's accuracy in identifying the structural states captured by the experimental particles, avoiding hallucinations. As seen in Fig. 10, both HetSIREN and the reconstructed map agree with high confidence about the conformational state captured in that specific region of the structural landscape. In addition, the comparison of the different local resolutions estimated by DeepRes also shows the improved volume representation capabilities of the network, which is capable of decoding high-resolution states directly from a single point in the latent space, helping in those common cases in which several rounds of discrete classification may end up with reduced data sets. This capability improved the average local resolution of around 0.7 Å.

Our findings suggest that when the Spike protein is maintained at 4 °C, it tends to adopt more open configurations, predisposing it to subsequent denaturation, consistent with previous studies[26,27] at different conditions. Conversely, at 37 °C, the range of molecular motions tends toward a more closed and less accessible conformation. Even when the Spike protein is in a 1 Up state at this temperature, its opening range is markedly reduced compared to that at 4 °C. This phenomenon may directly influence the ability of the virus to evade the immune system while maintaining its capacity to initiate successful infections by modulating the equilibrium with ACE2[32]. Given the importance of the Spike protein in vaccine development, which typically involves recombinant expression of attenuated versions stored at low temperatures, our study underscores the importance of further structural analyses with methods such as HetSIREN. A deeper understanding of the conformational dynamics under specific conditions could have profound implications for the design of new vaccine formulations[33].

Discussion

Continuous heterogeneity is a significant breakthrough in the CryoEM field, as shown by its increasing popularity and successful applications to better understand macromolecular conformational rearrangements through experimental CryoEM data[4,5,7–12].

In this regard, we have introduced a deep learning-based heterogeneous reconstruction and refinement method called HetSIREN. HetSIREN addresses the conformational variability problem in real space by encoding particle images into a latent space based on their specific structural state, followed by a decoder capable of translating the latent space into high-resolution volumes.

HetSIREN presents several critical innovations that set it apart to all current methods. In a nutshell, by working entirely in real space, HetSIREN has been able to use and even modify altogether (in the field

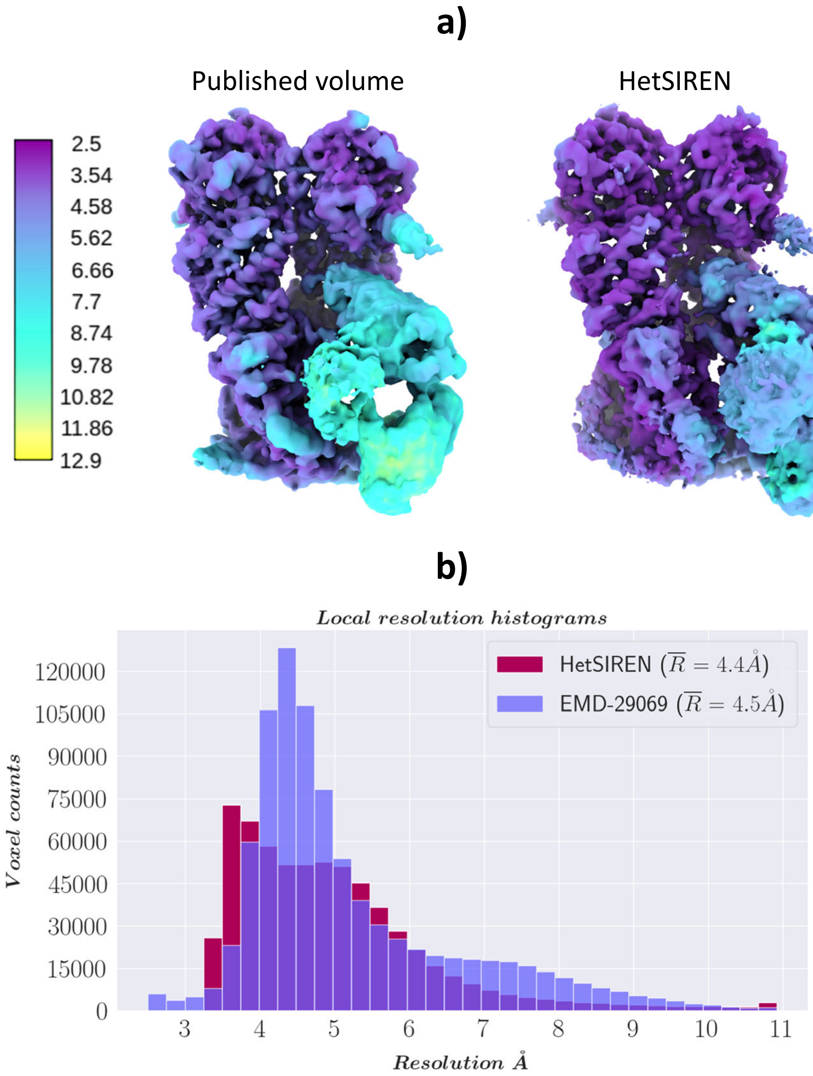


Fig. 6 | Resolution analysis of HetSIREN decoded maps compared with the deposited map from[21]. (**a**) shows on the left the deposited map (EMD-29069) and the HetSIREN map on the right, both colored by their local resolution estimated with DeepRes[20]. The comparison shows an improvement in the local resolution of the map decoded by HetSIREN, mainly occurring in the flexible region composed of GR and FKBP51. (**b**) quantitatively compares the estimated local resolutions based on local resolution histograms. Similarly to (**a**), the local resolution of HetSIREN shows a displacement of the voxel resolutions to the high-resolution domain. Source data are provided as a Source Data file.

of CryoEM) meta-sinusoidal activation fields with many enhanced analytics capabilities to current approaches[13,34]. Furthermore, we implemented a "disentanglement" procedure concerning pose and CTF so that we introduce a constraint in latent space that makes it focus on structural differences and not into pose-induced or CTF-induced differences (indeed, the way structural changes translate into changes at the image projection level is very different depending on the particle projection direction, as indicated in ref. 14). HetSIREN also introduces a range of regularizers, such as L_1 and Total Variation minimization.

The real-space decoded map is further analyzed to enhance the biomolecule signal and minimize the presence of common artifacts and errors in CryoEM, such as noise or negative values in the reconstruction, while preserving the structural features in the map. Furthermore, HetSIREN allows one to customize the reconstruction region with a mask, which can be applied to exclude unwanted structures from the decoded volumes (such as membranes or nanodiscs) or to focus the heterogeneity analysis on a specific region of space.

In addition to estimating structural states from initially supplied projection geometry information, HetSIREN can refine the initial particle pose and in-plane shift according to each image's specific conformation. To this end, it produces two extra latent spaces, one to analyze particle configurations and the other to refine the pose and in-plane shift of the input particles into the network. In this way, the alignment matrix refinement is also considered during the training phase, helping to generate better CryoEM maps from the decoder.

In conclusion, HetSIREN adds a new approach to the growing heterogeneity analysis family of methods. It does so by introducing unique characteristics that set it apart, including the application of SIREN-based hypernetworks to improve the quality of the decoded maps, the ability to disentangle the pose and CTF information in conformational landscapes, the possibility to customize the analysis process with user-defined reconstruction masks as well as the inclusion of explicit regularization terms to enhance the structural features of the decoded maps while reducing noise and other artifacts. The net result of all these innovations is two-fold: First, HetSIREN conformational landscapes are much more structure-focused than with current approaches in the field, and second, HetSIREN is capable of obtaining maps from individual points of the landscape with improved resolution compared to what is currently achievable in the field.

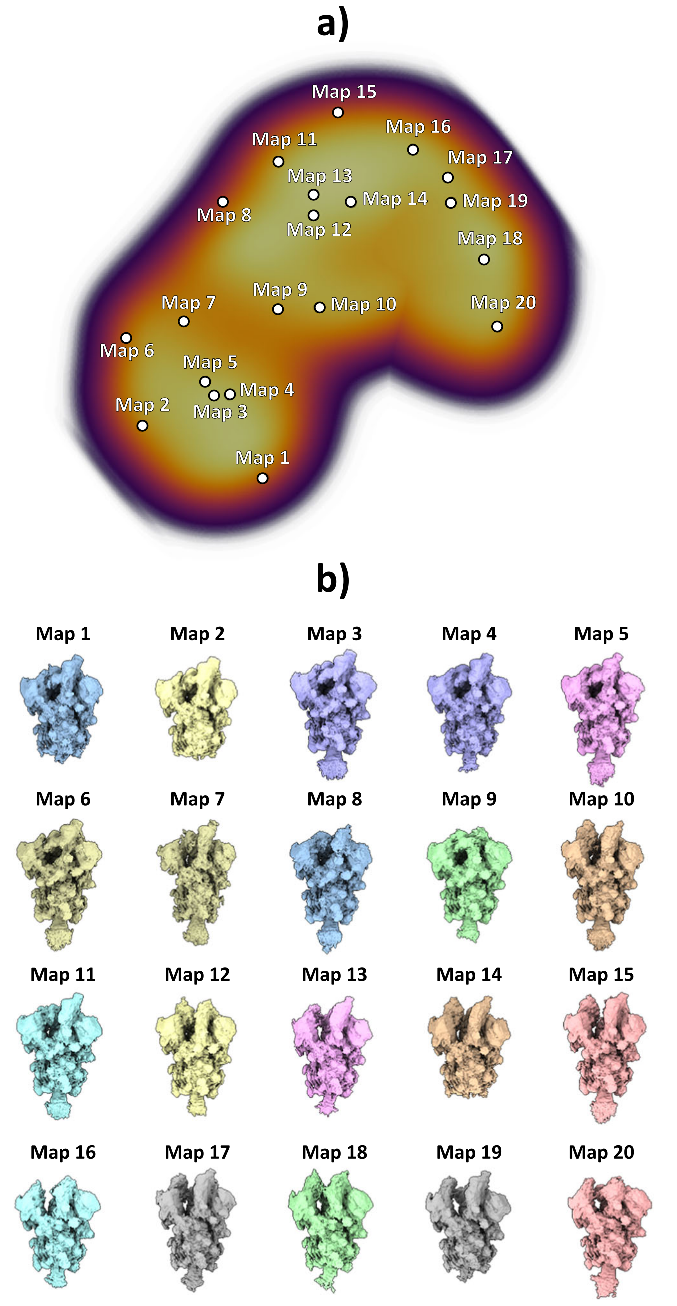


Fig. 7 | Conformational landscape analysis of the main motions detected by HetSIREN on the SARS-CoV-2 Spike protein at 4°C. Panel a) shows the UMAP[19] representation of the landscape obtained from the original 8D latent space encoded by HetSIREN from the particle images after training. Each dot in the landscape corresponds to the position of the cluster representative obtained from a KMeans clustering of the original HetSIREN latent space. Panel b) shows side views of the decoded HetSIREN volumes obtained from the latent space coefficients assigned to every representative shown in Panel a). The maps provide a sensible exploration of the different states identified by HetSIREN, including 1 Up and 2 Up conformations.

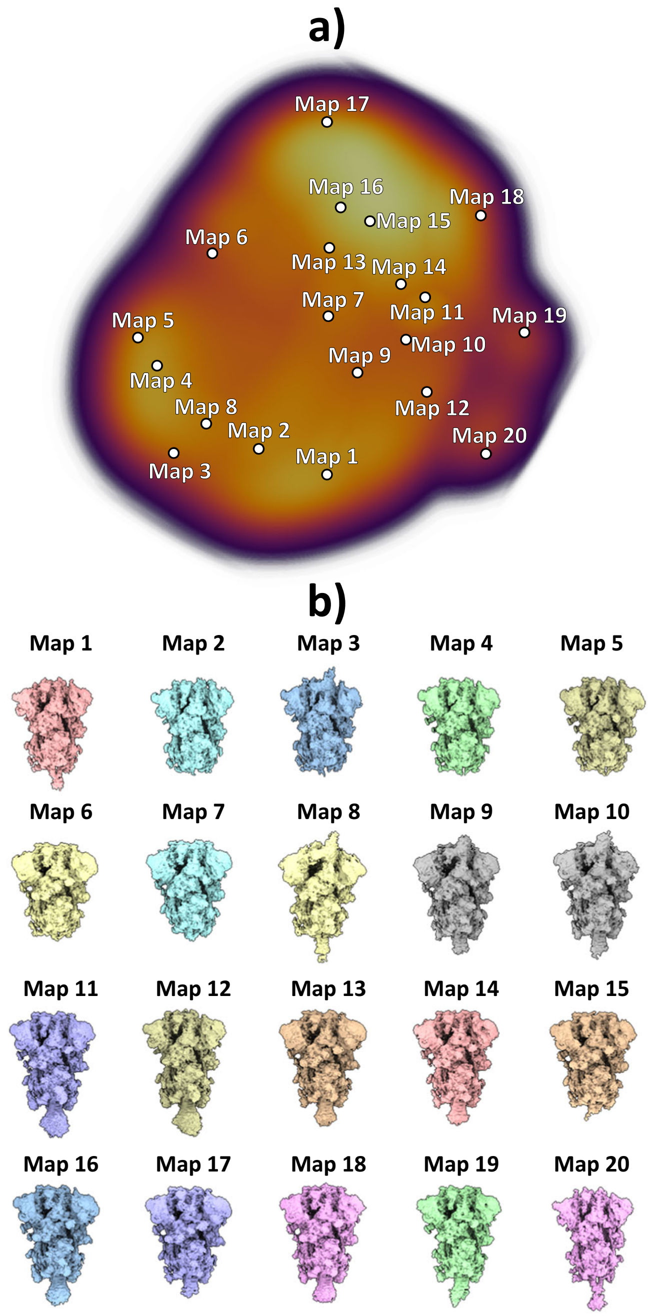


Fig. 8 | Conformational landscape analysis of the main motions detected by HetSIREN on the SARS-CoV-2 Spike protein at 37°C. Panel a) shows the UMAP[19] representation of the landscape obtained from the original 8D latent space encoded by HetSIREN from the particle images after training. Each dot in the landscape corresponds to the position of the cluster representative obtained from a KMeans clustering of the original HetSIREN latent space. Panel b) shows side views of the decoded HetSIREN volumes obtained from the latent space coefficients assigned to every representative shown in Panel a). The maps provide a sensible exploration of the different states identified by HetSIREN, including 3 Down and 1 Up conformations.

Methods

This section starts with a general presentation of the image formation model in CryoEM and then details the architecture of the HetSIREN network and training strategies.

In addition, Supplementary Table 1 summarizes the performance analysis of the proposed method in terms of the usage of computing resources. Performance metrics were evaluated with default parameters on an RTX Ada 6000 generation GPU.

Image formation model in CryoEM

One of the main goals of single particle analysis is to determine the 3D structure of a biomolecule through a set of 2D images generated by orthogonally integrating the electrostatic potential during micrograph acquisition. Therefore, the image formation model can be represented as a rotation operator, R_n, a translation operator, T_n, and a posterior projection, P, of the underlying 3D structure V_n. The subindex n

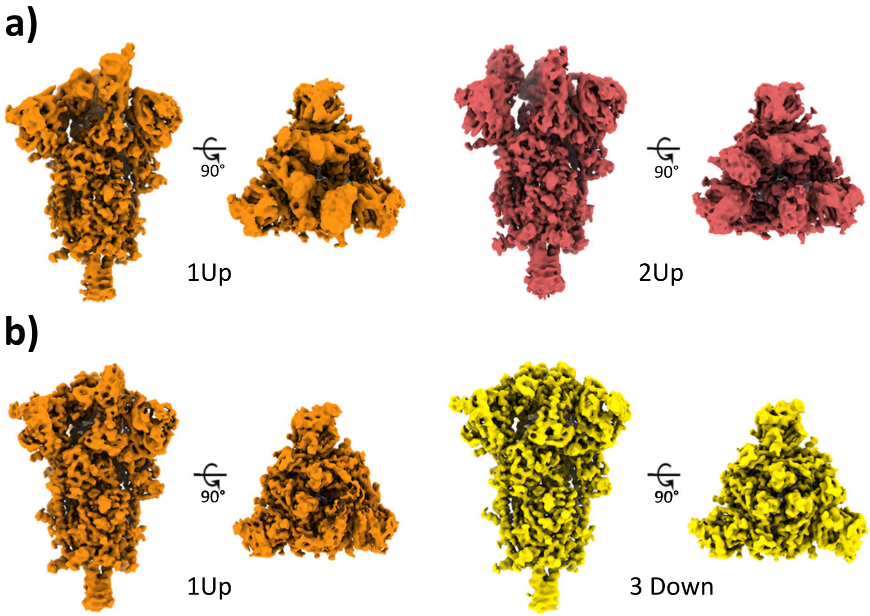


Fig. 9 | Structural differences of the SARS-CoV-2 Spike protein at different temperatures. Representative maps of the main conformational states detected by the HetSIREN network when the Spike protein is kept at 4°C (**a**) or 37°C (**b**). At 4°C (**a**) the Spike always shows at least one (orange) or two (red) RBDs in its open position. At 37°C (**b**) the Spike is mainly detected in its 3 Down conformation (yellow), with a lower population of particles exhibiting a 1 Up state (orange).

emphasizes that each image has a different rotation, translation, and underlying structure.

