Interactive segmentation of membrane and membrane mimic densities in cryo-EM maps

Alok Bharadwaj[a], Lotte Veerbeek[a], and Arjen J. Jakobi[a] ✉

[a]Department of Bionanoscience, Kavli Institute of Nanoscience, Delft University of Technology, 2629 HZ Delft, The Netherlands

Contextual density arising from biological membranes or membrane mimics such as detergent micelles or lipid nanodiscs, is ubiquitous in cryo-EM maps of integral and peripheral membrane proteins, or membrane-associated complexes, from single-particle analysis and subtomogram averaging. Here, we describe SURFER, a lightweight, GPU-accelerated extension for UCSF ChimeraX that enables rapid semantic differentiation of membrane or membrane mimic density from macromolecular signal. SURFER operates on unfiltered maps to automatically generate segmentation masks that can be flexibly applied to any aligned target map, including reference volumes for local refinement or post-processed maps optimised for visualisation and interpretation. By allowing contextual density to be selectively toggled or subtracted, SURFER enables direct interactive comparison between representations with and without membrane-related features to facilitate consistent analysis and reporting of membrane context in cryo-EM structures. An interactive graphical interface within ChimeraX supports convenient threshold adjustment, connectivity filtering, and immediate visual evaluation of the resulting segmentation.

Correspondence: a.jakobi@tudelft.nl

1. Introduction

Cryogenic electron microscopy (cryo-EM) and electron tomography (cryo-ET) now routinely yield three-dimensional electrostatic potential maps for membrane proteins and membrane-associated complexes across a wide resolution range[1,2]. Beyond the ordered macromolecular core, many cryo-EM structures of membrane proteins contain contextual density originating from sample preparation or the native environment, including detergent micelles[3,4], amphipols[5], lipid nanodiscs[6,7], membrane bilayers[8], and other partially occupied or conformationally heterogeneous components[2]. Such contextual density is typically weakly ordered and dominated by low resolution signal. Although it can provide important information on the organisation of the membrane environment and the positioning of transmembrane regions, its incoherent signal frequently causes the density to fragment into discontinuous regions that complicate visualisation and segmentation using simple threshold-based approaches.

Map post-processing has a strong influence on how such contextual density is represented. Conventional global sharpening[9] and many learned enhancement approaches[10–13] prioritise recovery of high-resolution contrast in well-ordered regions that often form the bulk of the macromolecular signal. While effective for enhancing macromolecular features, these approaches often further fragment low-resolution density or suppress it altogether[14,15]. In contrast, map post-processing strategies that aim to preserve local signal statistics can retain contextual density at intensity levels comparable to those of the macromolecule[16–18]. Building on these principles, LocScale-2.0 was introduced to improve the interpretability of cryo-EM maps through confidence-guided map optimisation while maintaining a more unbiased representation of both ordered and disordered components[19]. As a result, membrane and membrane mimic density often appears more continuous and interpretable in LocScale-2.0 optimised maps at a common intensity threshold.

While retention of contextual density is advantageous for assessing membrane topology and macromolecular environment, it also introduces challenges during interactive map inspection and visual presentation. In particular, lipid or detergent density may obstruct the view on transmembrane regions or dominate visualisation at commonly used contour levels. When this density is brought onto a comparable intensity scale to the macromolecular core, interpretation and visualisation create a practical need for tools that enable semantic differentiation between contextual membrane signal and ordered macromolecular features. In addition to visual presentation, segmentation of membrane mimic density can be practically useful during 3D refinement. From an image formation perspective, a cryo-EM micrograph represents a tomographic projection of the specimen modulated by the point spread function of the electron microscope[20]. When membrane proteins are embedded in a lipid bilayer, nanodisc or detergent micelle, these components can produce a strong contribution to the projection signal, which in some orientations may exceed that of the embedded protein itself. The strong low-resolution signal of detergent micelles or lipid nanodiscs, where detergent and lipid molecules are expected to adopt unrelated positions away from the immediate environment of the protein, can negatively impact iterative refinement procedures that assume a largely uniform signal, and can lead to overfitting. Although data-driven regularisation strategies such as non-uniform refinement[21] and local signal filtering[22] reduce sensitivity to disordered regions, selective masking and focused refinement workflows remain widely used in practice for restricting alignment and reconstruction to regions of interest or for removing non-target signal by subtraction[23–27].

Segmentation of cryo-EM density maps has traditionally relied on intensity-based thresholding[28], morphological

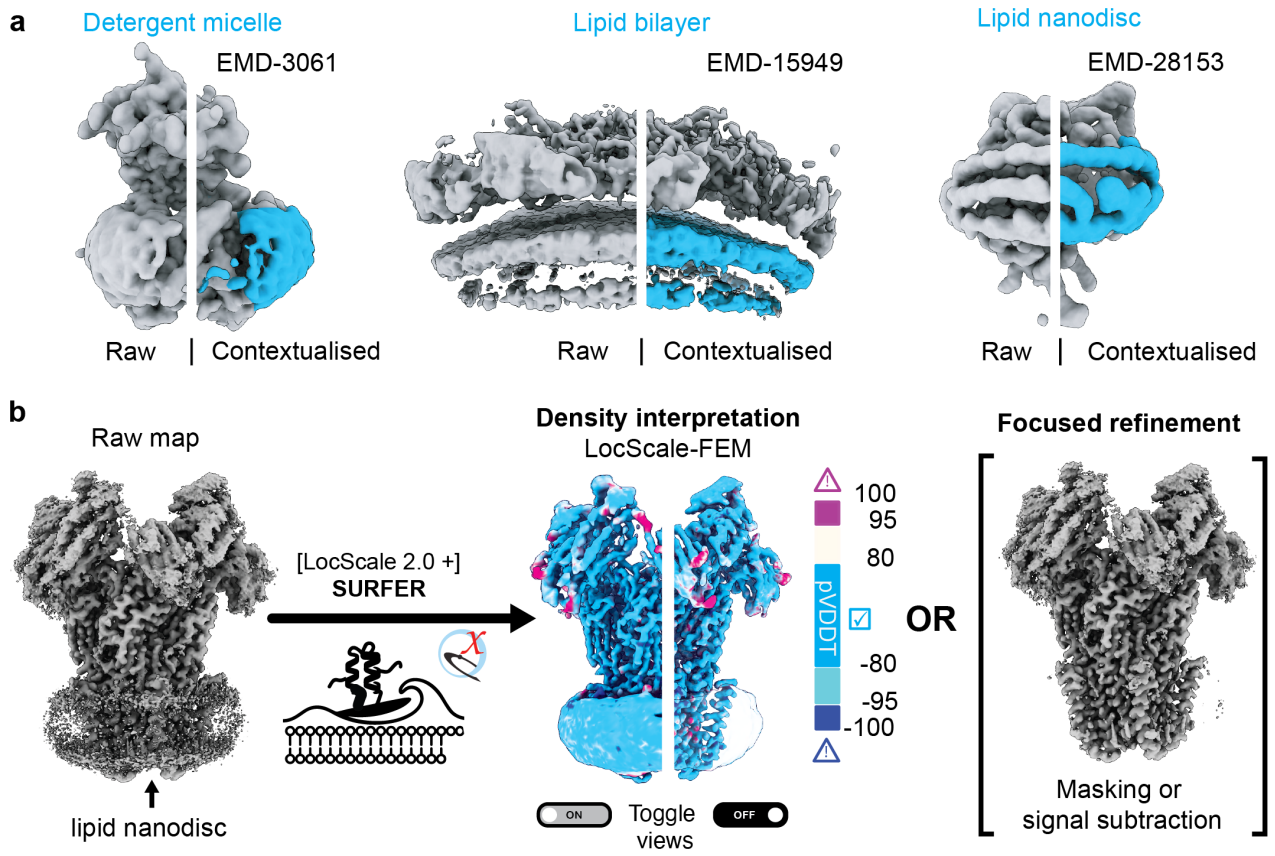

Fig. 1. Contextualisation of lipid densities and applications of SURFER | (a) Raw and contextualised cryo-EM density maps for γ-secretase in amphipol A8-35 (EMDB: EMD-3061), COPII coat on lipid bilayers (EMDB: EMD-15949) and the mechanosensitive channel TMEM63A in lipid nanodiscs (EMDB: EMD-28153). (b) Schematic illustration of LocScale-SURFER segmentation and substraction of lipid micelle for visualisation with LocScale 2.0-optimised maps, or for particle subtraction for 3D refinement. The example shown is the pentameric calcium-sensitive channel DeCLIC from *Desulfofustis glycolicus* reconstituted in lipid nanodiscs (EMDB: EMD-19995). The LocScale-FEM pVDDT confidence score is mapped onto the optimised map.

watershed approaches[29,30], and region-based methods that combine connectivity with multi-scale smoothing or statistical merging[31,32]. These approaches form the basis of widely used tools such as Segger and related workflows for partitioning macromolecular assemblies into compact subregions[31]. More recent methods incorporate additional structural constraints, for example graph-based representations and symmetry information[33]. While effective for separating well-defined macromolecular components, these methods are primarily designed to identify regions characterised by relatively sharp intensity gradients or compact connectivity. Lipid and detergent-associated density, by contrast, is typically smooth, weakly bounded, and sometimes spatially contiguous with the macromolecular envelope, making its segmentation sensitive to contour levels and connectivity criteria.