Although it is common practice to formulate the image formation model in Fourier space to take advantage of the central slice theorem, HetSIREN directly decodes the 3D structure $\boldsymbol{V}_n$ in real space. Equation 1 shows the formulation of the image formation model in real space as used in this work.

$$I_n = PSF_n * (P^\circ \mathrm{T_n}^\circ \mathrm{R_n})(V_n) + \epsilon_n \qquad (1)$$

where $\boldsymbol{\epsilon}_n$ is a term representing the noise added to the image, PSF_n is the point spread function that captures the optical characteristics of the CryoEM microscope, and * is the convolution operator.

Sinusoidal Representation Network (SIREN)

HetSIREN is based on a (modified) Sinusoidal Architecture Network (SIREN)[13] to increase the quality of the high-frequency characteristics of the decoded 3D structures.

The main contribution of SIREN architectures is using sinusoidal functions as activation functions of the neural network[13]. This activation has been shown to have faster convergence and higher representation fidelity than other popular representations of signals in deep learning, such as ReLU with positional encoding, traditionally used in CryoEM[6]. In addition, the gradient computation in a sine activation can be easily modulated, as it is represented by another SIREN activation function with a phase shift, which allows for finer tuning of the gradient computations to improve the representation capabilities of neural networks.

As described in the next section, the decoder architecture in HetSIREN relies on the real-space representation of 3D signals modulated by a set of SIREN activations in its hidden layers. A dense layer with linear activation follows this to compute the map values.

Whereas SIREN-based networks have improved signal approximation capabilities in many applications[13], they suffer from a decrease in performance when representing a whole set of different structures[35], which is a critical requirement in an application such as heterogeneity analysis. Consequently, the success of SIREN in this application has required the development of a modified architecture so that traditional meta-sinusoidal representations rely on two different layers sharing their weights to improve the inpainting capabilities on whole datasets (Supplementary fig. 3). In this meta-representation, one of the layers (commonly represented by a dense representation with ReLU activation) is used to compute the weights that will be posteriorly passed to the second SIREN layer to perform the forward pass through the network. During our experiments, we found the meta-architecture to have improved performance at the expense of slightly higher memory consumption.

It should be noted that it is possible to increase the number of ReLU layers used to compute the shared weights of its associated meta-sinusoidal layer to increase the inpainting capabilities of the model at the expense of more restricted time and GPU memory constraints.

HetSIREN network architecture

HetSIREN network follows an autoencoder architecture detailed in Supplementary fig. 4.

The proposed encoder implements two different architectures that lead to a latent space of a number of dimensions determined by the user and set by default to 10. The architectures implemented in the encoder include a multilayer perceptron network (MLP) with three hidden layers of 1024 neurons or a residual convolutional architecture. In practice, the two encoder architectures have obtained similar latent-space representations. However, MLPs have a higher chance of overfitting, while the convolutional architecture is more robust at the expense of slightly longer training times. By default, the convolutional architecture is chosen (Supplementary fig. 5), although it can be modified in the Scipion protocol to use the MLP model if desired.

The feature vectors extracted from the latent space $\boldsymbol{z}$ are then sent through the decoder, which performs the mapping $\Delta\boldsymbol{V}_n = \boldsymbol{f}(\boldsymbol{z}_n)$ (i.e., the feature vector of the n-th particle is mapped to a specific map). The density values $\Delta\boldsymbol{V}_n$ are then added to a reference map V_0 to obtain the final heterogeneous reconstruction:

$$\boldsymbol{V}_n = \boldsymbol{V}_0 + \boldsymbol{\Delta V}_n \qquad (2)$$

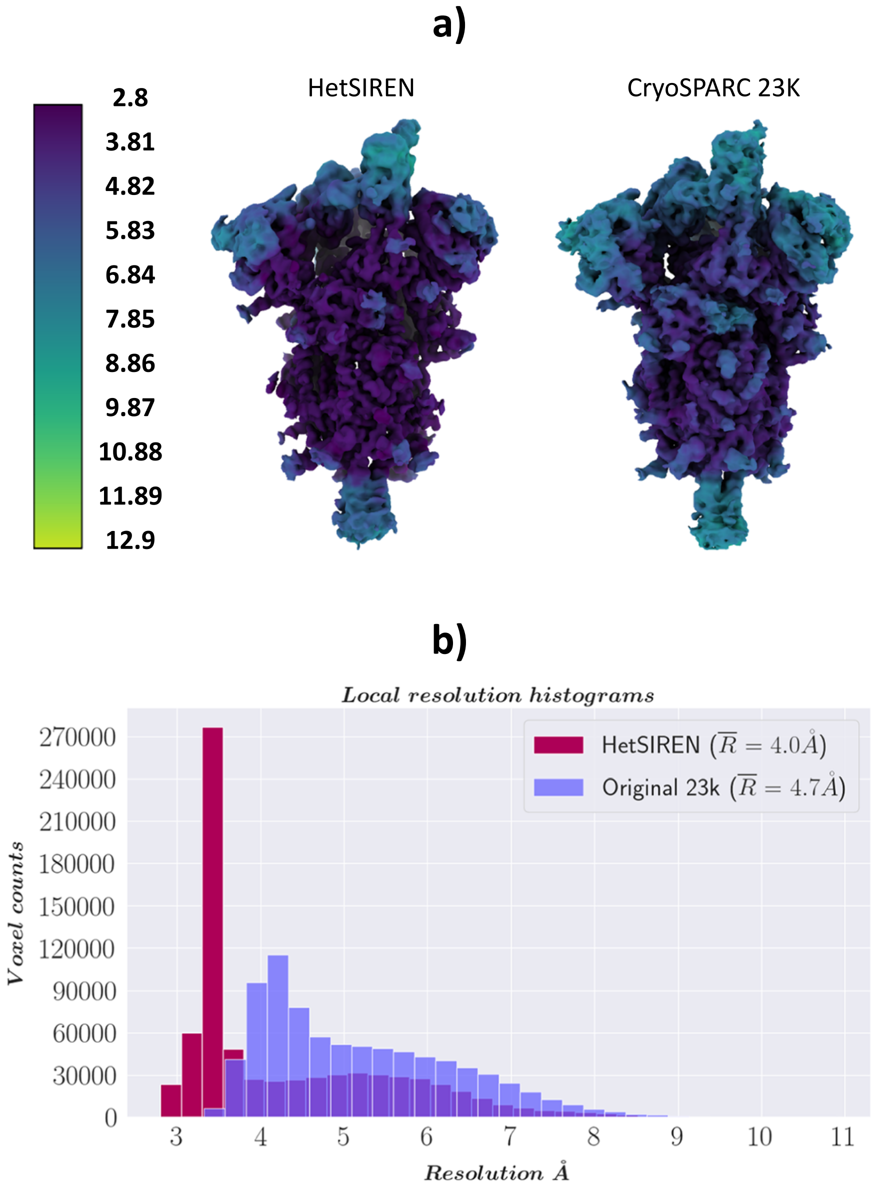


Fig. 10 | Resolution analysis of HetSIREN compared against the map refined with standard procedures. (**a**) shows the HetSIREN decoded map for one of the 1 Up conformation clusters (left), followed by the CryoSparc reconstruction of the 23766 particles closest to that cluster (right). Maps are colored according to their local resolution estimated with DeepRes[20]. (**b**) Quantitatively compares the estimated local resolutions based on local resolution histograms. Similarly to (**a**), the local resolution of HetSIREN (dark red) shows a displacement of the voxel resolutions to the high-resolution domain, translating into an improvement of 0.7 Å in the mean resolution to the CryoSparc reconstruction with the 23776 particles (light purple). Source data are provided as a Source Data file.

If the reference volume is empty $\boldsymbol{V}_0 = 0$, the decoder will directly generate the volume representing a given conformation $\boldsymbol{V}_n = \Delta\boldsymbol{V}_n$. Suppose that the reference volume is a homogeneously reconstructed density map. In that case, the network will produce the changes that will be applied to the reference to represent a new conformational state as shown in Eq. 2.

As described in the previous section, the decoder comprises a series of hidden meta-sinusoidal residual layers followed by a dense layer with linear activation that recovers $\Delta\boldsymbol{V}_n$. To keep the memory footprint of the decoder as low as possible, the number of hidden layers is fixed to three, with a total number of neurons and hyperneurons (i.e., the number of neurons in the dense ReLU layers composing the meta-sinusoidal layers) equal to the latent space dimension. The last dense layer has as many features as the density values needed to recover $\Delta\boldsymbol{V}_n$.

The previous decoder is followed by a physics-based decoder that implements the image formation model defined in Eq. 1. The generated computer-simulated projections are then compared with the experimental images to backpropagate the final loss during the training phase.

Disetangling of poses and CTF from conformations

Ideally, conformational latent spaces should only capture information on the structural changes a biomolecule may undergo based on the experimental data collected. However, conformational latent spaces suffer from a strong coupling of several factors apart from structural information, such as image pose and (to a lesser extent) CTF information. Indeed, in how structural changes are translated into projection images, differences depend strongly on the particle's projection geometry (the pose). Therefore, the interpretability of the estimated landscapes is largely compromised unless the effect of the previous factors is explicitly decoupled from the conformational landscapes.

Following a similar process to the one conceptually proposed in ref. 14, HetSIREN includes a decoupling architecture to effectively disentangle pose and CTF from the structural information captured in its conformational landscape. By coupling the disentangled landscapes with the high-resolution volume decoder, HetSIREN allows us to explore conformational landscapes with an improved understanding of the structural features that participate in different motions.

Supplementary fig. 6 shows the modified encoder architecture, including the decoupling workflow for both the poses and the CTF information. During the training phase, experimental images are forwarded through the encoder and the decoder, generating a cleaned set of theoretical projections, the CTF corrupted theoretical projections, and the latent space vectors $\boldsymbol{z}$.

The pose decoupling step relies on a second forward pass through the decoder to generate a new set of cleaned projections. Before that, the pose information associated with the current batch of images is shuffled and passed to the decoder with the corresponding latent space vectors encoded from the experimental images. Therefore, the second forward pass will generate a new set of cleaned images from the same conformation as the first generated projections but with a different pose. A random pose could also be passed to the decoder for this step, but shuffling ensures that the original pose distribution is maintained. The two sets of cleaned images are then forwarded through the pose decoupling encoder and the conformational latent layer to generate two new sets of latent vectors $\boldsymbol{z}_t$ and $\boldsymbol{z}_{p,t}$. Since these new latent vectors encode the same conformation as the first-generated vectors but at different poses, we can impose a constraint to place them as close as possible in the latent space:

$$\text{Loss} = \lambda_p\left(|\boldsymbol{z} - \boldsymbol{z}_t|_2^2 + |\boldsymbol{z} - \boldsymbol{z}_{p,t}|_2^2\right) \tag{3}$$

Upon convergence, the network will learn to produce the same conformation independently of the image pose, effectively decoupling the structural and pose information.

The CTF decoupling process follows a principle similar to the pose decoupling workflow described above. In this case, apart from the CTF corrupted theoretical projections, a new set of CTF corrupted images is generated through a third forward pass through the decoder. Before this new forward pass, the CTFs associated with the batch are shuffled without touching the poses. The new sets of images are then passed to the CTF decoupling encoder and the conformational latent layer to generate two new sets of latent vectors $\boldsymbol{z}_c$ and $\boldsymbol{z}_{p,c}$. Similarly to the previous case, the new latent vectors and the first generated vectors should be as close as possible in the latent space, as they represent the same conformation up to the CTF corruption. Therefore, we can impose a new regularization factor as follows:

$$\text{Loss} = \lambda_c\left(|\boldsymbol{z} - \boldsymbol{z}_c|_2^2 + |\boldsymbol{z} - \boldsymbol{z}_{p,c}|_2^2\right) \tag{4}$$

Once convergence is achieved, the network will learn to produce the same conformation independently of the CTF, effectively decoupling the structural and CTF information.

In the last instance, combining the two decoupling workflows allows the generation of a decoupled latent space where the conformational information predominates, increasing the interpretability of the conformational landscape.

HetSIREN cost function

The possibility of working directly in real space when decoding theoretical volumes/images allows the inclusion of additional constraints in the training objective function. These extra terms prevent the network from learning unwanted or meaningless features in the experimental images, such as noise or normalization errors.

Before introducing the different terms included in the HetSIREN objective function, we provide a simplified guideline of the implemented training strategy. The optimization of network parameters is based on the Adam optimization method with a custom learning rate (set by default at 10^{-5}) and a batch size (set by default to 8). Experimental images are forwarded through the network, leading the network output to a set of theoretically decoded projections. The comparison of experimental and theoretical projections uses, by default, a standard Mean Square Error (MSE) function as follows:

$$\text{Loss} = \sum_b |\boldsymbol{I}_b - D(\boldsymbol{z}_b)|_2^2 \tag{5}$$

where I_b is an experimental image in the batch of images being considered, D represents the decoder network, and z_b is the latent space vector associated with I_b by the encoder network:

$$\boldsymbol{z}_b = E(I_b) \tag{6}$$

Besides the standard MSE cost function, we provide alternative means to compare the theoretical and experimental projections to further customize the network training in the Scipion protocol, such as the correlation between the image pair.

Real-space regularization

The low signal-to-noise ratios encountered in particle images extracted from the acquired micrographs are probably the primary source of errors and overfitting in CryoEM image processing.

In a homogeneous reconstruction, one can use averaging of many images during the reconstruction process to reduce the noise level as much as possible. Nevertheless, the previous solution does not apply to the heterogeneous reconstruction case, where, ideally, the reconstruction of 3D structures per particle is the primary goal.

One possible way to regularize the noise in the decoded volumes is to apply a low-pass filter in the reconstructed Fourier space at the expense of decreasing the resolution of the reconstructed structures, which we tend to avoid. Fortunately, noise can be regularized in real space without sacrificing the high-frequency information content discarded by the low-pass filter, as indicated in the next paragraph.

Returning to the image formation model described in Eq. 1, it is possible to observe that the effect of the noise term ϵ is the addition of additional unwanted density values to the voxels in the volume/image grid. Therefore, we may penalize the cost function with an L_1 regularization term:

$$\text{Loss} = \ldots + \lambda_1|\boldsymbol{V}_0 + \Delta\boldsymbol{V}_n|_1 \tag{7}$$

The previous term enforces HetSIREN to learn a $\Delta\boldsymbol{V}_n$ that minimizes noise while preserving the structural high-frequency features of the decoded volumes. Depending on the conditions of the data set and the desired denoising level, it is possible to modify the regularization term λ_1 to improve the quality of the decoded structures. By default, the regularization parameter is set to 1.0, which we have practically found to introduce a good balance between the loss function terms for all the datasets currently tested.