Similar challenges are well recognised in cryogenic electron tomography (cryo-ET) image analysis, where membranes are commonly segmented using dedicated methods developed for low signal-to-noise and diffuse boundaries, including tensor voting[34], deep learning-based voxel classification[35–38] and combinations of deep learning-based segmentation with parametric fitting and interactive curation[39]. Such hybrid strategies are often required to accommodate the wide range of membrane geometries encountered in practice. Comparable variability is also observed in single-particle reconstructions of membrane proteins, where contextual density may arise from membrane mimics such as micelles, bicelles, nanodiscs, or planar bilayers and can differ substantially across datasets. Semantic differentiation of such density therefore benefits from combining prior information on membrane location with data-driven voxel classification.

To aid in the analysis of contextual lipid and detergent densities in single-particle reconstructions and sub-tomogram averages, we have developed SURFER (Segmentation of Unstructured Regions and Filtering for Enhanced Representation), a lightweight framework for machine-learning-based segmentation of lipid and detergent densities. SURFER automatically identifies, segments, and optionally subtracts lipid and detergent densities corresponding to micelles, lipid bilayers, or lipid nanodiscs, thereby allowing users to focus alternately on the macromolecular core or the assembly in its membrane context (Figure 1). While originally designed as a companion to LocScale-2.0 map optimisation, the segmentation mask generation of SURFER operates on unfiltered maps and application of its segmentations mask is therefore compatible with any raw or post-processed map.

SURFER is implemented as a ChimeraX[40] bundle with both graphical user interface and command-line functionality Here, we describe the design and training of the segmentation model, evaluate its performance across a diverse set of cryo-EM density maps from EMDB-deposited membrane protein structures, and illustrate how SURFER enables contextualisation of lipid and membrane mimic densities and robust micelle subtraction in cryo-EM structures.

2. Results

2.1 Overview.

Our research was motivated by a lack of accessible tools specifically optimised for segmentation of membrane or detergent micelle-related densities in three-dimensional (3D) electrostatic potential maps from cryo-EM single-particle analysis and subtomogram averaging. SURFER is a tool designed to fill this gap (Figure 1a). The method uses voxel-wise classification with a hybrid convolutional–transformer architecture to distinguish contextual density from the ordered macromolecular signal. Segmented contextual density can be then be interactively visualised or subtracted from a target map to enable direct comparison between representations that retain or suppress membrane-related features, or for the generation of masks for focused refinement (Figure 1b).

In this section, we first describe the generation of a curated training dataset and the design of the segmentation model. We then evaluate segmentation performance across a diverse set of deposited cryo-EM reconstructions and analyse the geometric variability of lipid and detergent environments captured in the training data. Finally, we illustrate practical use cases of SURFER, including interactive visualisation and subtraction of contextual density, and assess the robustness of the approach under different segmentation thresholds and filtering strategies.

2.2 Training dataset generation and characteristics.

We assembled a labeled dataset of segmented lipid and detergent densities from curated map–model pairs of integral membrane and membrane-associated proteins derived from the Electron Microscopy Data Bank (EMDB)[41], restricted to entries with associated atomic models encompassing the full transmembrane region.

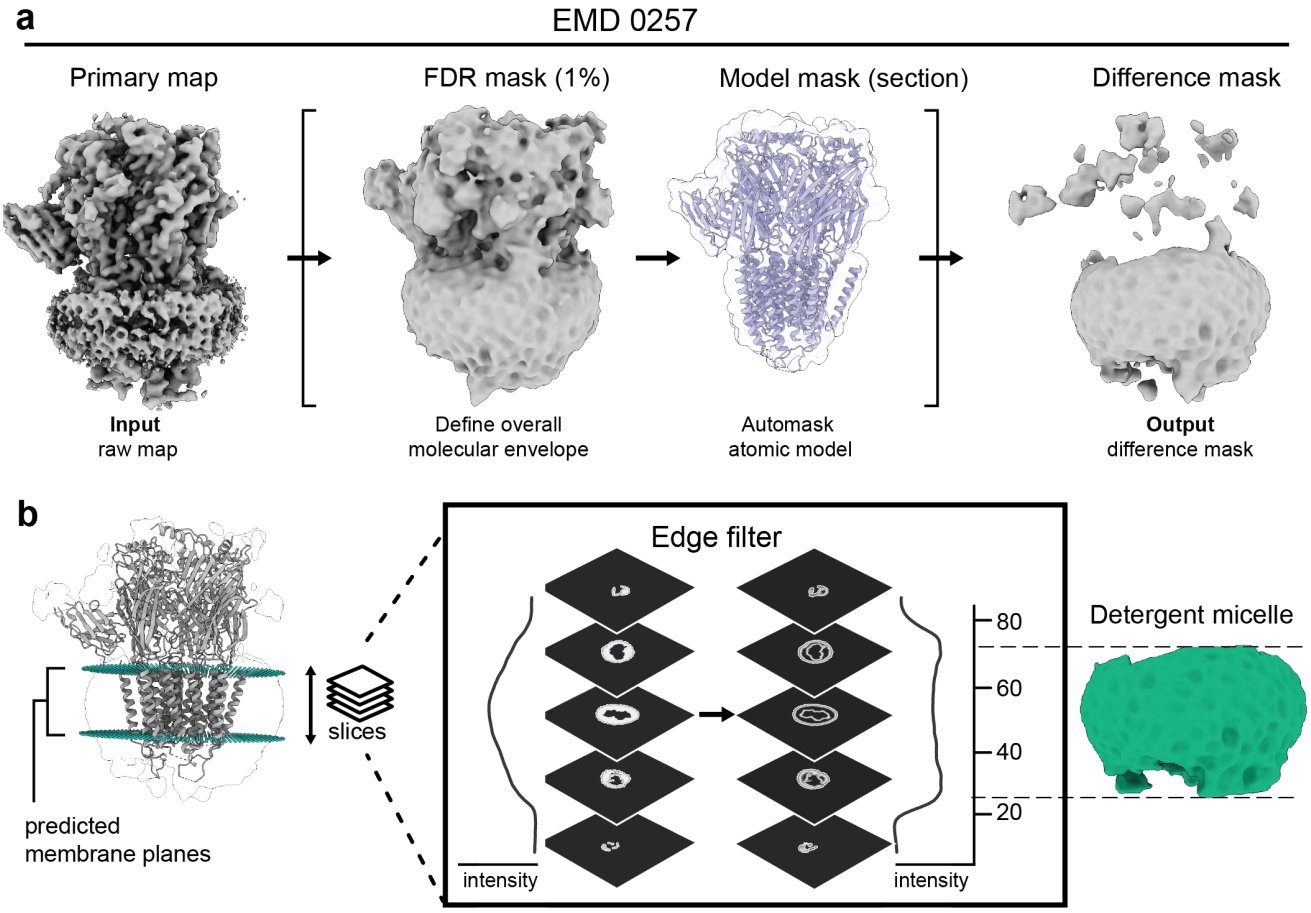

Fig. 2. Training dataset generation for membrane/detergent segmentation | (a) Difference mask generation for isolation of lipid density regions illustrated with the *E. coli* potassium transporter KdpFABC (EMDB: EMD-0257). The left shows the raw average of the unfiltered half maps from which the FDR confidence mask is generated. The ordered volume is estimated from the deposited atomic model (PDB ID: 6hug) and subtracted from the FDR-enclosed volume to generate a difference mask. Note that the raw difference mask is not restricted to the detergent micelle but can encompass other unmodelled signal. (b) Predicted membrane planes relative to the atomic model and the raw difference mask and procedure of edge detection to refine plane positioning at the detergent boundary. The segmented detergent micelle volume is shown on the far right.