Negative value mitigation

Ensuring proper background and noise normalization in CryoEM images introduces an artifact in the reconstructed structures represented as a set of negative values scattered along the volume grid. Although this is not usually a significant concern, we found that heterogeneous reconstruction benefits from regularizing this artifact, allowing the neural network to focus on the protein signal instead of compensating the generated values with unwanted negative voxels.

Therefore, the objective function is further modified, including a L_1 regularization term that penalizes adding negative values to the volume.

$$\text{Loss} = \ldots + \lambda_2 | \min(\boldsymbol{V}_0 + \Delta\boldsymbol{V}_n, 0)|_1 \tag{8}$$

Similarly to the previous case, controlling the regularization strength through the regularization parameter λ_2 is possible, which is set by default to 1.0.

Enforcing density continuity in decoded maps

Even in the presence of the denoising regularization term, overfitting might remain an issue, especially when working with large-volume grids to achieve high-resolution 3D structures. However, the nature of the overfitting differs from the addition of noise introduced previously: The decoded voxel values might appear as artifacts scattered along the protein signal so that the projection still matches the proper structure but without providing meaningful structural features at the volume level.

Therefore, ensuring proper density continuity is essential to allow the network to learn high-frequency structural details while maintaining a proper biomolecular shape. Our model controls density continuity through Total Variation (TV) losses, which account for different continuity features. The rationale behind TV is to promote overall smoothness in the image by reducing noise and minor fluctuations while allowing for sharper edges. Combining both penalizations encourages the decoded volume to have smooth transitions with a reduced likelihood of abrupt changes in pixel values while preserving edges that might otherwise be overly smooth.

Our implementation of TV regularizations is:

$$\text{Loss} = \ldots + \lambda_3 |\nabla(\boldsymbol{V}_0 + \Delta\boldsymbol{V}_n)|_1 + \lambda_4 |\nabla(\boldsymbol{V}_0 + \Delta\boldsymbol{V}_n)|_2^2 \tag{9}$$

where $\nabla\boldsymbol{V}$ represents the spatial gradient of $\boldsymbol{V}$.

As in previous regularization terms, the regularization strength can be controlled through the parameters λ_3 and λ_4, respectively; by default, both terms are set to 0.1.

Multiresolution loss to achieve high resolution in a single training

The two main objectives of heterogeneous reconstructions are to provide a meaningful latent space that orders the conformational states captured by the particles according to their similarity and to decode high-resolution maps with the decoder to explore and explain the latent space. The previous workflow generally involves training the network with the original images at full resolution. However, in practice, the previous approach is not always ensured to converge to a satisfactory solution due to the large amount of local minima present when minimizing the objective function.

One possible approach to overcome the local minima problem is to warm up the neural network. This implies initial training on downsampled images, which smooths the solution space, thus minimizing the chances of getting stuck on spurious local minima. The pretrained network is then fine-tuned on the unsampled data, allowing it to reach high-resolution structures without escaping from the initial solution.

However, the previous approximation requires at least two different training steps, which overall impacts the learning time needed by the network. In HetSIREN, we propose a multiresolution training approach that allows for robustly obtaining a high-resolution structure on a single model training step, which implies a significant improvement in the training time compared to the previously described strategy. The multiresolution approach minimizes the MSE between different pairs of images at different resolutions, allowing the network to explore both the smooth solution space defined by the filtered image pairs and the original solution space. In this way, the loss function becomes:

$$\text{Loss} = \sum_b \sum_\omega |L_\omega(I_b - D(\boldsymbol{z}_b))|_2^2 \tag{10}$$

where L_ω is a lowpass filter followed by a downsampling operator and ω is chosen from a discrete set of cutoff frequencies.

During multiresolution training, the original experimental images are forwarded through the network and used to decode volumes at the same pixel size as the experimental images. A bank of filters and downsampling operations is posteriorly applied to the experimental and decoded projections to generate the multi-resolution pairs. Finally, the pairs are compared through an MSE error loss and combined before backpropagation occurs.

It should be noted that the previously described regularization terms are only computed with the original decoded map at full resolution. In this way, the network focuses on improving the features of the unsampled volumes at the original pixel size.

In our tests, we discovered that using only a set of frequencies at full resolution (R) and half resolution ($2R$) typically achieves the desired outcomes, streamlining the trade-off between the MSE costs of the original and downsampled image pairs.

Focused reconstruction in HetSIREN

On many occasions, 3D structure reconstruction in CryoEM is carried out on the entire volume grid that contains the biomolecule of interest (or, for memory-saving purposes, on the sphere inscribed in the cubic grid). However, the region of interest might be more complex in some scenarios. For example, nanodiscs or membranes in the reconstructed maps are usually undesired as they might affect the proper reconstruction of the embedded structural features.

In heterogeneous reconstruction, the motivation for designing such a mask follows a similar reasoning, with the addition of focusing the latent space so that only the conformational changes in a region of interest are captured. Therefore, HetSIREN allows custom reconstruction masks that determine the area to be reconstructed by the neural network.

The implementation of focused reconstruction in HetSIREN is based on a mask M designed and input by the user in the form of the Scipion protocol. The network configuration is modified to accommodate the focused reconstruction process if a mask is provided. The main change applied to the network is to limit the number of voxels considered in the volume decoder D to only those present in the mask as follows:

$$\Delta\boldsymbol{V}_n = D_M(\boldsymbol{z}) \tag{11}$$

Here, D_M represents the new decoder focused on mask voxels. The previous modification allows us to generate theoretical projections only for the structural changes within the mask.

Even though the network will learn to modify only the regions within the mask, we found it helpful to consider the voxels out of these regions when generating the theoretical 2D projections. Thanks to the $\Delta\boldsymbol{V}$ implementation in HetSIREN, it is possible to project the entire 3D volume along a given particle projection direction once the desired reconstruction region has been refined according to the decoded values. In this way, obtaining a set of theoretical projections with homogeneous information is possible apart from the area enclosed by the mask defined by the user.

Once the 2D projections have been generated, it is possible to use the cost function and regularization previously described to train the network. However, the previous cost functions will not ensure that the region being refined/reconstructed will follow a similar voxel value distribution to the one in the original map. Therefore, when focusing on the landscape, an additional regularization parameter is added to ensure that the application of $\Delta\boldsymbol{V}$ respects the voxel value distribution of the reference volume. Being $\boldsymbol{v}^{D,n}$ and $\boldsymbol{v}^{D,r}$ the vector storing the voxel values within the region of interest for the HetSIREN and reference volumes, respectively, the new regularization reads:

$$\begin{aligned} Loss = \ldots + \lambda_5 \Bigg(& \sum_b \left(\max\left(\boldsymbol{v}_b^{D,n}\right) - \max\left(\boldsymbol{v}_b^{D,r}\right)\right)^2 \\ & + \sum_b \left(\min\left(\boldsymbol{v}_b^{D,n}\right) - \min\left(\boldsymbol{v}_b^{D,r}\right)\right)^2 \\ & + \sum_b \left(\mu_b^{D,n} - \mu_b^{D,r}\right)^2 \\ & + \sum_b \left(\sigma_b^{D,n} - \sigma_b^{D,r}\right)^2 \Bigg) \end{aligned} \tag{12}$$

where μ and σ represent the mean and standard deviation values stored in the vectors.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The atomic coordinates and cryo-EM density maps for the SARS-CoV-2 Spike protein at 4 °C and 37 °C were deposited in the Protein Data Bank and EM Data Bank with codes 9GDX and 9GDY and EMD-51279 and EMD-51280, respectively. The synthetic and ribosome dataset analyzed in this work can be downloaded as a Scipion test dataset through the following command: scipion3 testdata --download FlexHub_Tutorials (assuming Scipion is already installed in the system). The atomic model used in the synthetic dataset is deposited in the Protein Data Bank with code 4AKE. The source data underlying Figs. 4b, 6b, 10b and Supplementary fig. 7a, b, d are provided as a Source Data file. Source data are provided with this paper.

Code availability

HetSIREN algorithm is freely available through Scipion 3.0[16] under the plugin scipion-em-flexutils[36] https://github.com/scipion-em/scipion-em-flexutils and the package Flexutils-Toolkit[37] https://github.com/I2PC/Flexutils-Toolkit. The protocol corresponding to the algorithm described in this manuscript is flexutils - flexible align - HetSIREN. Tutorials on how to setup and use HetSIREN are provided in the following webpage https://scipion-em.github.io/docs/release-3.0.0/docs/user/tutorials/flexibilityHub/main_page.html#tutorials.

References

1. Carroni, M. & Saibil, H. R. Cryo electron microscopy to determine the structure of macromolecular complexes. *Methods* **95**, 78–85 (2016).
2. Scheres, S. H. W. et al. Modeling Experimental Image Formation for Likelihood-Based Classification of Electron Microscopy Data. *Structure* **15**, 1167–1177 (2007).
3. Toader, B., Sigworth, F. J. & Lederman, R. R. Methods for Cryo-EM Single Particle Reconstruction of Macromolecules Having Continuous Heterogeneity. *J. Mol. Biol.* **435**, 168020 (2023).
4. Dashti, A. et al. Trajectories of the ribosome as a Brownian nanomachine. *Proc. Natl Acad. Sci. USA* **111**, 17492–17497 (2014).
5. Jin, Q. et al. Iterative elastic 3D-to-2D alignment method using normal modes for studying structural dynamics of large macromolecular complexes. *Structure* **22**, 496–506 (2014).
6. Zhong, E. D., Bepler, T., Berger, B. & Davis, J. H. CryoDRGN: reconstruction of heterogeneous cryo-EM structures using neural networks. *Nat. Methods* **18**, 176–185 (2021).
7. Chen, M. & Ludtke, S. J. Deep learning-based mixed-dimensional Gaussian mixture model for characterizing variability in cryo-EM. *Nat. Methods* **18**, 930–936 (2021).
8. Luo, Z., Ni, F., Wang, Q. & Ma, J. OPUS-DSD: deep structural disentanglement for cryo-EM single-particle analysis. *Nat. Methods* **20**, 1729–1738 (2023).
9. Schwab, J., Kimanius, D., Burt, A., Dendooven, T. & Scheres, S. H. W. DynaMight: estimating molecular motions with improved reconstruction from cryo-EM images. *Nat. Methods* **21**, 1855–1862 (2024).
10. Punjani, A. & Fleet, D. J. 3DFlex: determining structure and motion of flexible proteins from cryo-EM. *Nat. Methods* **20**, 860–870 (2023).
11. Herreros, D. et al. Approximating deformation fields for the analysis of continuous heterogeneity of biological macromolecules by 3D Zernike polynomials. *IUCrJ* **8**, 992–1005 (2021).
12. Vuillemot, R. et al. MDSPACE: Extracting Continuous Conformational Landscapes from Cryo-EM Single Particle Datasets Using 3D-to-2D Flexible Fitting based on Molecular Dynamics Simulation. *J. Mol. Biol.* **435**, 167951 (2023).
13. Sitzmann, V., Martel, J. N. P., Bergman, A. W., Lindell, D. B. & Wetzstein, G. Implicit Neural Representations with Periodic Activation Functions. *arXiv* **2006.09661** (2020).
14. Klindt, D. A., Hyvärinen, A., Levy, A., Miolane, N. & Poitevin, F. Towards interpretable Cryo-EM: disentangling latent spaces of molecular conformations. *Front Mol. Biosci.* **11**, 1393564 (2024).
15. Harastani, M., Eltsov, M., Leforestier, A. & Jonic, S. HEMNMA-3D: Cryo Electron Tomography Method Based on Normal Mode Analysis to Study Continuous Conformational Variability of Macromolecular Complexes. *Front. Mol. Biosci.* **8**, 663121 (2021).
16. de la Rosa-Trevín, J. M. et al. Scipion: A software framework toward integration, reproducibility and validation in 3D electron microscopy. *J. Struct. Biol.* **195**, 93–99 (2016).
17. Wong, W. et al. Cryo-EM structure of the Plasmodium falciparum 80S ribosome bound to the anti-protozoan drug emetine. *Elife* **2014**, e03080 (2014).
18. Herreros, D. et al. Scipion Flexibility Hub: an integrative framework for advanced analysis of conformational heterogeneity in cryoEM. *Acta Crystallogr D. Struct. Biol.* **79**, 569–584 (2023).
19. McInnes, L., Healy, J., Saul, N. & Großberger, L. UMAP: Uniform Manifold Approximation and Projection. *J. Open Source Softw.* **3**, 861 (2018).
20. Ramírez-Aportela, E., Mota, J., Conesa, P., Carazo, J. M. & Sorzano, C. O. S. DeepRes: a new deep-learning- and aspect-based local resolution method for electron-microscopy maps. *IUCrJ* **6**, 1054–1063 (2019).
21. Noddings, C. M., Johnson, J. L. & Agard, D. A. Cryo-EM reveals how Hsp90 and FKBP immunophilins co-regulate the glucocorticoid receptor. *Nat. Struct. Mol. Biol.* **30**, 1867–1877 (2023).
22. Jollife, I. T. & Cadima, J. Principal component analysis: a review and recent developments. *Philos. Transac. Royal Soc. A: Math., Phys. Eng. Sci.* **374**, 20150202 (2016).
23. Zhou, P. et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nat. 2020 579:7798* **579**, 270–273 (2020).
24. Yan, R. et al. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science* **367**, 1444–1448 (2020).
25. Ginex, T. et al. The structural role of SARS-CoV-2 genetic background in the emergence and success of spike mutations: The case of the spike A222V mutation. *PLoS Pathog.* **18**, e1010631 (2022).
26. Edwards, R. J. et al. Cold sensitivity of the SARS-CoV-2 spike ectodomain. *Nat. Struct. Mol. Biol. 2021 28:2* **28**, 128–131 (2021).
27. Yang, T. J., Yu, P. Y., Chang, Y. C. & D. Hsu, S. Te. D614G mutation in the SARS-CoV-2 spike protein enhances viral fitness by desensitizing it to temperature-dependent denaturation. *J. Biol. Chem.* **297**, 101238 (2021).

28. Chen, C. Y., Chang, Y. C., Lin, B. L., Huang, C. H. & Tsai, M. D. Temperature-Resolved Cryo-EM Uncovers Structural Bases of Temperature-Dependent Enzyme Functions. *J. Am. Chem. Soc.* **141**, 19983–19987 (2019).
29. Chang, Y. C., Chen, C. Y. & Tsai, M. D. Preparation of High-Temperature Sample Grids for Cryo-EM. *J. Vis. Exp.* **2021**, https://doi.org/10.3791/62772 (2021).
30. Krieger, J. M., Sorzano, C. O. S. & Carazo, J. M. Scipion-EM-ProDy: A Graphical Interface for the ProDy Python Package within the Scipion Workflow Engine Enabling Integration of Databases, Simulations and Cryo-Electron Microscopy Image Processing. *Int. J. Mol. Sci.* **24**, 14245 (2023).
31. Jamali, K. et al. Automated model building and protein identification in cryo-EM maps. *Nat. 2024 628:8007* **628**, 450–457 (2024).
32. Benton, D. J. et al. The effect of the D614G substitution on the structure of the spike glycoprotein of SARS-CoV-2. *Proc. Natl. Acad. Sci. USA* **118**, (2021).
33. Bruch, E. M. et al. Structural and biochemical rationale for Beta variant protein booster vaccine broad cross-neutralization of SARS-CoV-2. *Sci. Rep.* **14**, 1–16 (2024).
34. Wolterink, J. M., Zwienenberg, J. C. & Brune, C. Implicit Neural Representations for Deformable Image Registration. *Proc. Mach. Learn. Res.* **172**, 1349–1359 (2022).
35. Chauhan, V. K., Zhou, J., Lu, P., Molaei, S. & Clifton, D. A. A brief review of hypernetworks in deep learning. *Artif. Intell. Rev.* **57**, 1–29 (2024).
36. Herreros, D. et al. Real-space heterogeneous reconstruction, refinement, and disentanglement of CryoEM conformational states with HetSIREN, scipion-em-flexutils, https://doi.org/10.5281/zenodo.14980837 (2025).
37. Herreros, D. et al. Real-space heterogeneous reconstruction, refinement, and disentanglement of CryoEM conformational states with HetSIREN, Flexutils-Toolkit, https://doi.org/10.5281/zenodo.14980835 (2025).
38. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods 2017 14:3* **14**, 290–296 (2017).