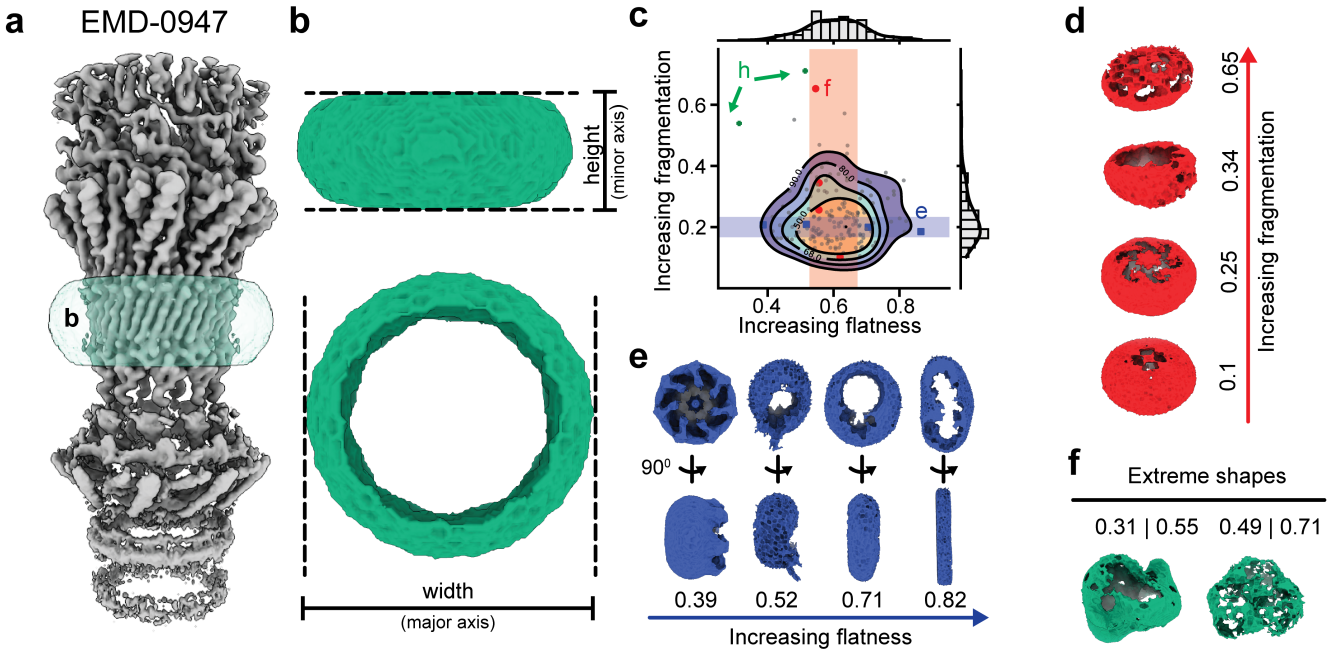

Fig. 3. Geometric variability of membrane mimic density in the training dataset | (a-b) Geometrical properties of segmented micelle volume shown for the enterobacterial curli secretion complex CsgG-CsgF (EMDB: EMD-0947). (c) Distribution of shape features from the training dataset (N=170). Semi-transparent rectangles denote bands of largest variation along fragmentation and flatness ratios respectively. (d-e) Representative micelle densities contained within the high variability bands. (f) Out-of-distribution outliers marked by arrows in (c). Numbers indicate flatness and fragmentation ratio, respectively.

To isolate disordered lipid and detergent density, we generated difference masks by determining the molecular envelope via statistical thresholding with false discovery rate (FDR) control[18,19,42] and subtracting the ordered volume occupied by the atomic coordinate model. The resulting masks contain density not explained by the atomic model, including lipid or detergent-related density together with other unmodelled structural features (Figure 2a). We then further limited this volume to membrane or membrane mimic density by predicting the planes circumscribing micelles, lipid nanodiscs or bilayer membranes around the transmembrane region using Positioning of Proteins in Membranes 3.0 (PPM 3.0), which positions membrane proteins by minimising the free energy of transfer into an anisotropic solvent model[43,44]. While PPM 3.0 supports both planar and curved membranes, we restricted analyses here to planar geometries and reserve curved membranes for future work. Comparison of the predicted boundaries to the cryo-EM data revealed that PPM systematically underestimated lipid bilayer and micelle thickness. To correct this, we refined boundary positions by computing intensity gradients normal to the predicted planes after enhancing intensity transitions with a Sobel filter (Figure 2b). By scanning through the volume along the plane normal, the locations of maximal ascending and descending gradients can then be identified and used to define the final membrane or micelle boundaries. We applied this workflow to extract lipid and detergent densities from the averaged half maps of 254 sequence-diverse membrane protein structures deposited in the EMDB, of which 186 high-quality segmentations were retained after curation (Figure 2b, Figure S1). In most cases, these densities appear smooth because the majority of lipids or detergent molecules in membranes, micelles, or nanodiscs are mobile, resulting in averaging out of structural detail such that no individual lipid or detergent molecules can be identified. This typically means that the micelle, nanodisc or membrane components in these complexes are disordered beyond a resolution of about 10–20 Å (Figure 3a-b). The shape of these lipid or detergent structures is highly pleomorphic, depending on both the shape and symmetry of the embedded membrane protein or complex and the properties and composition of the surrounding micelle or membrane (Figure S1). We characterised this variability by measuring flatness and fragmentation levels across the micelle dataset (Figure 3c-f). The variation in flatness reflects the different physicochemical properties and molecular packing geometry of the detergent or lipid molecules in micelles, bicelles and nanodiscs. Fragmentation likely arises from low signal-to-noise ratios in the raw data. Estimating molecular boundaries when signal strength approaches noise levels is inherently challenging and can produce discontinuous, rough surfaces. While some micelles exhibit uniform fragmentation in all directions, others display anisotropy, potentially reflecting preferential protein-lipid interactions along specific orientations. Together, the flatness and fragmentation distributions shown in Figure 3c provide a compact representation of the geometric variability captured in the training dataset. For a newly analysed membrane protein structure, projecting the segmented contextual density into this feature space offers a simple way to assess whether its morphology falls within the range represented during training. Membrane mimics

whose geometric descriptors lie within the main density of the training distribution are expected to be well supported by the learned model (see Section 2.3), whereas pronounced outliers may provide a practical indication that a given membrane geometry is underrepresented in the training data and could constitute a more challenging, potentially out-of-distribution case.

2.3 Segmentation model and performance.
We used the annotated dataset to train a three-dimensional Swin-Conv-UNet (SCUNet) for voxel-wise segmentation of cryo-EM density maps (Figure S1). The network follows a U-Net–style encoder–decoder architecture with skip connections[45] and is composed of Swin-Conv blocks that combine convolutional processing with window-based vision transformer modules[46,47]. Within each block, convolutional layers capture local density features, while transformer blocks aggregate information within fixed-size spatial windows. Although SCUNet was originally introduced for image restoration tasks[47], we adapt the same principle here for binary voxel-wise classification of lipid membrane or detergent micelle density versus macromolecular signal as the combination of convolutional layers and shifted window-based transformer modules enable a straighforward way for context aggregation of spatially extended density features[48]. Details of the network architecture, training protocol, and optimisation behaviour are summarised in Table S3, Figure S2a-d and described in the Methods.

2.4 Threshold dependence and connectivity filtering.
SURFER produces a continuous-valued voxel-wise confidence score indicating the likelihood that a given voxel belongs to membrane or membrane mimic density. To obtain a segmentation mask for visualisation or density subtraction, this output must be binarised at a chosen threshold. Because membrane mimic density typically occupies only a small fraction of the molecular volume and is dominated by weakly bounded signal, SURFER produces probabilistic outputs with rather broad confidence distributions near the class boundaries (Figure S2e). As a result, converting these probabilities into binary masks is sensitive to the chosen threshold, and the threshold that yields optimal segmentation can vary across datasets. At low segmentation thresholds, this sensitivity can lead to false-positive inclusion of low-resolution density that is not part of the membrane or membrane mimic. A good example illustrating this effect is the nanodisc-embedded transient receptor potential (TRP) ion channel NOMPC[49]. NOMPC contains 29 ankyrin repeat (AR^{1-29}) domains, of which AR^{1-7} are highly flexible and resolved only at low resolution comparable to that of the surrounding lipid nanodisc. As shown in Figure 4a and 4b, SURFER-generated segmentation masks at low thresholds include substantial off-target density corresponding to these flexible AR^{1-7} domains. Increasing the threshold progressively suppresses this off-target signal; however, rather aggressive thresholds are required before it is fully eliminated, at which point the nanodisc density itself begins to fragment and thus results in incomplete subtraction (Figure 4c). A practical way to alleviate this trade-off is to exploit the characteristic connectivity of membrane and membrane mimic density. Such density typically forms extended and spatially contiguous regions, whereas most off-target density will mainly be composed of smaller and disconnected components. SURFER therefore includes an optional connectivity filtering step in which only the largest connected component of the binarised segmentation mask is retained prior to subtraction. As illustrated in Figure 4c, including this connectivity filter can help exclude spurious off-target density while preserving the membrane mimic envelope at moderate thresholds, and thus allow clean subtraction of the lipid nanodisc.

To more generally examine the effect of binarisation threshold on segmentation quality, we evaluated voxel-wise segmentation accuracy as a function of threshold using the F1 score, computed on the held-out test dataset for which 16 curated reference segmentations were available. The F1 score, defined as the harmonic mean of precision and recall, provides a measure of segmentation performance that is particularly informative in the presence of strong class imbalance as is the case here (Figure S2d). As shown in Figure 4d, the F1 score exhibits a pronounced dependence on the chosen segmentation threshold. For any given reconstruction, there exists a threshold at which the best balance is achieved between recovering the membrane mimic density and avoiding false-positive inclusion of unrelated low-resolution signal. At low thresholds, recall is high but precision is reduced due to inclusion of off-target low-resolution signal, leading to moderate overall performance. Increasing the threshold suppresses these false positives and improves precision, resulting in a peak F1 score at an intermediate threshold. At still higher thresholds, recall decreases as the predicted membrane mimic density becomes fragmented or incomplete, again reducing the F1 score. The location and shape of this optimum varies between reconstructions but for most reconstructions F1 scores are relatively flat across a range of segmentation thresholds centred on 0.5. This behaviour highlights two points: First, even though a generic 0.5 threshold appears well-suited as a reasonable best guess, segmentation quality can probably not be reliably optimised using a single global threshold across datasets. Second, the threshold dependence observed in the quantitative evaluation mirrors the practical challenges encountered during visual inspection, where users may have to balance removal of smooth contextual density against preservation of weak but structurally relevant features. These observations motivate the interactive threshold selection and connectivity-based filtering implemented in SURFER, which allow users to adapt the segmentation to the specific characteristics of each reconstruction rather than relying on a fixed global threshold (Figure 4e & S2f).

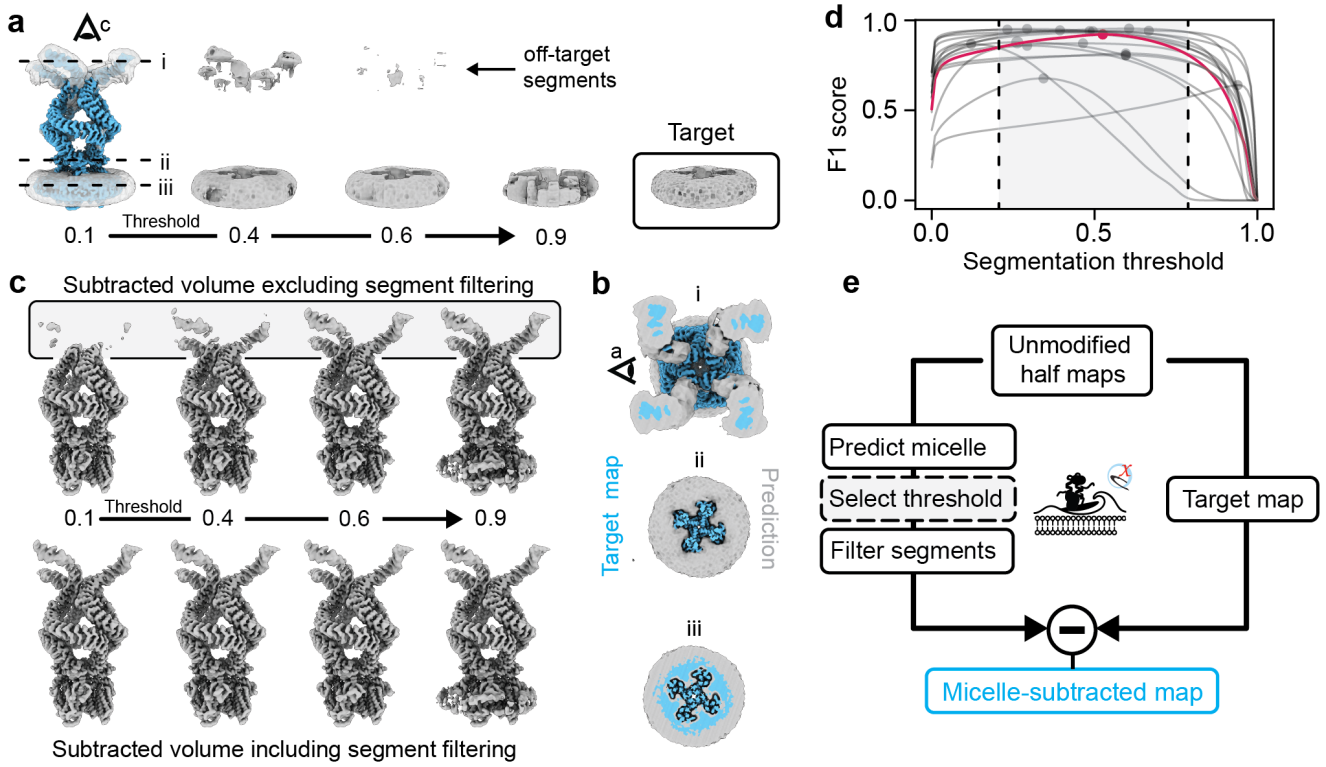

Fig. 4. Threshold dependence and connectivity filtering | (a) Illustration of off-target segments in membrane-distal regions of nanodisc-embedded *D. melanogaster* NOMPC (EMDB: EMD-8702). Density contours for different binarisation thresholds are shown. (b) Effect of filtering based on largest connected volume on the micelle-subtracted volume. The top row shows how subtraction of the predicted segment without filtering also removes density from the off-target segments. The bottom row shows how segment filtering restricts substraction to only the micell density. (c) 2D section views at the three section planes indicated in (a) shows the overlap of the unfiltered micelle prediction with the target map. (d) F1 score as a function of binarisation threshold evaluated on the held-out test dataset. Segmentation performance peaks at an intermediate threshold, reflecting the trade-off between recall at low thresholds and precision at high thresholds. The position and sharpness of this optimum vary across reconstructions. The curve for EMD-8702 is shown in red. (e) Schematic illustration of the interactive thresholding workflow implemented in SURFER, highlighting how users can explore the effect of threshold choice on segmentation quality and apply connectivity filtering prior to density subtraction.

2.5 Interactive segmentation and context exploration. SURFER is deployed as an interactive plugin within the UCSF ChimeraX environment (Figure 5a). ChimeraX is increasingly used for atomic coordinate and density map display and manipulation, making it a natural platform for a user-friendly contextual segmentation workflow. The plugin provides a ChimeraX interface to the SURFER tool, enabling users to perform contextual segmentation directly during map inspection and analysis (Figure 5b). However, all SURFER functions can also be programmatically accessed via the command line if desired. The workflow comprises an automated segmentation step followed by interactive threshold refinement and application. The SURFER tools allows users to specify a pair of unfiltered, independent half maps as input. From these, SURFER computes a FDR-controlled confidence mask to define the molecular boundary and subsequently predicts detergent- or membrane-associated density within this region. The predicted contextual density can be interactively thresholded to generate a binary segmentation mask, which can then be applied to any target map aligned with the input to selectively remove micelle or membrane signal (Figure 5b). Target maps may include the unfiltered input map itself, for example, to generate masks for local 3D refinement, or post-processed maps intended for display and figure generation, such as contextualised maps produced by LocScale-2.0 FEM map optimisation. The ChimeraX interface allows users to toggle between representations with and without contextual density to facilitate direct visual comparison. Depending on map size, the procedure typically completes within seconds for smaller reconstructions and within approximately 10 minutes for the largest maps tested (Table 1).

Table 1. SURFER runtime for selected EMDB entries

EMDB	Box size (voxels)	Pixel size (Å)	Box size (Å)	Volume ($\times 10^6$ Å^3)	Time
30017	240	1.03	247.2	15.1	36–38 s
3061	180	1.40	252.0	16.0	30–40 s
8702	400	1.20	480.0	110.6	3–4 min
13234†	336	1.70	571.2	186.3	4–5 min
19999†	460	1.63	749.8	421.5	10–12 min

† Map does not contain micelle density; used to quantify runtime only.