Acknowledgements

The authors acknowledge the financial support from the Ministry of Science, Innovation and Universities (BDNS n. 716450) to the Instruct Image Processing Center (I2PC) as part of the Spanish participation in Instruct-ERIC, the European Strategic Infrastructure Project (ESFRI) in the area of Structural Biology, Grant [PID2022-136594NB-I00] (COSS and JMC) funded by MICIU /AEI/10.13039/501100011033/ and "ERDF A way of making Europe", by the "European Union", Spanish State Research Agency AEI/10.13039/501100011033, through the "Severo Ochoa" Programme for Centres of Excellence in R\&D [CEX2023-001386-S] (COSS and JMC), Comunidad Autónoma de Madrid" through Grant: S2022/BMD-7232 (COSS and JMC), European Union (EU) and Horizon 2020 through grant: HighResCells (ERC - 2018 - SyG, Proposal: 810057) (COSS and JMC) European Union (EU), and Horizon Europe through grant: Fragment Screen Proposal: 101094131 (COSS and JMC). We thank Medigen Vaccine Biologics Corp. (MVC, No. 68, Shengyi 3rd Rd., Zhubei City, Hsinchu County 302, Taiwan) for providing the SARS-CoV-2 spike protein used in this work. The SARS-CoV-2 spike protein cryo-EM work was supported by Academia Sinica (AS-KPQ-109-TPP2) (MDT) and Taiwan Cryo-EM Consortium (grant no. NSTC 113-2740-B-006-004) (MDT). Cryo-EM experiments were performed at the Academia Sinica Cryo-EM Facility supported by Academia Sinica (AS-CFII-108-110) (MDT), and initial image processing steps were performed at the Academia Sinica Grid-computing Center. We also thank Dr. Mohamad Harastani for helping generate the adenylate kinase dataset.

Author contributions

D.H. developed and tested the HetSIREN method presented throughout the manuscript. C.P.M. and J.K. processed and analyzed the datasets presented in the manuscript. D.I. prepared cryo-EM grids, collected data, and pre-processed the SARS-CoV-2 datasets. C.N. and DA provided and helped to understand the GR:Hsp90:FKBP51 dataset. M.D.T. provided and helped with the understanding of the SARS-CoV-2 Spike datasets. C.O.S.S. and J.M.C. jointly supervised this work. D.H. and C.P.M. wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-025-59135-0.

Correspondence and requests for materials should be addressed to David Herreros.

Peer review information *Nature Communications* thanks Muyuan Chen and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary Information

Real-space heterogeneous reconstruction, refinement, and disentanglement of CryoEM conformational states with HetSIREN

David Herreros*[1], Carlos Perez Mata[1,2], Chari Noddings[3], Deli Irene[4], James Krieger[1], David A. Agard[5,6], Ming-Daw Tsai[4], Carlos Oscar Sanchez Sorzano+[1], and Jose Maria Carazo+[1]

[1]Centro Nacional de Biotecnologia-CSIC, C/ Darwin, 3, 28049, Cantoblanco, Madrid, Spain
[2]PKF Attest innCome, Orense 81, 28020, Madrid
[3]Altos Labs, Redwood City, CA, USA
[4]Institute of Biological Chemistry, Academia Sinica, Taipei 115, Taiwan
[5]Department of Biochemistry Biophysics, University of California, San Francisco, CA, USA
[6]Chan Zuckerberg Imaging Institute, Redwood City, CA, USA
+These authors jointly supervised this work: C.O.S. Sorzano and J.M. Carazo

* Corresponding author
E-mail: dherreros@cnb.csic.es

Supplementary Methods

Cryo-EM sample preparation and data collection for the SARS-CoV-2 samples

0.5 mg/ml of purified Beta variant Spike protein sample in 1X PBS buffer at pH 7 was diluted by 100 mM sodium citrate tribasic dihydrate at pH 5 to a final concentration of 0.25 mg/ml and a final pH value of 5.5, a condition in which the preferred orientation was minimized. A 4 μl sample was applied to Quantifoil holey carbon grids R1.2/1.3 200 mesh for 4°C, and to Quantifoil gold grids R2/2 200 mesh for 37°C, with prior incubation at the respective temperatures for 10 min. The grids had been glow discharged with negative polarity at 25 mA for 30 seconds using an EMS 100 Glow discharge apparatus. They were used within 30 min to minimize the hydrophilic recovery of the grids. After application of the sample, the grids were incubated for 10 seconds in 100% humidity at 4°C or 37°C in a Mark IV vitrobot device (Thermo Fisher Scientific) and then blotted for 3 seconds with blot force 0 before being flash frozen in liquid ethane cooled by liquid nitrogen.

For the 4°C and 37°C samples, 11,137 and 7,064 movie micrographs were automatically collected on a Bio-quantum-K3 detector (Gatan, Inc.) at a nominal magnification of 81,000x which resulted in a pixel size of 1,061 Å by using a Titan Krios microscope (Thermo Fisher Scientific) operating at 300 keV with a GIF Quantum energy filter with a slit width of 20 eV. 50 frames per movie were collected at 1 e-/ $Å^2$ per frame for a total dose of 50 e-/ $Å^2$ on the sample by using counting mode at a defocus range between -1.5 μm to -2.2 μm.

Standard image processing workflow for the SARS-CoV-2 samples

All image processing steps were performed within the Scipion software framework (1). For both samples, particles were previously pooled through standard 2D classification approaches in CryoSPARC (7) conducted by the laboratory of Prof. Ming-Daw Tsai. These particles were then directly imported into Scipion, with 662,379 and 468,911 particles for the samples at 4°C and 37°C, respectively. The selected particles were downsampled to 1.4 Å/px. These particles generated four *ab initio* models imposing C3 symmetry in CryoSPARC (7). All particles were subjected to non-uniform refinement using the best initial model low-pass filtered to 30 Å as a reference. This refinement was followed by an angular consensus protocol (8), retaining the best 615,000 and 410,000 particles for the samples at 4°C and 37°C, respectively. We then symmetry-relaxed these C3 symmetry-refined particles into C1 (9) while performing a 3D classification into 10 classes as implemented in Relion (10). We employed a 3D clustering consensus protocol to retain stable and statistically significant particles across the entire datasets to minimize the variability in class distribution over replicates of the same protocol. We inspected the particle clusters with a p-value < 0.05, and to confirm the assignment of particles to the different conformations, we generated initial models and non-uniform refined them independently. We then rejected the clusters of particles resulting in junk volumes and selected only the best clusters, corresponding to 479,908 and 309,062 particles for the samples at 4°C and 37°C, respectively. After this standard image processing workflow, we merged all clusters for each independent sample and subjected the corresponding particles to HetSIREN.

Model building and refinement of the SARS-CoV-2 samples

Firstly, we manually docked the model into the density as a rigid body, followed by real space fitting using the Fit in Map routine in UCSF Chimera (11) for the complete Spike structure, which includes chains A, B, and C. We used previously deposited structures as starting models, matching each detected conformation: PDB IDs 7WEV and 7VX1 for Beta variant in the 3 Down and 1 Up states, respectively (12). For the 2 Up state, we computationally modified the 1 Up model (7VX1) by removing one of the RBD Down chains and replacing it with a previously duplicated and individually fitted RBD Up chain. Real-space refinement was then performed in Phenix (13) with the enabled global minimization, local grid search, ADP, and rigid body options. We defined each chain's NTD, RBD, and S2 domains as independent rigid bodies, resulting in 9 rigid bodies in total. To preserve the general arrangement of the different domains within the Spike protein, the starting model was used as a reference model with restraints and secondary structure restraints. The resulting models were then manually inspected in Chimera (11) and Coot (14) to check the fit to the density. The quality of the obtained models was assessed using MolProbity (15) as implemented in Phenix (16) and the Worldwide PDB (wwPDB) OneDep System (https://deposit-pdbe.wwpdb.org/deposition). Refinement statistics are listed in Supplementary Table 3 and 4.

Characterization of the decoupling effect on conformational landscapes

The decoupling architecture introduced in HetSIREN minimizes the effects that the pose and CTF have on the organization of the different images in a latent space. Ideally, a conformational latent space that considers only the structural differences in the particle images should be learned. This way, the conformational landscape of the biomolecule under study can be properly reflected.

To better reflect these effects in the conformational latent spaces, we propose two scenarios in which the pose and CTF's downstream effects dominate.

The first test case consists of analyzing the simulated adenylate kinase dataset presented in the first section of the manuscript. During the simulation, a different CTF corruption was applied to each projection individually, trying to make the CTF as prominent as possible against the pose and the conformational variability captured in the images.

The previous images were used to train two different HetSIREN networks. The first network has a standard architecture without decoupling, while the second network includes the CTF decoupling architecture but not the pose decoupling part. Comparing the two landscapes allows one to better observe how the CTF decoupling architecture affects the latent space organization.

The results of this analysis are summarized in Supplementary Figure 8. The landscape colors represent a clustering of the different images according to their CTF information to simplify the visualization of the organization of the images based on the CTF information. Supplementary Figure 8a shows the landscape learned by the network with no decoupling architecture, leading to a significant landscape spreading to accommodate different ”bands” with similar CTFs. This is a strong deviation from the gold standard landscape, which should be a straight line, as discussed in the manuscript section "Simulated adenylate kinase landscape and landscape disentanglement." In contrast, the CTF-decoupled latent space shown in Supplementary Figure 8b shows a more condensed latent space, better reflecting the ideal latent space. By combining images with the same conformation and variable CTF information, HetSIREN effectively learns to decouple the CTF effect, minimizing its effect on the organization of the latent space and leading to a more prominent structural component.

The second test relies on analyzing clean images without CTF, which allows us to better assess the pose effect on the conformational landscape. To that end, we simulated 2000 images from two SARS-CoV-2 Spike electron density maps in one-up and three-down conformations. This simulated dataset describes a very simple conformational latent space, ideally consisting of two isolated points representing the two discrete states used to simulate the images.

Similar to the previous test, two HetSIREN networks were trained with the new image dataset: the first network has a standard architecture with no decoupling parts, while the second includes only the pose decoupling architecture to analyze its effect on the landscape. The results of this analysis are summarized in Supplementary Figure 9. Supplementary Figure 9a shows the landscapes obtained from the training dataset, colored according to clustering into four groups of the projection sphere to better visualize the pose. The non-decoupled landscape suffers from a similar effect to the CTF case, deviating from the ideal "two dots" latent space due to a strong organization induced by the pose. In contrast, the decoupled landscape presented is significantly condensed towards the ideal "two dots" representation, showing the ability of the architecture to effectively learn that images with similar conformation and different poses should be close in the latent space. A different experiment is proposed in Supplementary Figure 9b, where the previously trained networks are used to predict the landscape of a new dataset composed of the original images after applying noise to their poses. As can be seen from this result, the non-decupled network predicts a disordered landscape, placing the particles in completely different locations compared to the landscape shown in panel a), even if the images are the same. In contrast, the decupled landscape is not so much affected by the new poses, as it has learned to predict that the images represent two distinct conformations independently of their pose.

Comparison of SIREN and ReLU activation in HetSIREN

Applying different activation functions to the outputs of the layers in a neural network may induce differences in the accuracy and performance of a neural network. In this manuscript, we propose the application of sine activation functions well known in the deep learning field as SIRENs. Even though SIREN activation functions usually outperform other popular activations like ReLU, it is interesting to evaluate their effect on HetSIREN and its architecture.

To properly assess the differences between SIREN and ReLU in HetSIREN, we propose a test with the EMPIAR 10028 (2) dataset analyzed throughout the manuscript, which will be used to train two networks: the first one consists of HetSIREN with ReLU activations without adding the decoupling and the additional cost functions proposed in this work to isolate the effect of the activation function. The second network follows the same principles as the first one, changing the activation function to the SIREN activations presented in this work.

After training the two networks, two different conformations were selected from the conformational landscapes, decoding two volumes representing two distinct compositional states found in the dataset: one of the conformations loses completely its 40S subunit, while the second has a smaller loss of mass in the 40S subunit of the ribosome. The comparison of these two states is presented in Supplementary Figure 10.

As can be seen from the decoded volumes, both ReLU and SIREN perform similarly in our network architecture regarding the structural details of the structures. However, a significant difference is highlighted in the upper images arising from the change in the activation function.

While ReLU activations prevent the network from learning how to completely remove the 40S subunit of the ribosome, SIREN activations lead to a more sensible representation of this evasive state thanks to a clearer removal of the subunit.

Cost function ablation studies

As discussed in the Methods manuscript section, HetSIREN implements different cost functions directly affecting the decoded volume representation, trying to guide the network toward learning more accurate and interpretable 3D volumes from the images.

One of the effects of the proposed cost functions is to directly tackle the noise present in the images, allowing the network to focus on the signal instead of learning how to add noise to the decoded volumes. To better assess the effect of the previous denoising, we proposed an ablation test starting from the simulated adenylate kinase images already described in the manuscript. The tests first analyze the set of noise-free images, followed by a progressive addition of noise. In all these steps, two HetSIREN networks were trained: one did not include the additional denoising cost functions, unlike the second network, which is allowed to learn how to denoise the decoded volumes. It is important to highlight that the three denoising cost functions proposed in the manuscript (L1 regularization and the two versions of the total variation) are evaluated together as they complement each other to reduce the denoise while preserving the relevant details.

The results obtained from the previous analysis are summarized in Supplementary Figure 11. As explained before, the first step is analyzing the original 500 noise-free images. The projections of the decoded volumes show that both HetSIREN networks could identify the correct structure. However, it is possible to observe a non-uniform background when the denoising cost functions are not included, probably generated as a CTF effect. As expected, the network properly detected the conformational change captured by the images.

The analysis continuous with a new set of noisy images simulated to have a medium noise. When medium noise is added, it is possible to observe a more drastic effect on the two neural networks. The network without denoising adds a considerable amount of noise to the decoded volume in an attempt to match the denoise of the projection, unlike the network with denoising that manages to get a noise-free volume similar to one obtained with the baseline images. Similarly to the previous case, the detected conformational change is the expected one.

Lastly, a dataset with a high level of noise was analyzed. When no denoising is considered, the decoded volumes lack any meaningful signal and are completely dominated by noise. However, the network with denoising manages not only to detect the right signal but also to produce a noise-free decoded volume with features similar to those of the baseline dataset. This result shows the strong effect of handling the noise directly with the network, allowing it to learn accurate and meaningful structure representations even in highly noisy conditions. Moreover, the network with denoising also manages to properly detect the expected conformational change, showing the ability of the network to perform heterogeneous reconstruction with small and noisy datasets.

The next set of cost functions to be evaluated are those related to the focused reconstruction/refinement introduced in the manuscript. The main purpose of these const functions is to regularize the neural network so that it learns to refine the map while preserving the original voxel value characteristics in the reference volume given to the network. Similarly to the case before, it is required to consider these cost functions simultaneously to properly evaluate their effect, as their combination is needed to properly represent the voxel value distribution in the original volume.

For this test, we trained two neural networks using the EMPIAR 10028 dataset to simplify comparing the results with those presented in the manuscript. The only difference between the two networks trained is the consideration of the cost functions related to the focused reconstruction process. The results from this test are summarized in Supplementary Figure 12. As can be seen from the figure, when the cost functions are not considered, the network introduces a strong artifact in the decoded volume. This artifact arises from the freedom the network has to place any possible voxel value in a given position in the grid, completely breaking the relation of the decoded values with the original distribution of voxel values in the reference volume. In contrast, the regularized network effectively learns to refine the region of interest while considering that the range of values of the decoded region should be as similar as possible to the reference volume. Thus, the regularized network does not present the artifact previously described, improving the representation and interpretability of the decoded volume.