Figure 5c illustrates this workflow for the human γ-secretase complex (EMDB: EMD-3061)[26]. In this example, the target map is a LocScale-2.0 feature-enhanced map (LocScale-FEM), in which ordered macromolecular regions and more poorly resolved contextual density including the detergent micelle are represented on a comparable intensity scale. Application of SURFER

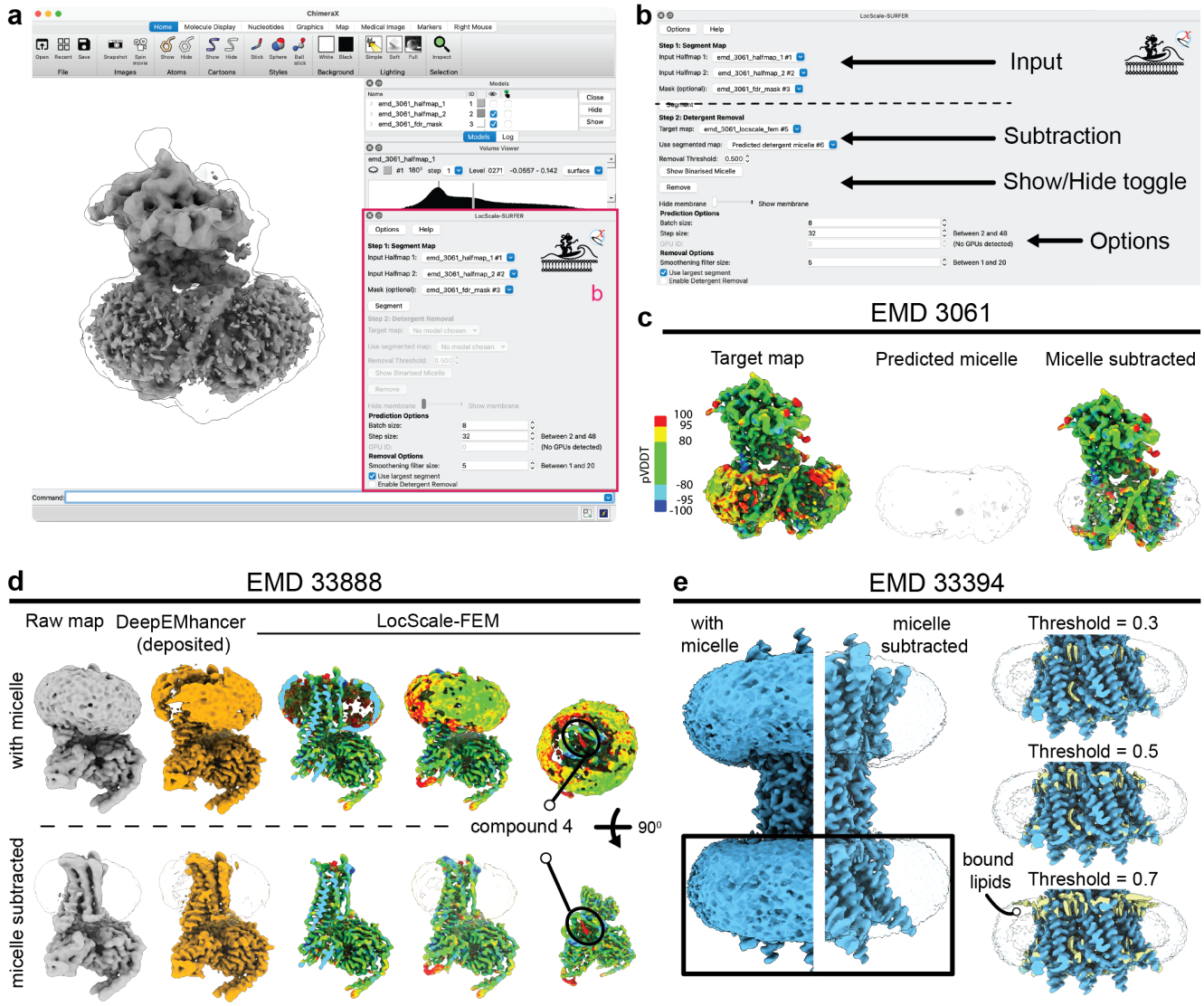

Fig. 5. Interactive segmentation and contextual density control with SURFER in UCSF ChimeraX | (a) ChimeraX interface illustrating SURFER applied to the γ-secretase complex (EMDB: EMD-3061). The averaged unfiltered map is shown in grey, with the outline of the FDR-based confidence mask superposed. (b) Close-up of the SURFER tool interface. SURFER accepts a pair of unfiltered half maps as input and optionally a molecular boundary mask. Predicted micelle density can be visualised across a range of binarisation thresholds. Once a threshold is selected, the segmented micelle can be subtracted from any target map aligned to the raw input. In this example, a LocScale-FEM optimised map[19] is used as the target. (c) LocScale-FEM target map, SURFER-predicted micelle density, and the corresponding micelle-subtracted map for EMD-3061. (d) Application of SURFER to the relaxin family peptide receptor 4 (EMDB: EMD-33888). Shown are the raw map, the deposited DeepEMhancer-optimised map, and the LocScale-FEM map before (top) and after (bottom) micelle subtraction. Density corresponding to the bound RXFP4 ligand is preserved int he subtracted map. (e) Application of SURFER to the connexin43 (Cx43/GJA1) gap junction intercellular channel (EMDB: EMD-33394). A LocScale-2.0 optimised target map is shown before and after micelle subtraction. Density associated with tightly bound lipids is retained in the micelle-subtracted map, although its retention depends on the chosen binarisation threshold.

segmentation to the LocScale-FEM map enables selective removal of the surrounding micelle while preserving the protein-associated transmembrane features that benefit from LocScale-2.0 optimisation such as the TMD2 (Figure 5c). We also tested the ability of the segmentation workflow in distinguishing smooth contextual density arising from mobile detergent molecules or lipids from density corresponding to ordered lipids or small-molecule ligands at the protein–micelle interface. Figure 5d shows the result for the relaxin family peptide receptor 4 (RXFP4; EMDB: EMD-33888), which contains a weakly bound amidrazone scaffold compound located at the transmembrane region[19,50]. We performed micelle subtraction on the raw reconstruction, the author-deposited DeepEMhancer-optimised map, and a LocScale-FEM map of RXFP4. Comparison of the maps before and after micelle subtraction shows that, in this case, SURFER effectively removes the surrounding micelle while preserving the ligand density at the automatically determined segmentation threshold. There are, however, cases in which the segmentation mask at the default threshold may encroach upon less well-ordered density in transmembrane regions. Figure 5e illustrates application of SURFER to the hexameric connexin43 (Cx43/GJA1) gap junction intercellular channel (EMDB: EMD-33394) in lipid nanodiscs, again using a

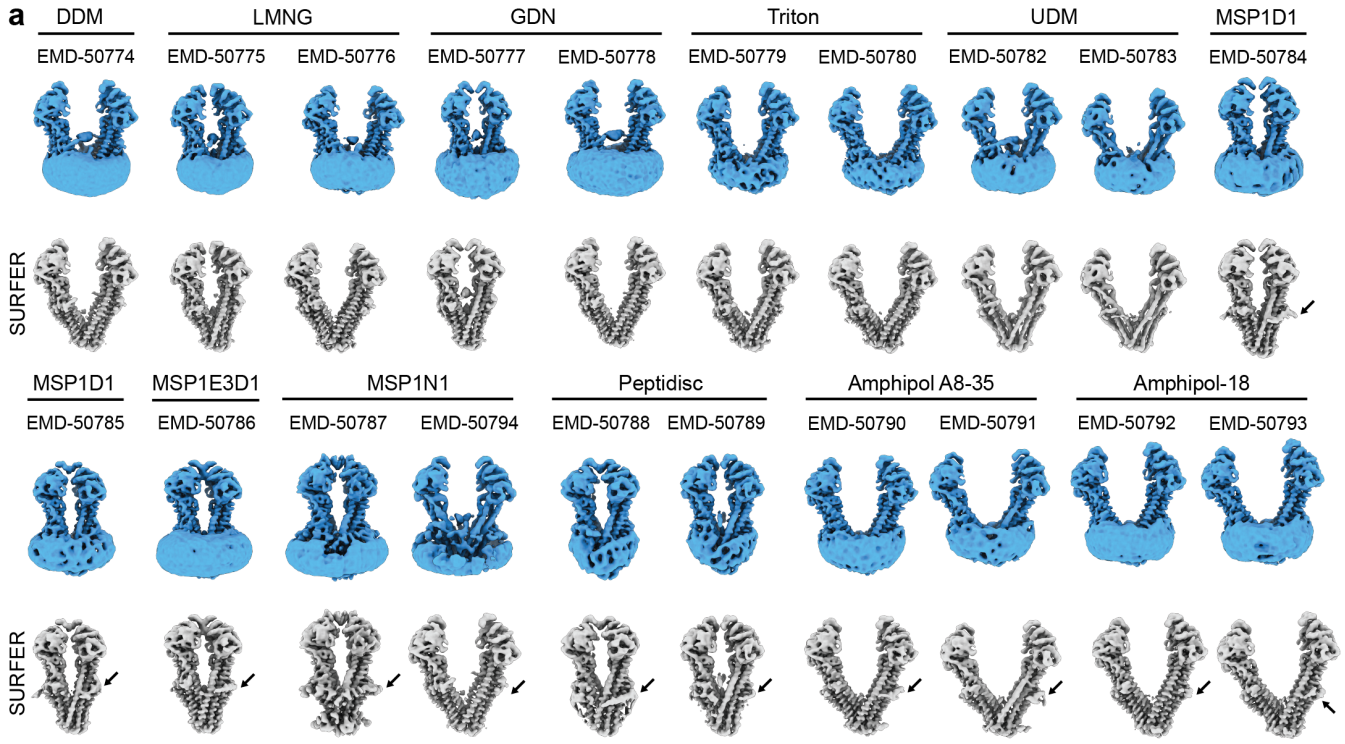

Fig. 6. Segmentation and subtraction of diverse membrane mimics for MsbA | (a) LocScale-FEM maps of MsbA solved in 12 different membrane mimics (EMDB: EMD-50774 through EMD-507974) before (blue) and after (grey) segmentation and micelle subtraction with SURFER. Arrows indicate retained density of membrane scaffold proteins, petidiscs and amphipols.