Apart from the denoising and focused reconstruction-related losses previously evaluated, HetSIREN includes an internal sharpening arising as a post-processing effect from the way the decoded volumes are constructed, which is added to the enhancement effects of the additional cost functions. However, this internal sharpening does not prevent further post-processing of the decoded volume to further enhance the structural features in the volume, which is an essential step to properly understand and interpret a given biomolecular structure.

To better reflect the previous idea, we propose a comparison of HetSIREN when the internal sharpening and a further post-processed version of the decoded volume with DeepEMhancer (17) and EMReady (18). To that end, we compared one of the HetSIREN volumes decoded for the EMPIAR-10028 dataset previously discussed in the manuscript. The comparison is presented in Supplementary Figure 13. We propose as a baseline of the comparison the CryoSPARC volume reconstructed from this dataset. The comparison shows how the internal sharpening of HetSIREN significantly improved the structural features in the volume, which are similar to the ones present in the CryoSPARC volume post-processed with DeepEMhancer. In addition, the sharpening post-processing of the HetSIREN volume enhances even further the structural features compared to its non-sharpened version and the sharpened CryoSPARC volume.

References

[1] J.M. de la Rosa-Trevín, A. Quintana, L. del Cano, A. Zaldívar, I. Foche, J. Gutiérrez, J. Gómez-Blanco, J. Burguet-Castell, J. Cuenca-Alba, V. Abrishami, J. Vargas, J. Otón, G. Sharov, J.L. Vilas, J. Navas, P. Conesa, M. Kazemi, R. Marabini, C.O.S. Sorzano, and J.M. Carazo. Scipion: A software framework toward integration, reproducibility and validation in 3D electron microscopy. *Journal of Structural Biology*, 195(1):93–99, 2016.

[2] W. Wong, X. Bai, A. Brown, I.S. Fernandez, E. Hanssen, M. Condron, Y.H. Tan, J. Baum, and S.H.W. Scheres. CryoEM structure of the *Plasmodium falciparum* 80s ribosome bound to the anti-protozoan drug emetine. *eLife*, 3:e03080, 2014.

[3] L. McInnes, J. Healy, N Saul, and L. Großberger. Umap: Uniform manifold approximation and projection. *Journal of Open Source Software*, 3(29):861, 2018.

[4] CM Noddings, J.L. Johnson, and D. Agard CryoEM reveals how Hsp90 and FKBP immunophilins co-regulate the glucocorticoid receptor. *Nature Structural & Molecular Biology*, 30:1867–1877, 2023.

[5] I. Jolliffe and J. Cadima. Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374:20150202, 2016.

[6] K. Jamali, L. Käll, R. Zhang, A. Brown, D. Kimanius, and S. H. W. Scheres. Automated model building and protein identification in cryo-EM maps. *Nature*, 628(8007):450–457, Apr 2024.

[7] A. Punjani, J.L. Rubinstein, D.J. Fleet, and M.A. Brubaker. CryoSPARC: algorithms for rapid unsupervised CryoEM structure determination. *Nature Methods*, 14:290–296, 2017.

[8] J.M. de la Rosa-Trevín, J. Otón, R. Marabini, A. Zaldívar, J. Vargas, J.M. Carazo, and C.O.S. Sorzano. Xmipp 3.0: An improved software suite for image processing in electron microscopy. *Journal of Structural Biology*, 184:321–328, 2013.

[9] V. Abrishami, S. L. Ilca, J. Gomez-Blanco, I. Rissanen, J. M. de la Rosa-Trevín, V. S. Reddy, J. -M. Carazo, and J. T. Huiskonen. Localized reconstruction in Scipion expedites the analysis of symmetry mismatches in cryo-EM data. *Prog. Biophys. Mol. Biol.*, 160:43–52, 2021.

[10] D. Kimanius, L. Dong, G. Sharov, T. Nakane, and S.H.W. Scheres. New tools for automated CryoEM single-particle analysis in RELION-4.0. *Biochemical Journal*, 478:4169–4185, 2021.

[11] E.F. Pettersen, T.D. Goddard, C.C. Huang, E.C. Meng, G.S. Couch, T.I. Croll, J.H. Morris, and T.E. Ferrin. Ucsf chimerax: Structure visualization for researchers, educators, and developers. *Protein Science*, 30(1):70–82, 2021.

[12] Y. Wang, C. Xu, Y. Wang, Q. Hong, C. Zhang, Z. Li, S. Xu, Q. Zuo, C. Liu, Z. Huang, and Y. Cong. Conformational dynamics of the Beta and Kappa SARS-CoV-2 spike proteins and their complexes with ACE2 receptor revealed by cryo-EM. *Nat. Commun.*, 12(1):7345, 2021.

[13] D. Liebschner, P.V. Afonine, M.L. Baker, G. Bunkóczi, V.B. Chen, T.I. Croll, B. Hintze, L.W. Hung, S. Jain, A.J. McCoy, N.W. Moriarty, R.D. Oeffner, B.K. Poon, M.G. Prisant, R.J. Read, J.S. Richardson, D.C. Richardson, M.D. Sammito, O.V. Sobolev, D.H. Stockwell, T.C. Terwilliger, A.G. Urzhumtsev, L.L. Videau, C.J. Williams, and P.D. Adams. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix *Acta Cryst. D*, 75:861–877, 2019.

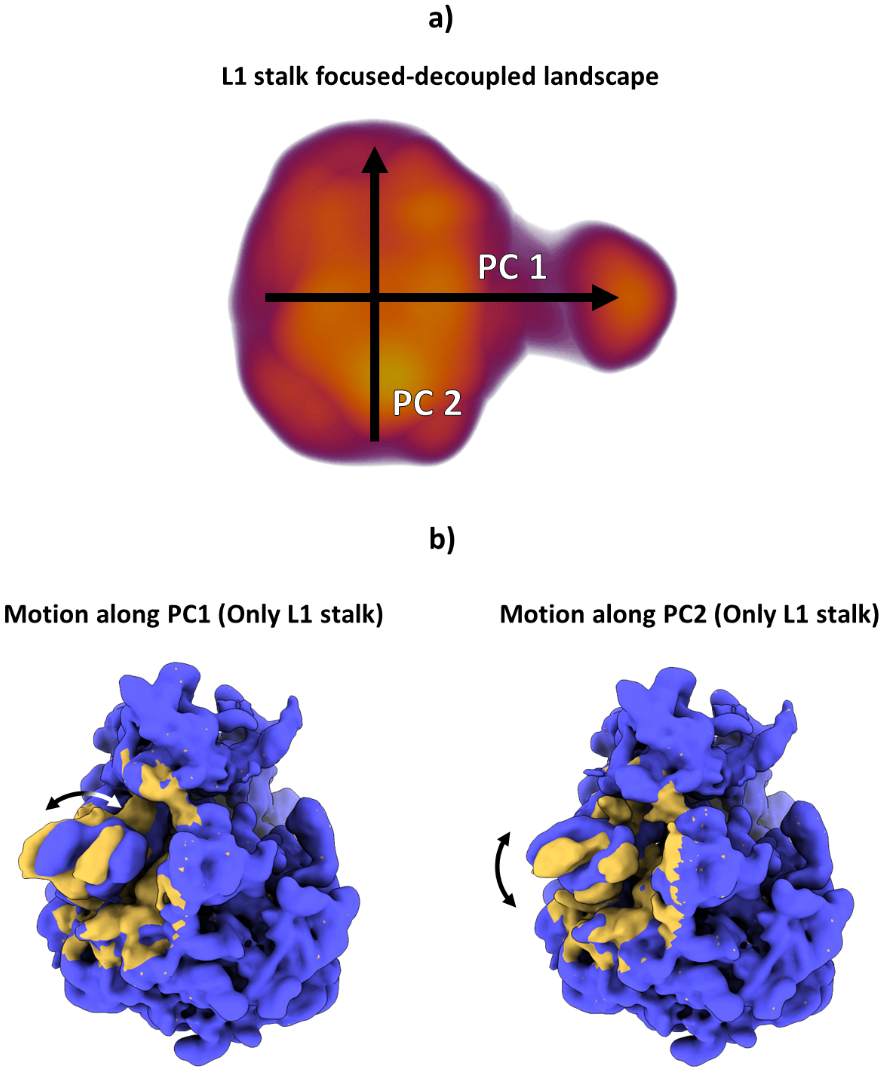
[14] P. Emsley and K. Cowtan, Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.*, vol. 60, no. Pt 12 Pt 1, pp. 2126–2132, 2004.

[15] V. B. Chen, W. B. Arendall III, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson, and D. C. Richardson. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.*, 66(Pt 1):12–21, 2010.

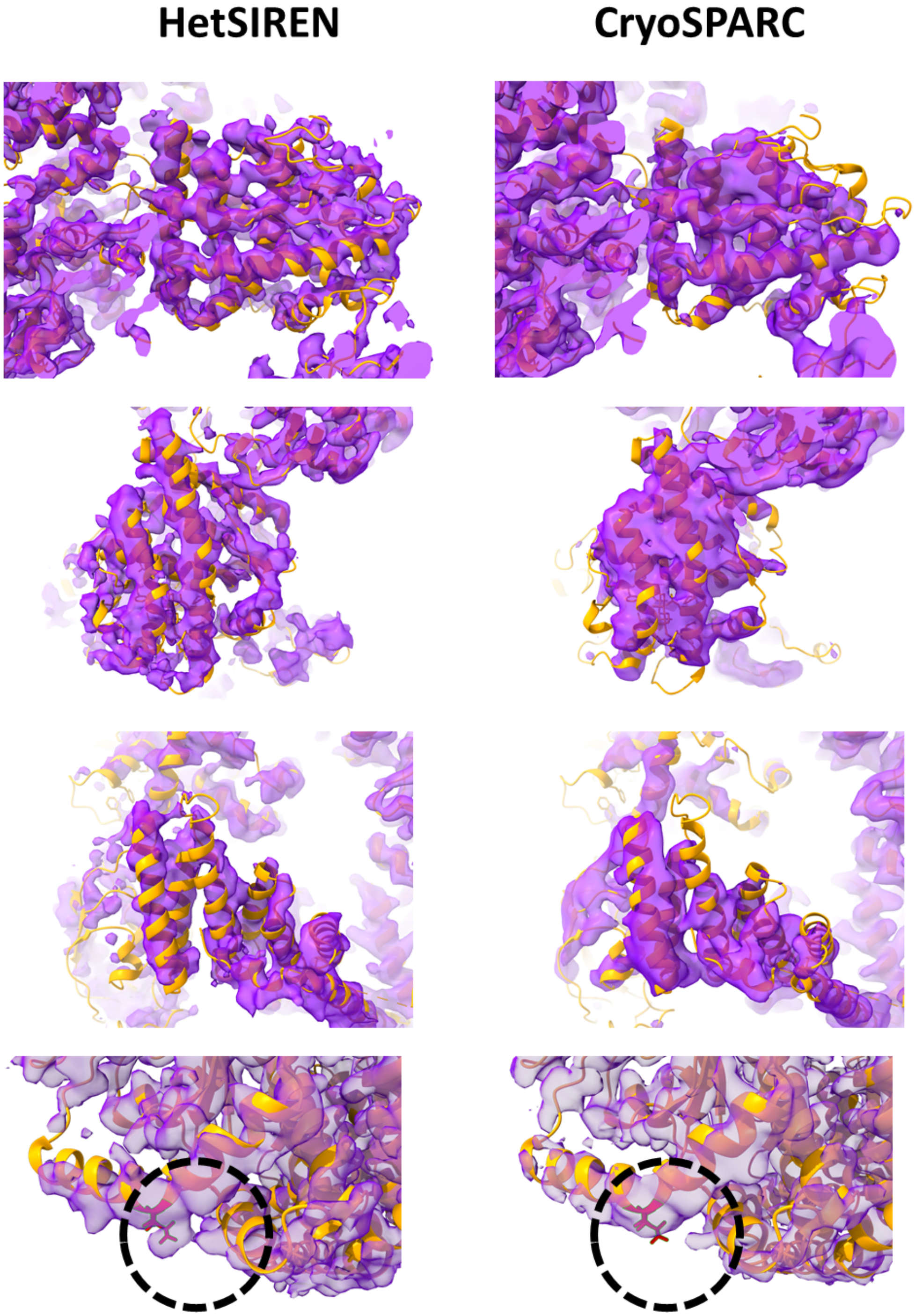
[16] C. J. Williams, J. J. Headd, N. W. Moriarty, M. G. Prisant, L. L. Videau, L. N. Deis, V. Verma, D. A. Keedy, B. J. Hintze, V. B. Chen, S. Jain, S. M. Lewis, W. B. Arendall III, J. Snoeyink, P. D. Adams, S. C. Lovell, J. S. Richardson, and D. C. Richardson. MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci.*, 27(1):293–315, 2018.

[17] R. Sanchez-Garcia, J. Gomez-Blanco, A. Cuervo, J.M. Carazo, C.O.S Sorzano, and J. Vargas. DeepEMhancer: a deep learning solution for cryo-EM volume post-processing. *Communications Biology*, 4:874, 2021.

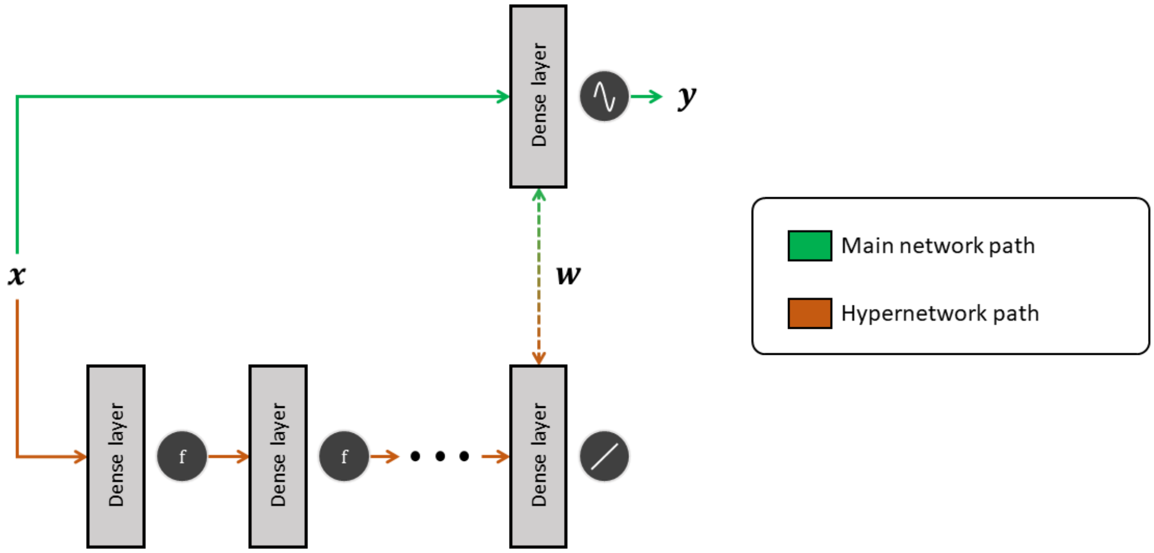
[18] He J., Li T., and Huang S.Y. Improvement of cryo-EM maps by simultaneous local and non-local deep learning. *Nat. Commun.*, 14:3217, 2023.


a)
L1 stalk focused-decoupled landscape
PC 1
PC 2
b)
Motion along PC1 (Only L1 stalk)
Motion along PC2 (Only L1 stalk)