LocScale-2.0–optimised target map[19,51]. The Cx43/GJA1 structure of EMD-33394 (PDB: 7xqf) contains 24 modelled copies of the cholesterol ester cholesteryl hemisuccinate (CHS) and 12 copies of the phospholipid 1-palmitoyl-2-oleoyl-phosphatidylethanolamine (POPE) tightly associated with the transmembrane domains of both hemichannels. When binarising the segmentation mask at the default threshold of 0.5 or lower, in this case, the retention of ordered lipid features depends on the chosen binarisation threshold for the segmentation mask, reflecting their partial spatial overlap with the segmented contextual density. It thus appears that interactive threshold adjustment can be necessary to balance suppresion of smooth contextual density and retention of structured, protein-associated lipid density.

2.6 Generalisation across membrane mimics.

To assess the ability of SURFER to segment contextual density across a broad range of membrane environments, we investigated a set of 20 cryo-EM structures of the bacterial ATP-binding cassette transporter MsbA determined in 12 distinct membrane mimetic systems (EMDB: EMD-50774–EMD-50794). This dataset provides a controlled benchmark in which in which the macromolecular core adopts a limited set of conformations, while the surrounding contextual density varies substantially in geometry, intensity and other characteristics. The series includes classical detergents (DDM, LMNG, GDN, Triton, UDM), nanodiscs assembled with different membrane scaffold proteins (MSP) variants, peptidiscs, and amphipols[4]. These membrane mimics produce contextual density ranging from compact, approximately spherical micelles to flattened nanodiscs and irregular peptidisc belts (Figure 6a). We applied SURFER segmentation to all 20 reconstructions using the default binarisation threshold and subtracted the predicted lipid or detergent density from each map (Figure 6a). Across the set of 20 maps, SURFER consistently identified the dominant detergent or lipid envelope surrounding MsbA despite pronounced differences in micelle shape and thickness. Differences in segmentation quality appear primarily associated with variations in signal continuity and resolution of the membrane mimic density rather than systematic misclassification. Subtraction of the predicted contextual density preserved the structured transmembrane helices and cytosolic domains of MsbA across the series. In several nanodisc and amphipol reconstructions, density corresponding to membrane scaffold proteins or polymer belts remained partially visible (arrows in Figure 6a). These results indicate that SURFER generalises across a broad range of experimentally distinct membrane mimetic systems without mimic-specific parameter adjustment and captures shared geometric and intensity features of contextual lipid and detergent density.

Discussion and conclusion

Automated segmentation of contextual membrane and membrane mimic density in cryo-EM maps poses chal-

lenges distinc from from those encountered when segmenting density of ordered macromolecular components. Primarily inspired by LocScale-2.0[19], which optimises map representations to emphasise contextual density, SURFER was built to provide automated segmentation of membrane-related signal together with interactive control over its visualisation. While not explicitly tested for this purpose, the procedure may in principle facilitate mask generation for membrane subtraction or focused refinement. This could be advantageous for small integral membrane proteins, where dominant low-resolution micelle or nanodisc signal may influence a lignment. However, caution is warranted as subtracting signal from particle images or reconstructions risks removal of structurally relevant signal. While the observed robustness of SURFER segmentation is encouraging in this respect, systematic evaluation of its suitability for subtraction-based and focused refinement s trategies r emains t he s ubject of future work.

The results presented here show that SURFER can reliably identify membrane mimetic density across a wide range of membrane protein structures and membrane mimic geometries. In many cases, default segmentation parameters appear sufficient t o a ccurately separate and subtract contextual density while preserving ordered transmembrane regions. In more challenging situations, interactive adjustment of the binarisation threshold may be required to balance removal of smooth contextual signal against retention of structured features. The segmentation and subtraction procedure typically completes within a few minutes on a conventional laptop, making it well suited for interactive use.

The integration of SURFER into the UCSF ChimeraX[40] environment allows users to directly assess the plausibility of the segmentation, compare maps with and without contextual density, and generate masks for downstream applications without leaving the visualisation environment. SURFER appears to generalise reasonably well across different membrane mimics but is less accurate for membrane geometries that are underrepresented in the training data, such as strongly curved lipid bilayers. In such cases, segmentation quality is reduced and greater user intervention may be required. Extending the training dataset to include a broader range of membrane geometries is therefore a natural direction for future work.

Overall, SURFER provides a straightforward and accessible approach for contextual density segmentation in cryo-EM maps. By enabling selective suppression or visualisation of membrane and detergent density during interactive analysis, it complements existing map processing and visualisation tools and facilitates more objective interpretation of membrane protein structures.

ACKNOWLEDGEMENTS

We thank Tom Goddard for useful suggestions concerning UCSF ChimeraX integration, as well as Arne Möller, Kilian Schnelle and the participants of the CCP-EM Icknield workshops for testing early versions of the method. This work was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (ERC-StG-852880 to AJ) and CCP-EM. We acknowledge the use of computational resources of the DelftBlue supercomputer provided by the Delft High Performance Computing Centre.

DATA AND CODE AVAILABILITY

SURFER is open-source software released under a BSD license and is available at https://github.io/cryotud/locscale-surfer, where detailed documentation is also provided. SURFER is also available directly from the ChimeraX toolshed at https://cxtoolshed.rbvi.ucsf.edu. The SURFER segmentation model (v0.1) can be accessed via Zenodo (DOI: https://zenodo.org/records/15488062). Scripts used for analysis and for the generation of figures shown in this manuscript can be found at https://github.com/cryotud/publications.

Methods

Software environment. Model training and inference were implemented in PyTorch (v2.9) with CUDA 12.6. Data processing and analysis were performed in Python (v3.11). Cryo-EM map and atomic model operations were carried out using `EMmer`[18], a Python library for cryo-EM map and model processing built on `mrcfile`[52] and `gemmi`[53].

Training dataset generation.

Selection of map–model pairs. Training data were derived from cryo-EM reconstructions containing detergent micelles or membrane-associated density. Half-map pairs for 500 EMDB entries were initially collected. From these, 254 reconstructions containing visible lipid membranes, nanodiscs or detergent belts were identified through automated filtering and manual inspection. From this initial set, only entries with corresponding atomic models deposited in the Protein Data Bank (PDB)[54] and annotated membrane orientations in the OPM database[55] were retained. After target-mask generation and manual quality control (see below), 186 map–model pairs were retained. Of these, 144 were used for training and 26 for validation (85:15 split), while 16 were reserved as an independent test set. The full curated set is listed in Table S1 and S2 .

Molecular envelope and difference mask generation. For each reconstruction in the curated training and validation set, a mask defining the molecular envelope was generated using false discovery rate (FDR)–controlled thresholding[19,42]. Unfiltered half maps were averaged prior to FDR estimation. Noise statistics were estimated from cubic regions extracted from the edges of the reconstruction volume. The noise box size was set to the larger of 20^3 voxels or 10% of the box dimension of the reconstructed volume. To obtain smooth and conservative molecular boundaries, maps were low-pass filtered to 5 Å before FDR thresholding. All masks were binarised at an FDR of 1% and used consistently throughout training and validation. Density contained in the FDR mask but not explained by the atomic model was identified by subtracting a model-derived mask from the FDR mask. Model masks were generated from simulated model maps using `gemmi`[53] by including all voxels within 3.0 Å of any atom in the deposited structure. The resulting difference mask frequently contained thin shell artefacts at protein boundaries due to minor discrepancies between model and envelope estimation. These artefacts were removed by uniform filtering followed by re-binarisation.

Prediction and refinement of membrane boundaries. Approximate membrane orientations were obtained from OPM[55]. Because OPM coordinates are not aligned to deposited PDB models, all OPM structures were rigid-body aligned to their corresponding PDB entries using `gemmi`.
Membranes were assumed to be planar for training data generation; curved membranes were excluded. Each membrane was represented by two parallel planes of the form

$$ax + by + cz = d.$$

Initial plane parameters were estimated from three non-collinear points sampled from the predicted membrane surface. To refine the plane positions, edges in the difference mask were enhanced using a Sobel filter. Slices were extracted along the direction normal to the membrane planes, and the number of foreground voxels per slice was computed to generate an intensity profile. Peaks in the gradient of this profile were used to determine the refined membrane boundaries. The difference mask was subsequently cropped between the two refined planes.

Geometric characterisation of contextual density. To quantify geometric variability across detergent and membrane-associated densities, two complementary descriptors were computed from binarised contextual volumes: a flatness ratio and a fragmentation ratio. Volumes were binarised at a threshold of 0.5 prior to feature extraction.

Flatness ratio. The flatness ratio provides a measure of the aspect ratio of the micelle or membrane envelope. A best-fitting ellipsoid was computed for each binarised volume using `pyradiomics`[56]. The height of the micelle was defined as the minor principal axis of the ellipsoid and the width as the major principal axis. The flatness ratio was defined as

$$\text{Flatness} = 1 - \frac{\text{height}}{\text{width}}$$

Values approaching 1.0 indicate highly flattened, bilayer-like geometries, whereas lower values correspond to more isotropic or weakly anisotropic detergent belts.

Fragmentation ratio. Spatial continuity of the segmented density was quantified using a surface-area-to-volume metric that describes the degree of fragmentation. Binarised volumes were converted to triangular meshes using the marching cubes algorithm[57]. The surface area was computed as the sum of triangle areas, and the enclosed volume was obtained by tetrahedral decomposition.
The fragmentation ratio was defined as

$$\text{Fragmentation} = \frac{\text{surface area}}{\text{volume}}$$

Because area and volume are computed in physical units using the map pixel size, this ratio has units of $Å^{-1}$. Higher values indicate increased roughness or fragmentation.

Connectivity filtering. A binary segmentation sometimes contain multiple mutually disconnected regions, including unrelated low-resolution density. Connected-component analysis was done using `label` function implemented in `scikit-image` and only the largest connected component

was retained. The filtered volumes were manually inspected for dataset curation. Entries for which micelle density could not be reliably isolated were excluded. This curation step yielded the final dataset used for training, validation, and testing.

Segmentation network.

Network architecture. SURFER employs a 3D Swin-Conv U-Net (SCUNet)[47]. The encoder consists of three blocks, each composed of a Swin-Conv module followed by strided convolution for downsampling. Within each Swin-Conv module, window-based transformer layers operate in parallel with convolutional layers and their outputs are concatenated prior to resolution reduction. A single Swin-Conv module forms the bottleneck. The decoder mirrors the encoder using transposed convolutions for upsampling and skip connections to corresponding encoder features. Architectural details are summarised in Table S3 and Figure S2a.

Pre-processing and augmentation. All input maps were resampled to 1 Å/voxel and intensities were standardised to zero mean and a standard deviation of 0.1. Data augmentation was applied prior to cube extraction and included random rotations about Cartesian axes, translations up to ±10 voxels, B-factor modulation (random shifts between 0 and 400 $Å^2$; applied twice per map), and Gaussian blurring corresponding to effective resolutions between 5 and 20 Å.

Subvolume sampling. Training samples consisted of cubic subvolumes of size 48 Å. Cubes were extracted on a regular Cartesian grid with 26 Å spacing. Grid positions were classified as signal or noise using the resampled FDR mask. All signal-containing cubes were retained. Additional noise cubes were randomly sampled to achieve a signal-to-noise cube ratio of 4:1.

Training parameters. The network was trained using cube-wise binary cross-entropy (BCE) loss with L1 regularisation and the Adam optimiser with default momentum parameters ($\beta = [0.9, 0.999]$). Training was performed over eight epochs. The optimal weights for the L1 regularizer, learning rate, and batch size were determined empirically using the Optuna optimization framework[58]. The optimal values were found to be as follows: L1 weight decay of 0.0025, a learning rate of 5×10^{-4}, and a batch size of 64.

ChimeraX integration.

SURFER is distributed as a ChimeraX bundle and can be installed via the ChimeraX Toolshed or from the source repository. In ChimeraX, the tool is accessible under *Tools* → *Volume Data*. The bundle includes all required metadata and trained model weights within the source distribution. The workflow comprises two stages. First, unfiltered half maps are supplied for segmentation. If preferred, a user-defined mask can also be provided. Second, the predicted segmentation is applied to any aligned target map. Users select a binarisation threshold to define the final segmentation mask, which may be optionally smoothed using a uniform filter with a default width of 5 voxels. Contextual density can be interactively toggled to compare representations with and without membrane signal.

References

1. R. Nygaard, J. Kim and F. Mancia. Cryo-electron microscopy analysis of small membrane proteins. *Current Opinion in Structural Biology*, 64:26–33, [2020].
2. C.-T. Chien, M. Maduke and W. Chiu. Single-particle cryogenic electron microscopy structure determination for membrane proteins. *Current Opinion in Structural Biology*, 92: 103047, [2025].
3. G. G. Privé. Detergents for the stabilization and crystallization of membrane proteins. *Methods*, 41:388–397, [2007].
4. L. Hoffmann, A. Baier, L. Jorde, M. Kamel, J.-H. Schäfer, K. Schnelle, A. Scholz, D. Shvarev, J. E. M. M. Wong, K. Parey, D. Januliene, et al. The ABC transporter MsbA in a dozen environments. *Structure*, 33:916–923.e4, [2025].
5. C. Tribet, R. Audebert and J.-L. Popot. Amphipols: Polymers that keep membrane proteins soluble in aqueoussolutions. *Proceedings of the National Academy of Sciences*, 93: 15047–15050, [1996].
6. A. Nath, W. M. Atkins and S. G. Sligar. Applications of Phospholipid Bilayer Nanodiscs in the Study of Membranes and Membrane Proteins. *Biochemistry*, 46:2059–2069, [2007].
7. J. Frauenfeld, R. Löving, J.-P. Armache, A. F.-P. Sonnen, F. Guettou, P. Moberg, L. Zhu, C. Jegerschöld, A. Flayhan, J. A. G. Briggs, H. Garoff, C. Löw, Y. Cheng, et al. A saposin-lipoprotein nanoparticle system for membrane proteins. *Nature Methods*, 13:345–351,[2016].
8. L. Dietrich, A.-N. A. Agip, C. Kunz, A. Schwarz and W. Kühlbrandt. In situ structure and rotary states of mitochondrial ATP synthase in whole Polytomella cells. *Science*, 385(6713): 1086–1090, [2024].
9. P. B. Rosenthal and R. Henderson. Optimal Determination of Particle Orientation, Absolute Hand, and Contrast Loss in Single-particle Electron Cryomicroscopy. *Journal of Molecular Biology*, 333(4):721–745, [2003].
10. 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(1):1–8, [2021].
11. T. C. Terwilliger, O. V. Sobolev, P. V. Afonine, P. D. Adams and R. J. Read. Density modification of cryo-EM maps. *Acta Crystallographica Section D: Structural Biology*, 76:912–925, [2020].
12. J. He, T. Li and S.-Y. Huang. Improvement of cryo-EM maps by simultaneous local and non-local deep learning. *Nature Communications*, 14(1):3217, [2023].
13. J. Selvaraj, L. Wang and J. Cheng. CryoTEN: Efficiently Enhancing Cryo-EM Density Maps Using Transformers. *BioRxiv* 2024.09.06.611715,[2024].
14. N. A. Hoffmann, A. J. Jakobi, M. K. Vorländer, C. Sachse and C. W. Müller. Transcribing RNA polymerase III observed by electron cryomicroscopy. *The FEBS Journal*, 283: 2811–2819, [2016].
15. R. F. Berkeley, B. D. Cook and M. A. Herzik. Machine learning approaches to cryoEM density modification differentially affect biomacromolecule and ligand density quality. *Frontiers in Molecular Biosciences*, 11, [2024].
16. A. J. Jakobi, M. Wilmanns and C. Sachse. Model-based local density sharpening of cryo-EM maps. *eLife*, 6:e27131, [2017].
17. K. Ramlaul, C. M. Palmer and C. H. S. Aylett. A Local Agreement Filtering Algorithm for Transmission EM Reconstructions. *Journal of Structural Biology*, 205:30–40, [2019].
18. A. Bharadwaj and A. J. Jakobi. Electron scattering properties of biological macromolecules and their use for cryo-EM map sharpening. *Faraday Discussions*, 240:168–183, [2022].
19. A. Bharadwaj, R. d. Bruin and A. J. Jakobi. Confidence-guided cryo-EM map optimisation with LocScale-2.0. *BioRxiv*: 2025.09.11.674726v1, [2025].
20. A. Singer and F. J. Sigworth. Computational Methods for Single-Particle Electron Cryomicroscopy. *Annual Review of Biomedical Data Science*, 3:163–190,[2020].
21. A. Punjani, H. Zhang and D. J. Fleet. Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction.*Nature Methods*, 17:1214–1221 [2020].
22. K. Ramlaul, C. M. Palmer, T. Nakane and C. H. S. Aylett. Mitigating local over-fitting during single particle reconstruction with SIDESPLITTER. *Journal of Structural Biology*, 211: 107545, [2020].
23. Y. Liu and F. J. Sigworth. Automatic cryo-EM particle selection for membrane proteins in spherical liposomes. *Journal of Structural Biology*, 185:295–302, [2014].
24. K. H. Jensen, S. S. Brandt, H. Shigematsu and F. J. Sigworth. Statistical modeling and removal of lipid membrane projections for cryo-EM structure determination of reconstituted membrane proteins. *Journal of Structural Biology*, 194:49–60, [2016].
25. S. L. Ilca, A. Kotecha, X. Sun, M. M. Poranen, D. I. Stuart and J. T. Huiskonen. Localized reconstruction of subunits from electron cryomicroscopy images of macromolecular complexes. *Nature Communications*, 6:8843, [2015].
26. X.-c. Bai, C. Yan, G. Yang, P. Lu, D. Ma, L. Sun, R. Zhou, S. H. W. Scheres, et al. An atomic structure of human γ-secretase. *Nature*, 525:212–217, [2015].
27. E. Fernández-Giménez, M. Martínez, R. Sánchez-García, R. Marabini, E. Ramírez-Aportela, P. Conesa, J. M. Carazo, et al. Cryo-EM density maps adjustment for subtraction, consensus and sharpening. *Journal of Structural Biology*, 213:107780, [2021].