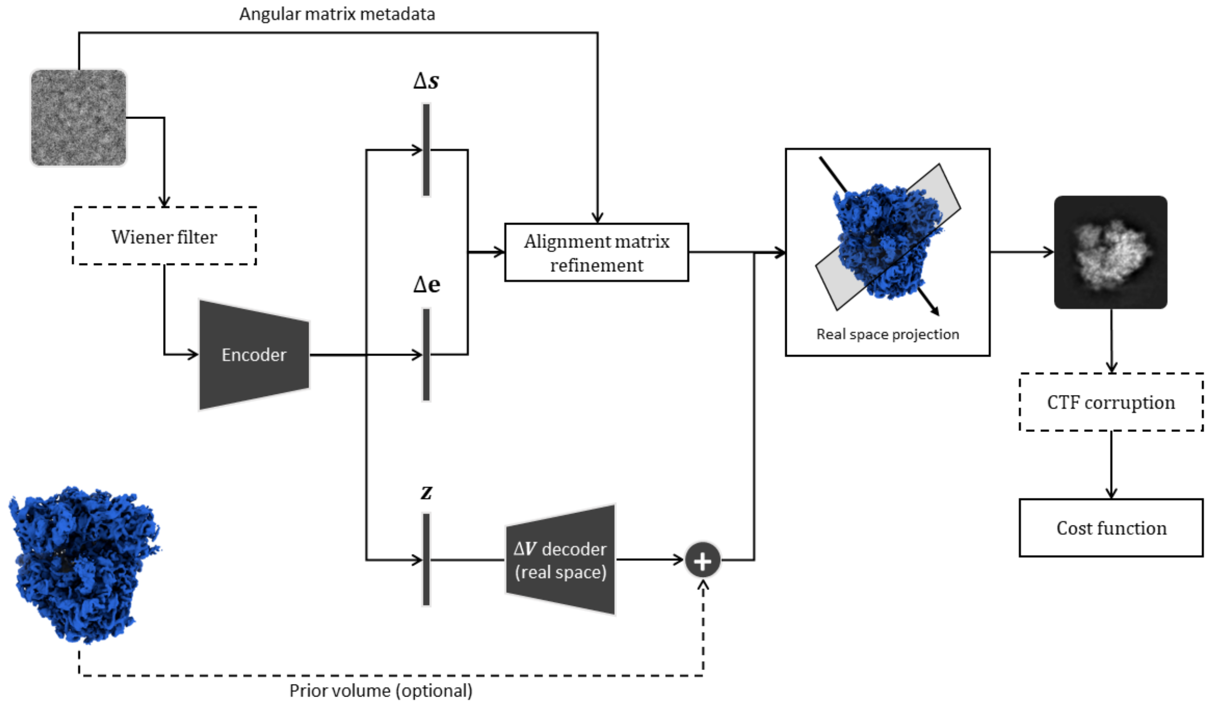
Supplementary Figure 1: Example of the L1 stalked-focused landscape estimated with HetSIREN. The landscape was estimated with the pose and CTF decoupling architecture by providing a spherical mask to the network enclosing the L1 stalk. This way, HetSIREN will only consider the L1 stalk region when determining the motions and conformational changes captured in the experimental particle images. Panel a) shows the UMAP (3) representation of the conformational latent space, including the approximate principal direction according to PCA (5). Panel b) shows the main L1 stalk motions detected by HetSIREN when sampling along the conformational latent space's first and second principal components. The motion detected shows a strong lateral and vertical displacement of the L1 stalk, which is much more easily identified here than when considering the whole particle, thanks to the ability to focus the landscape in this specific region.



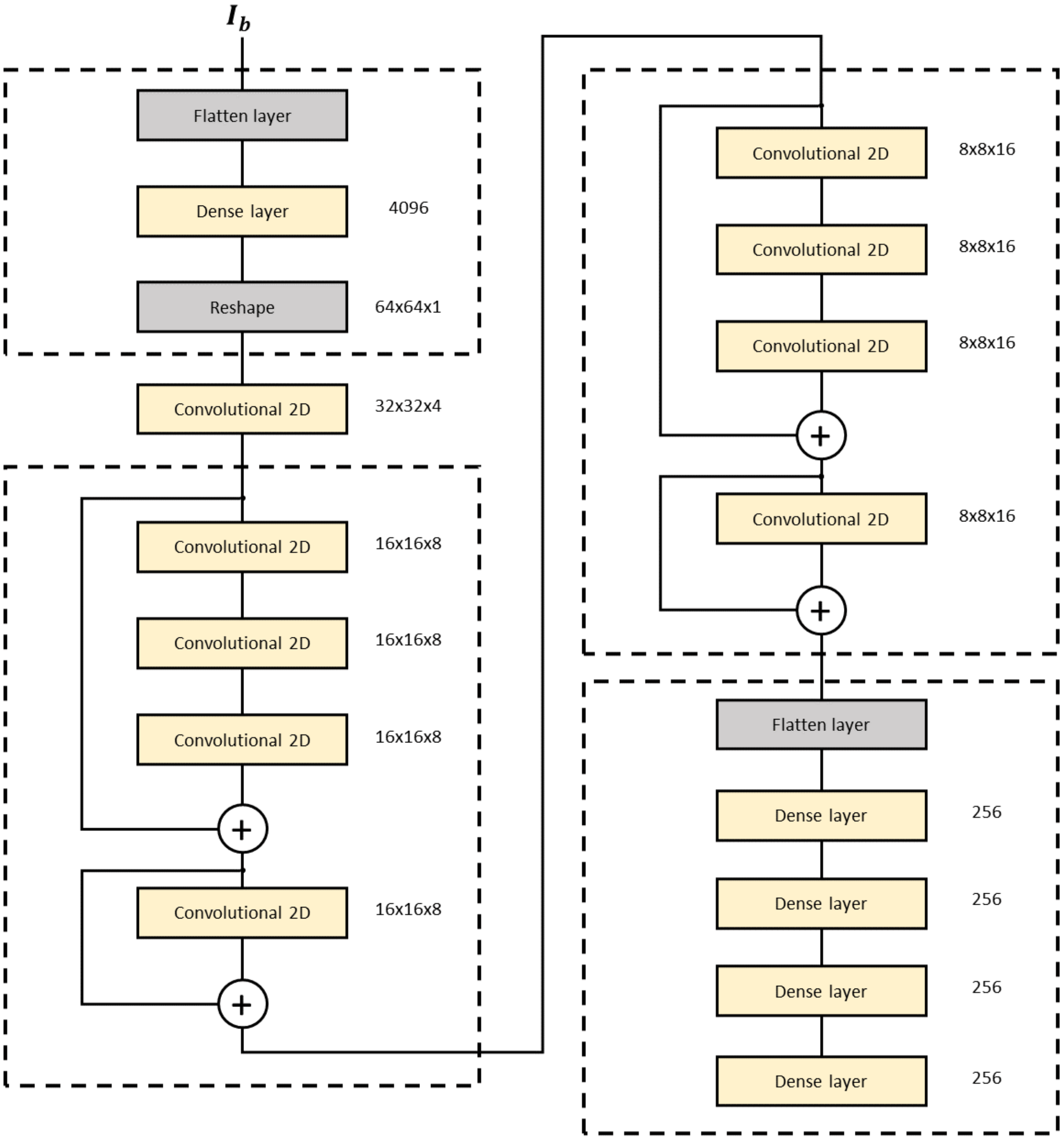
Supplementary Figure 2: Detailed comparison of HetSIREN and the deposited map from (4). The different panels present several zoom regions of the two volumes to better compare the resolution changes between HetSIREN and the deposited map. In addition, we highlight in the last row how HetSIREN has the ability to detect small structural details like side chains in the decoded volumes.



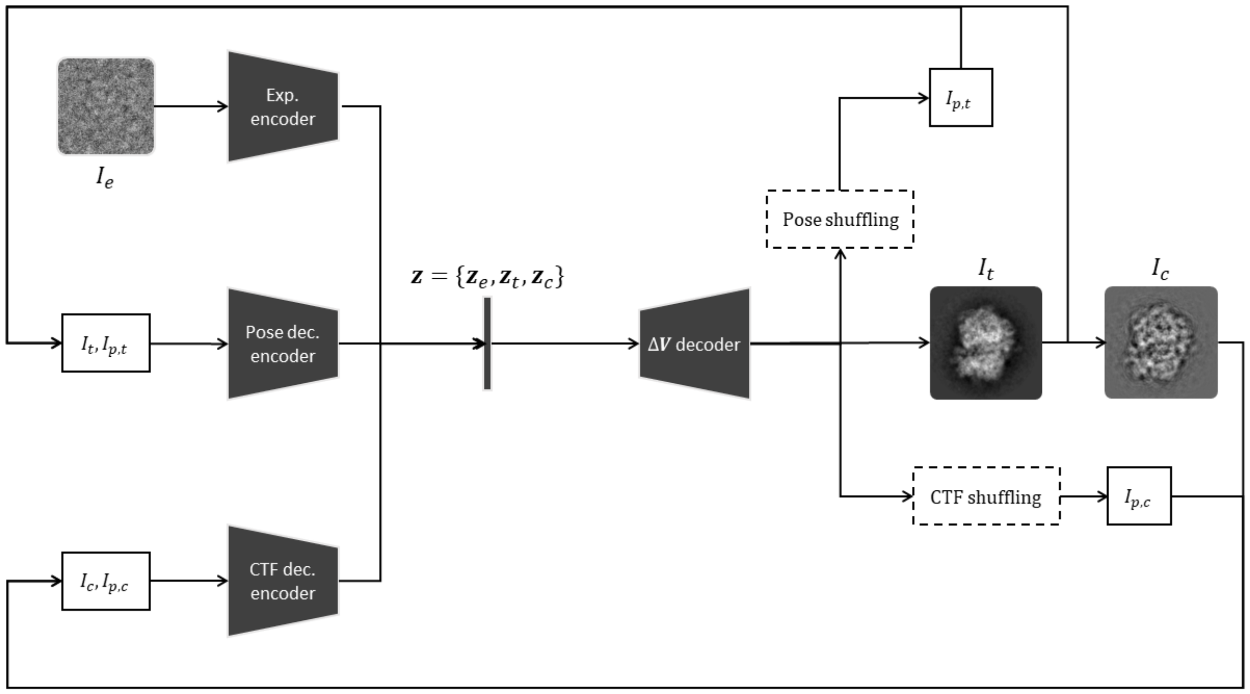
Supplementary Figure 3: Scheme of a meta-sinusoidal layer as implemented in the HetSIREN volume decoder network. The proposed architecture relies on a fully connected network with several layers (hypernetwork) whose weights will be updated during the backpropagation phase. The weights of the last layer in the fully connected network are then shared with the dense layer with the sine activation so that it can decode the appropriate outputs.



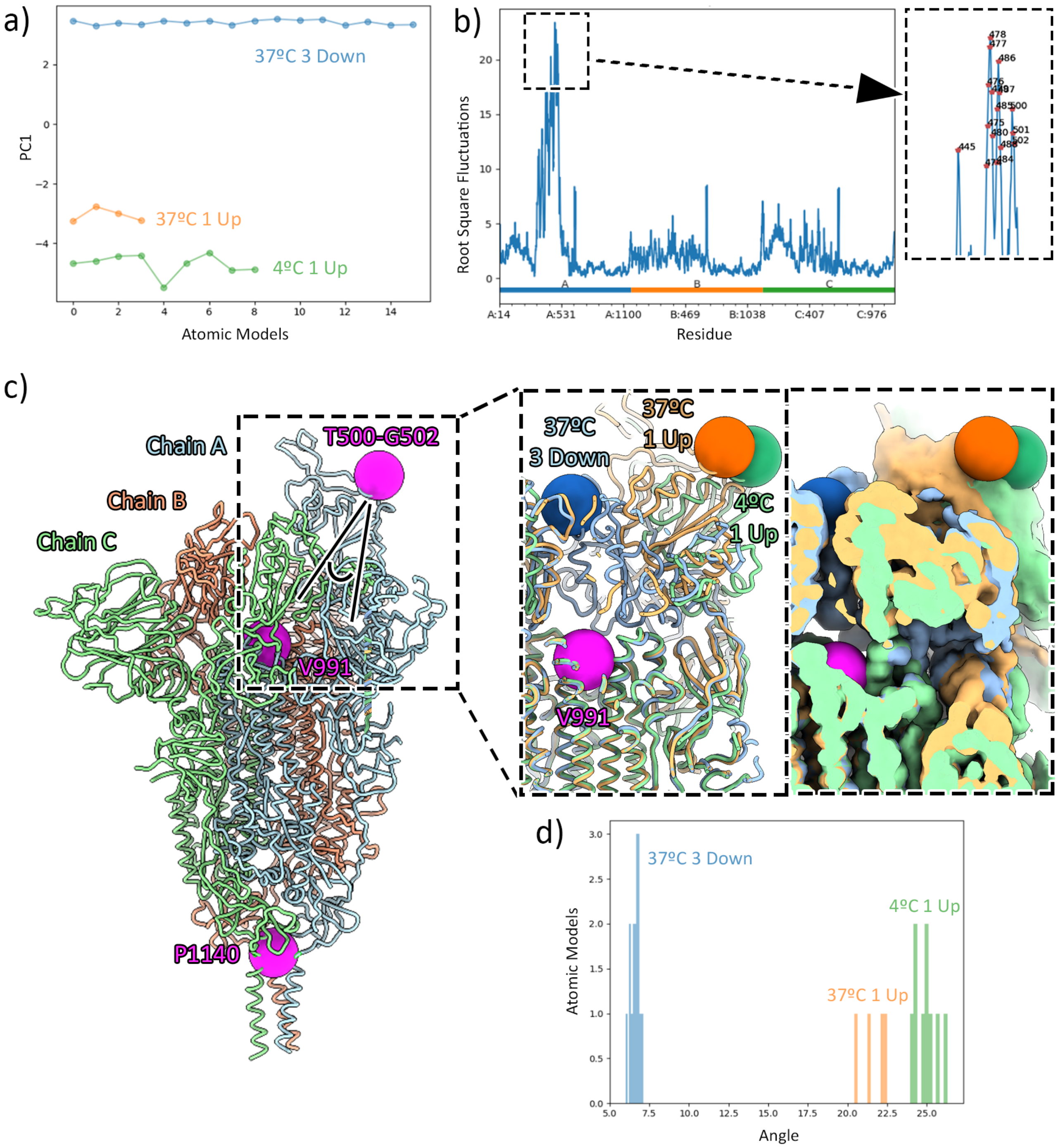
Supplementary Figure 4: Scheme of the HetSIREN network architecture and training strategy. In the scheme, the encoder has a dynamic architecture based on the user inputs (available choices include fully connected and convolutional architectures). The $\Delta\mathbf{V}$ decoder directly produces a full 3D volume in real space from the encoded latent space vectors $\mathbf{z}$. Depending on the availability of the prior volume, the decoded $\Delta\mathbf{V}$ could translate into a full reconstruction (without the prior volume) or a refinement. In addition to the conformational latent space $\mathbf{z}$, two additional bottleneck layers are estimated: a $\Delta\boldsymbol{s}$ layer to refine the in-plane shift of the particle and a $\Delta\boldsymbol{e}$ layer to refine the particle projection angle. The previous two vectors are combined to refine the estimated alignment matrices associated with the experimental image. Regarding the CTF, three possible scenarios are considered: particles have been previously corrected (no CTF considered inside the network), particles are CTF corrected before being fed to the encoder (Wiener filter box), or theoretical projections are CTF corrupted (CTF corruption box).



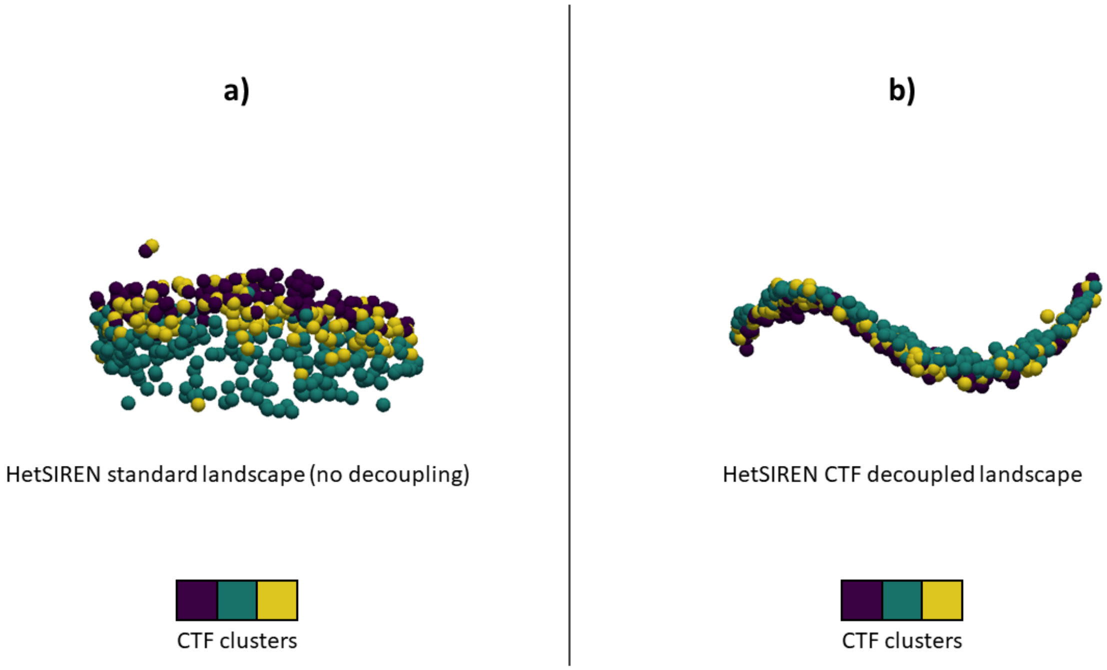
Supplementary Figure 5: Example of the default encoder architecture implemented in HetSIREN. The encoder relies on a resizing network followed by convolutional blocks with residual skips. The output images from the residual blocks are then passed to a fully connected block whose output is posteriorly converted into the three bottlenecks defined in HetSIREN.



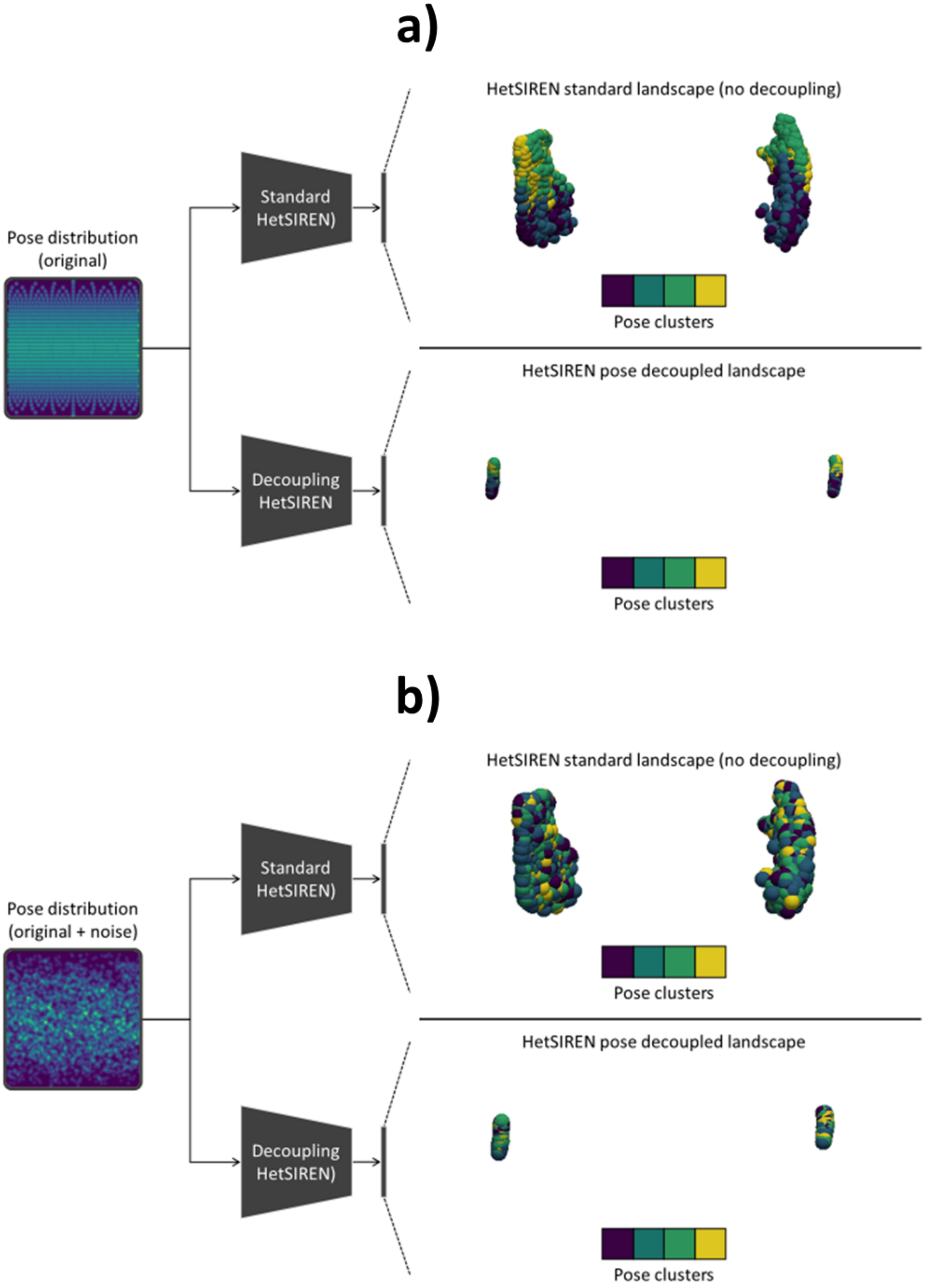
Supplementary Figure 6: HetSIREN poses and CTF decoupling architecture. The decoupling process starts with a batch of experimental images forwarded through the experimental encoder and the decoder to generate a batch of clean, and CTF corrupted projections I_t and I_c. In addition, the original poses and CTFs are shuffled to generate a new set of clean projections with the same conformation but variable pose and CTF $I_{p,t}$ and $I_{p,c}$. Once all the projections have been generated, the images I_t and $I_{p,t}$ are forwarded through the pose decoupling decoder. Similarly, the images I_c and $I_{p,c}$ are fed to the CTF decoupling decoder. In this way, it is possible to generate several sets of latent space vectors representing the same conformational state but with variable pose and CTF, which can be used to decouple the pose and CTF effects from the latent space as expressed in Equations **??** and **??**.



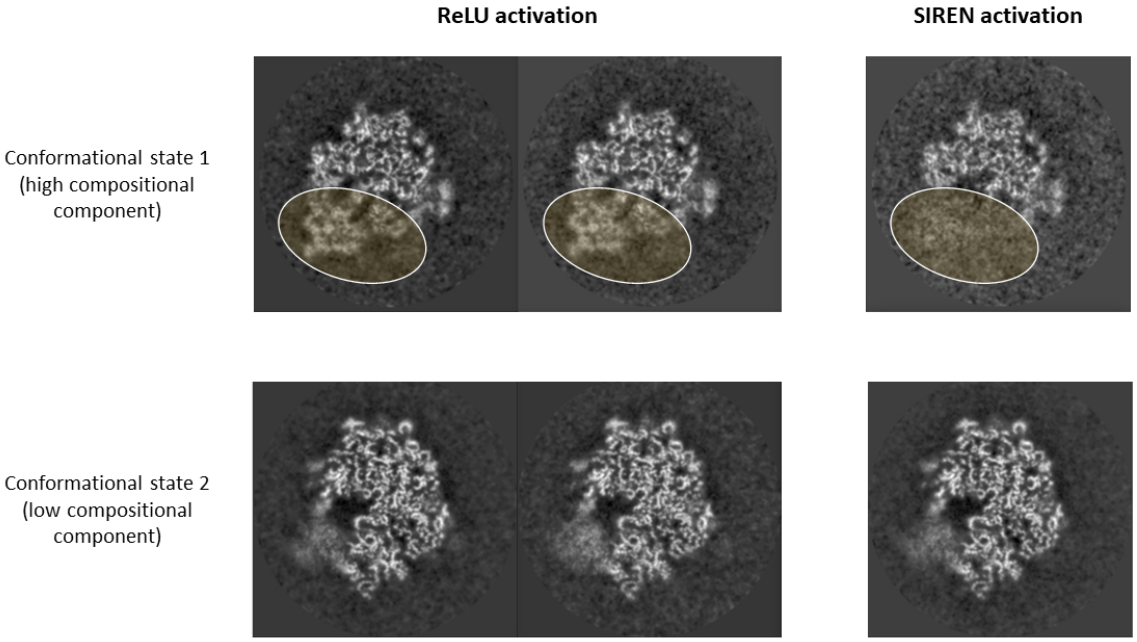
Supplementary Figure 7: Structural analyses of the SARS-CoV-2 Spike protein. Panel a) shows the PCA (5) for the atomic structures ensemble encompassing 16 models of 3 Down conformation at 37°C (blue), 4 models of 1 Up conformation at 37°C (orange), and 9 models of 1 Up conformation at 4°C (green). The Root Mean Square Fluctuations derived from PCA (5) for individual residues are shown in panel b). The inset shows a zoomed area of the residues exhibiting the highest mobility. Panel c) shows a representative atomic model for the 1 Up conformation (chain A in light blue, chain B in light red, and chain C in light green). The three magenta spheres represent the centroids used for the analyses of angle measurements (Thr500-Gly502 at the RBD and Val991 and Pro1140 at the top and bottom of the S2 domain, respectively). Angle is indicated in the dashed box. Insets show a detail of the differences between the three analyzed conformations (3 Down at 37°C in light blue, 1 Up at 37°C in light orange, and 1 Up at 4°C in light green) at atomic models (left) and cryoEM reconstructions (right) levels. Analyses of angle measurements are shown in panel d), matching the color code of inset c). Source data are provided as a Source Data file.



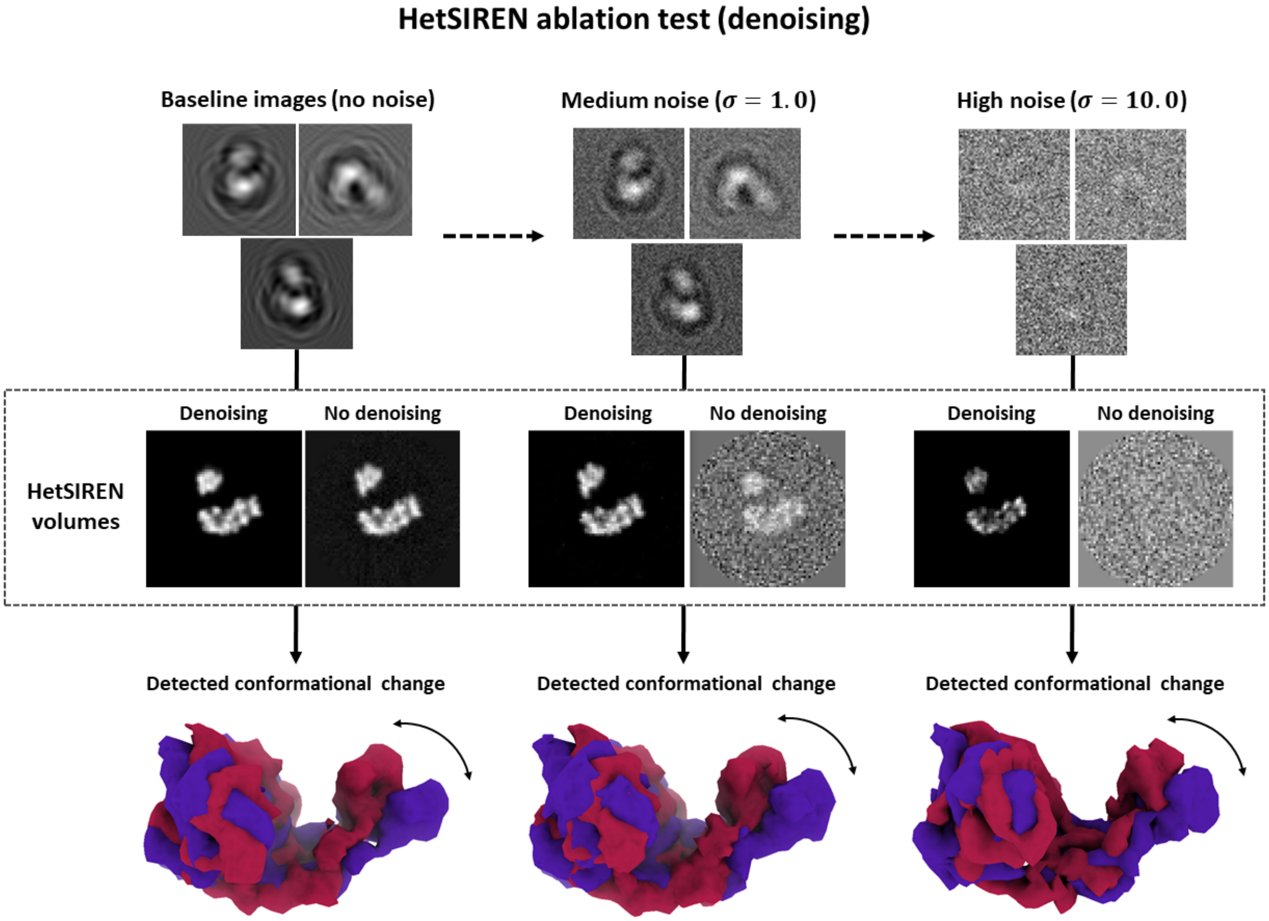
Supplementary Figure 8: Assessment of the CTF decoupling architecture on the latent space learned by HetSIREN. Panels a) and b) show two latent spaces obtained by training two different networks with images with variable CTF corruption. Panel a) shows the landscape encoded by the network with no decoupling architecture. Panel b) shows the landscape encoded by the network, including only the CTF decoupling part. The colors used to represent the landscapes correspond to a clustering of the CTF of the images into three different groups to simplify the visualization of this information. The comparison of the two panels shows how the decoupling effect effectively condenses the latent space, reducing the spreading induced by the strong organization of the latent space according to the CTF of the images.



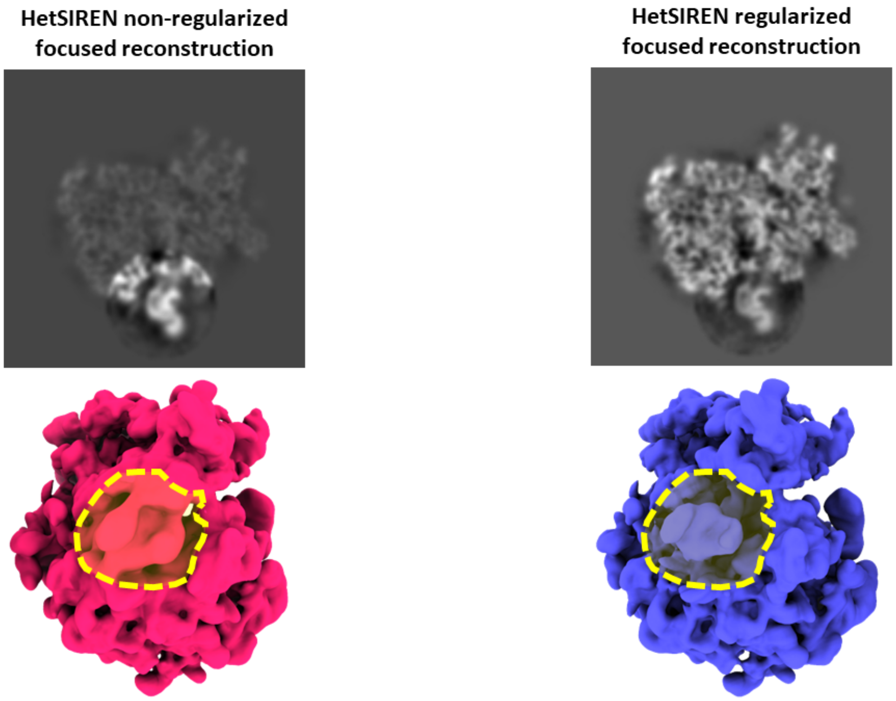
Supplementary Figure 9: Assessment of the pose decoupling architecture on the latent space learned by HetSIREN. The panels show the latent spaces obtained by training two different networks with images with variable poses and no CTF corruption. Panel a) shows the landscapes obtained with the training dataset with a uniform pose distribution. Panel b) shows the landscapes obtained after predicting from the training dataset after adding noise to the original poses. The colors used to represent the landscapes correspond to a clustering of the pose of the images into four different groups to simplify the visualization of this information. The comparison of the two panels shows how the decoupling effect effectively condenses the latent space, reducing the spreading induced by the strong organization of the latent space according to the pose of the images.



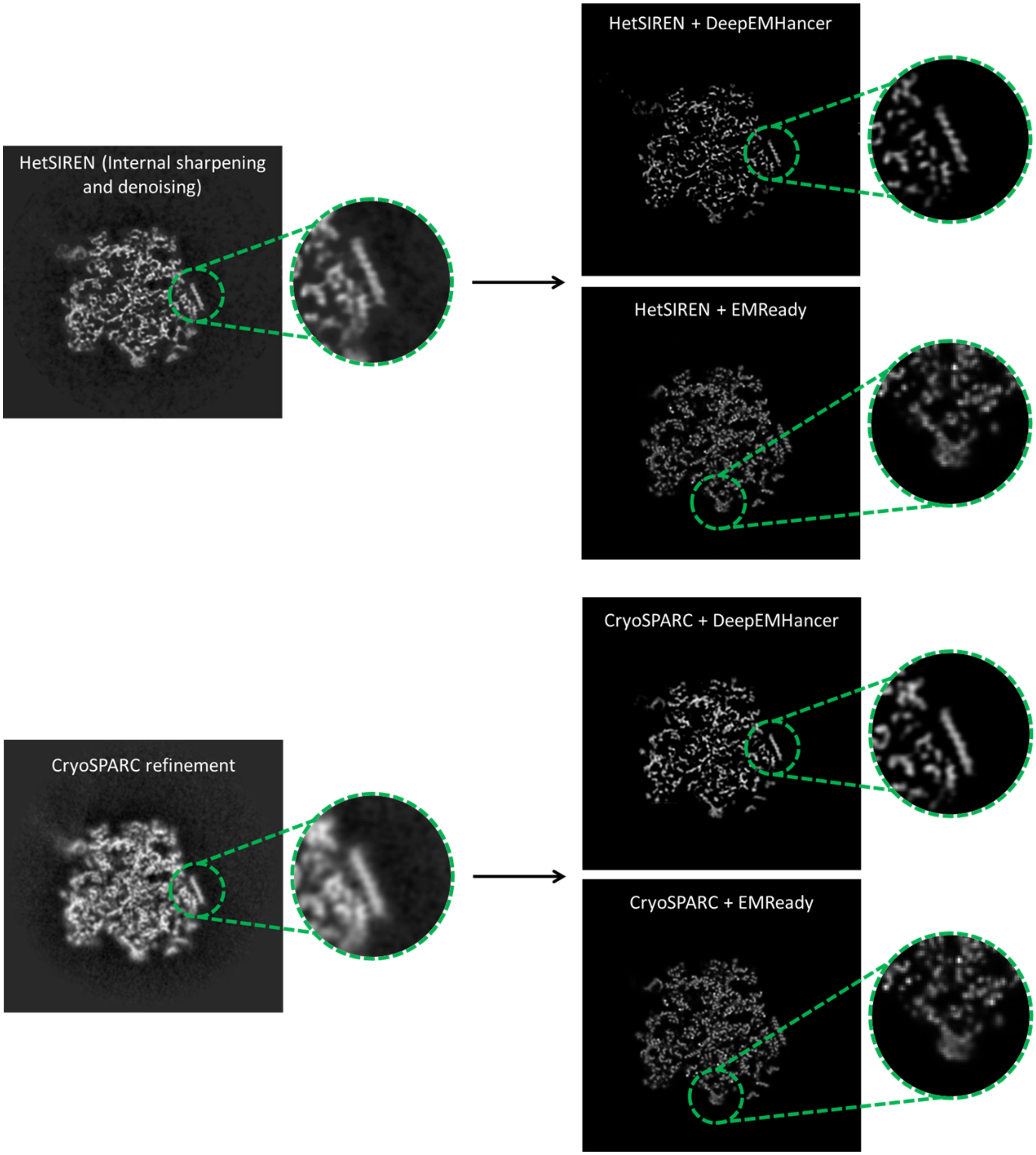
Supplementary Figure 10: Comparison of the decoded accuracy of HetSIREN when trained using two different activation functions in the decoded: ReLU and SIREN. The comparison shows that both activations have a similar performance in representing the structural details in a given state, although SIREN gives more freedom to the network to represent strong compositional variations, as highlighted in the upper images.


HetSIREN ablation test (denoising)
Baseline images (no noise)
Medium noise ($\sigma = 1.0$)
High noise ($\sigma = 10.0$)
HetSIREN volumes
Denoising
No denoising
Denoising
No denoising
Denoising
No denoising
Detected conformational change
Detected conformational change
Detected conformational change

Supplementary Figure 11: Ablation test to analyze the performance of the denoising cost functions implemented during the training phase of HetSIREN. The test evaluates the denoising capabilities of the network under different noise conditions: a set of ideal images, images with medium noise ($\sigma = 1$), and high noise ($\sigma = 10$). In all cases, two different networks were trained, whose only difference is the presence of the denoising cost functions in one of them. The 3D volumes shown are decoded with the denoising network in all cases.


HetSIREN non-regularized focused reconstruction
HetSIREN regularized focused reconstruction

Supplementary Figure 12: Evaluation of the focused reconstruction-related cost functions implemented in HetSIREN. The test evaluates the effect of adding the cost functions responsible for ensuring that the values in chimera volume follow a similar distribution. When this regularization is not applied, the decoded volume shows a clear artifact arising from a strong difference in the value distribution of the refined region and the rest of the volume. In contrast, the regularized network properly minimizes the previous artifact, yielding a more consistent volume.



Supplementary Figure 13: Comparison of HetSIREN and CryoSPARC reconstruction for the EMPIAR 10028 (2) dataset. The comparison shows first the original volumes obtained by both approaches. In the case of HetSIREN, the decoded volume includes the internal sharpening applied during the decoding step, as described in the manuscript. In addition, the figure shows the previous two volumes further post-processed by DeepEmhancer (17) and EMReady (18) to further enhance their structural features. This comparison reveals that the internal sharpening implemented in HetSIREN does not prevent further modification of the decoded volume, yielding a new representation with significantly enhanced structural features compared to both the CryoSPARC maps and its sharpened representation.

Performance metrics for HetSIREN				
Image size	Batch size	Epochs	GPU memory (GB)	Time 10^5 particles (hours)
128	16	50 (standard)	2.42	2.75
300	8	50 (standard)	17.3	14.2
300	8	50 (disentanglement)	17.3	14.5

Supplementary Table 1: Execution times and GPU memory consumption for HetSIREN. Metrics are referred to in the training phase.

Automatically modeled residues (ModelAngelo)	
CryoSPARC Map **Modelled Residues** **% Total Residues**	**RBD (residues 304-591)** 310 35.9%
HetSIREN map 4	365 42.2%
HetSIREN map 9	342 39.6%
HetSIREN map 13	401 46.4%
HetSIREN map 14	326 37.7%
HetSIREN map 15	346 40.0%

Supplementary Table 2: Comparison of automatically modeled residues performed by ModelAngelo.

Refinement statistics for SARS-CoV-2 Spike protein at 4°C

Refinement	**Map 1**	**Map 2**	**Map 3**	**Map 4**	**Map 5**	**Map 6**	**Map 7**	**Map 8**	**Map 9**	**Map 10**
Mask correlation coefficient	0.69	0.69	0.70	0.71	0.64	0.69	0.70	0.72	0.72	0.69
Model composition										
Non-hydrogen atoms	25,362	25,362	25,362	25,362	25,362	25,362	25,362	25,362	25,362	25,362
Protein residues	3,237	3,237	3,237	3,237	3,237	3,237	3,237	3,237	3,237	3,237
ADP (B-factors)										
min	56.88	65.32	54.68	67.54	47.65	52.98	59.54	50.81	63.09	62.66
max	428.31	325.45	462.43	318.58	363.74	620.24	539.35	297.54	358.72	455.62
mean	149.62	132.17	156.76	138.32	155.21	157.98	161.59	136.61	131.11	156.43
R.m.s deviations										
Bond lengths	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.009	0.008
Bond angles	1.428	1.490	1.442	1.500	1.367	1.384	1.446	1.484	1.597	1.455
Validation										
Molprobity score	1.65	1.77	1.62	1.81	1.53	1.60	1.58	1.64	1.79	1.69
Clashscore	5.91	6.35	5.95	6.31	4.97	5.77	5.55	6.49	7.64	6.99
Rotamer outliers (%)	1.17	0.74	1.06	0.99	0.78	0.67	0.88	0.99	1.24	0.60
Ramachandran plot										
Favoured (%)	95.97	93.05	95.94	95.69	96.03	95.91	96.00	95.91	95.72	95.66
Allowed (%)	3.97	6.95	4.00	4.25	3.91	4.03	3.94	4.03	4.22	4.28
Outlier (%)	0.06	0.00	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06

Refinement statistics for SARS-CoV-2 Spike protein at 4°C

Refinement	**Map 11**	**Map 12**	**Map 13**	**Map 14**	**Map 15**	**Map 16**	**Map 17**	**Map 18**	**Map 19**	**Map 20**
Mask correlation coefficient	0.68	0.70	0.71	0.72	0.70	0.67	0.69	0.69	0.69	0.71
Model composition										
Non-hydrogen atoms	25,362	25,362	25,362	25,362	25,362	25,362	25,362	25,362	25,362	25,362
Protein residues	3,237	3,237	3,237	3,237	3,237	3,237	3,237	3,237	3,237	3,237
ADP (B-factors)										
min	54.78	55.29	55.46	58.05	51.09	54.05	55.09	58.39	49.91	51.77
max	579.72	549.59	440.90	259.33	375.42	470.42	454.68	401.38	460.43	657.55
mean	156.63	171.29	150.96	127.42	140.63	155.53	153.43	144.97	159.05	153.25
R.m.s deviations										
Bond lengths	0.008	0.007	0.008	0.008	0.008	0.007	0.007	0.007	0.007	0.007
Bond angles	1.478	1.441	1.608	1.532	1.473	1.402	1.465	1.430	1.433	1.515
Validation										
Molprobity score	1.72	1.58	1.65	1.67	1.61	1.58	1.61	1.69	1.58	1.66
Clashscore	6.17	5.47	6.31	6.73	5.89	5.63	6.05	6.93	5.29	6.47
Rotamer outliers (%)	1.41	0.78	0.81	0.92	0.99	0.67	0.81	1.09	0.78	0.81
Ramachandran plot										
Favoured (%)	96.03	95.91	95.66	95.66	95.78	95.97	95.97	95.91	95.72	95.69
Allowed (%)	3.91	4.03	4.25	4.28	4.15	3.97	3.97	4.00	4.22	4.25
Outlier (%)	0.06	0.06	0.09	0.06	0.06	0.06	0.06	0.09	0.06	0.06

Supplementary Table 3: Refinement statistics for SARS-CoV-2 Spike protein at 4°C.

Refinement statistics for SARS-CoV-2 Spike protein at 37°C										
Refinement	**Map 1**	**Map 2**	**Map 3**	**Map 4**	**Map 5**	**Map 6**	**Map 7**	**Map 8**	**Map 9**	**Map 10**
Mask correlation coefficient	0.68	0.69	0.74	0.77	0.73	0.75	0.76	0.74	0.69	0.73
Model composition										
Non-hydrogen atoms	25,362	25,362	25,482	25,482	25,482	25,482	25,482	25,482	25,362	25,482
Protein residues	3,237	3,237	3,255	3,255	3,255	3,255	3,255	3,255	3,237	3,255
ADP (B-factors)										
min	66.33	69.60	73.86	84.37	69.29	71.30	68.22		73.47	63.30
max	469.35	418.14	418.14	331.04	299.72	277.98	319.64	261.50	412.59	336.72
mean	151.83	149.02	197.15	128.96	134.03	124.63	127.39	124.20	151.80	135.49
R.m.s deviations										
Bond lengths	0.008	0.008	0.008	0.009	0.009	0.008	0.009	0.009	0.007	0.008
Bond angles	1.603	1.612	1.717	1.840	1.829	1.745	1.777	1.797	1.566	1.738
Validation										
Molprobity score	1.69	1.65	1.61	1.80	1.62	1.66	1.59	1.63	1.68	1.63
Clashscore	6.63	6.21	6.06	7.53	6.99	6.26	6.36	6.49	6.69	5.96
Rotamer outliers (%)	1.02	0.88	1.16	1.72	0.88	1.33	1.05	1.19	0.92	1.33
Ramachandran plot										
Favoured (%)	95.44	95.50	96.49	96.71	96.49	96.58	96.52	96.61	95.60	96.71
Allowed (%)	4.47	4.40	3.42	3.20	3.42	3.32	3.39	3.29	4.31	3.20
Outlier (%)	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09

Refinement statistics for SARS-CoV-2 Spike protein at 37°C										
Refinement	**Map 11**	**Map 12**	**Map 13**	**Map 14**	**Map 15**	**Map 16**	**Map 17**	**Map 18**	**Map 19**	**Map 20**
Mask correlation coefficient	0.67	0.74	0.75	0.76	0.74	0.74	0.74	0.74	0.75	0.74
Model composition										
Non-hydrogen atoms	25,362	25,482	25,482	25,482	25,482	25,482	25,482	25,482	25,482	25,482
Protein residues	3,237	3,255	3,255	3,255	3,255	3,255	3,255	3,255	3,255	3,255
ADP (B-factors)										
min	67.35	73.33	75.37	80.29	70.45	66.29	71.37	86.52	76.29	78.77
max	450.55	359.03	350.66	288.60	290.02	330.95	288.01	405.81	421.83	304.79
mean	152.38	130.67	127.95	128.29	127.89	129.61	125.55	131.54	262.12	125.21
R.m.s deviations										
Bond lengths	0.007	0.009	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008
Bond angles	1.549	1.793	1.751	1.739	1.740	1.790	1.812	1.720	1.799	1.746
Validation										
Molprobity score	1.70	1.66	1.60	1.58	1.65	1.59	1.69	1.59	1.61	1.57
Clashscore	5.69	6.36	6.57	6.16	6.32	6.16	6.87	5.84	6.45	5.44
Rotamer outliers (%)	0.85	1.37	1.09	1.09	1.30	1.12	1.30	1.19	1.19	1.23
Ramachandran plot										
Favoured (%)	95.56	96.74	96.64	96.68	96.64	96.64	96.55	96.64	96.77	96.71
Allowed (%)	4.34	3.17	3.26	3.23	3.26	3.26	3.29	3.26	3.11	3.20
Outlier (%)	0.09	0.09	0.09	0.09	0.09	0.09	0.16	0.09	0.12	0.09

Supplementary Table 4: Refinement statistics for SARS-CoV-2 Spike protein at 37°C.