28. N. Otsu. A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9:62–66, [1979].
29. L. Najman and M. Schmitt. Watershed of a continuous function. *Signal Processing*, 38: 99–112, [1994].
30. N. Volkmann. A novel three-dimensional variant of the watershed transform for segmenta-tion of electron density maps. *Journal of Structural Biology*, 138:123–129, [2002].
31. G. D. Pintilie, J. Zhang, T. D. Goddard, W. Chiu and D. C. Gossard. Quantitative analysis of cryo-EM density map segmentation by watershed and scale-space filtering, and fitting of structures by alignment to regions. *Journal of Structural Biology*, 170:427–438, [2010].
32. R. Nock and F. Nielsen. Statistical region merging. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26:1452–1458, [2004].
33. G. Terashi, Y. Kagaya and D. Kihara. MAINMASTseg: Automated Map Segmentation Method for Cryo-EM Density Maps with Symmetry. *Journal of Chemical Information and Modeling*, 60:2634–2643, [2020].
34. A. Martinez-Sanchez, I. Garcia, S. Asano, V. Lucic and J.-J. Fernandez. Robust membrane detection based on tensor voting for electron tomography. *Journal of Structural Biology*, 186:49–61, [2014].
35. M. Chen, W. Dai, S. Y. Sun, D. Jonasch, C. Y. He, M. F. Schmid, W. Chiu, et al. Convolutional neural networks for automated annotation of cellular cryo-electron tomograms. *Nature Methods*, 14:983–985, [2017].
36. L. Lamm, R. D. Righetto, W. Wietrzynski, M. Pöge, A. Martinez-Sanchez, T. Peng, et al. MemBrain: A deep learning-aided pipeline for detection of membrane proteins in Cryo-electron tomograms. *Computer Methods and Programs in Biomedicine*, 224:106990,[2022].
37. I. de Teresa-Trueba, S. K. Goetz, A. Mattausch, F. Stojanovska, C. E. Zimmerli, M. Toro-Nahuelpan, D. W. C. Cheng, F. Tollervey, C. Pape, M. Beck, A. Diz-Muñoz, A. Kreshuk, J. Mahamid, et al. Convolutional networks for supervised mining of molecular patterns within cellular context. *Nature Methods*, 20:284–294, [2023].
38. A. Morales-Martínez, E. Garduño, J. M. Carazo, C. O. S. Sorzano and J. L. Vilas. Membrane and vesicle structure detection in cryo-electron tomography based on deep learning. *Journal of Structural Biology*, 217:108258, [2025].
39. M. Siggel, R. K. Jensen, V. J. Maurer, J. Mahamid and J. Kosinski. ColabSeg: An interactive tool for editing, processing, and visualizing membrane segmentations from cryo-ET data. *Journal of Structural Biology*, 216:108067, [2024].
40. E. C. Meng, T. D. Goddard, E. F. Pettersen, G. S. Couch, Z. J. Pearson, J. H. Morris, et al. UCSF ChimeraX: Tools for structure building and analysis. *Protein Science*, 32:e4792,[2023].
41. The wwPDB Consortium. EMDB—the Electron Microscopy Data Bank. *Nucleic Acids Research*, 52:D456–D465, [2024].
42. M. Beckers, C. M. Palmer and C. Sachse. Confidence maps: statistical inference of cryo-EM maps. *Acta Crystallographica Section D Structural Biology*, 76:332–339, [2020].
43. A. L. Lomize, I. D. Pogozheva and H. I. Mosberg. Anisotropic Solvent Model of the Lipid Bilayer. 2. Energetics of Insertion of Small Molecules, Peptides, and Proteins in Membranes. *Journal of Chemical Information and Modeling*, 51:930–946, [2011].
44. A. L. Lomize, S. C. Todd and I. D. Pogozheva. Spatial arrangement of proteins in planar and curved membranes by PPM 3.0. *Protein Science*, 31:209–220, [2022].
45. O. Ronneberger, P. Fischer and T. Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. In N. Navab, J. Hornegger, W. M. Wells and A. F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, [2015].
46. J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool and R. Timofte. SwinIR: Image Restoration Using Swin Transformer. In *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, pages 1833–1844, [2021].
47. K. Zhang, Y. Li, J. Liang, J. Cao, Y. Zhang, H. Tang, D.-P. Fan, R. Timofte, et al. Practical Blind Image Denoising via Swin-Conv-UNet and Data Synthesis. *Machine Intelligence Research*, 20:822–836, [2023].
48. N. Park and S. Kim. How Do Vision Transformers Work?, [2022]. arXiv:2202.06709 arXiv:2202.06709 [cs].
49. P. Jin, D. Bulkley, Y. Guo, W. Zhang, Z. Guo, W. Huynh, S. Wu, S. Meltzer, T. Cheng, L.Y. Jan, Y.-N. Jan, et al. Electron cryo-microscopy structure of the mechanotransduction channel NOMPC. *Nature*, 547(7661):118–122, [2017].
50. Y. Chen, Q. Zhou, J. Wang, Y. Xu, Y. Wang, J. Yan, Y. Wang, Q. Zhu, F. Zhao, C. Li, C.-W. Chen, X. Cai, R. A. D. Bathgate, C. Shen, H. Eric Xu, D. Yang, H. Liu, et al. Ligand recognition mechanism of the human relaxin family peptide receptor 4 (RXFP4). *Nature Communications*, 14(1):492, [2023].
51. H.-J. Lee, H. J. Cha, H. Jeong, S.-N. Lee, C.-W. Lee, M. Kim, J. Yoo, et al. Conformational changes in the human Cx43/GJA1 gap junction channel visualized using cryo-EM. *Nature Communications*, 14(1):931, [2023].
52. T. Burnley, C. M. Palmer and M. Winn. Recent developments in the CCP-EM software suite. *Acta Crystallographica Section D: Structural Biology*, 73(6):469–477, [2017].
53. M. Wojdyr. GEMMI: A library for structural biology. *Journal of Open Source Software*, 7(73): 4200, [2022].
54. wwPDB consortium. Protein Data Bank: the single global archive for 3D macromolecular structure data. *Nucleic Acids Research*, 47(D1):D520–D528, [2019].
55. M. A. Lomize, A. L. Lomize, I. D. Pogozheva and H. I. Mosberg. OPM: Orientations of Proteins in Membranes database. *Bioinformatics*, 22(5):623–625, [2006].
56. J. J. van Griethuysen, A. Fedorov, C. Parmar, A. Hosny, N. Aucoin, V. Narayan, R. G. Beets-Tan, J.-C. Fillion-Robin, S. Pieper, et al. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Research*, 77(21):e104–e107, [2017].
57. W. E. Lorensen and H. E. Cline. Marching cubes: A high resolution 3D surface construction algorithm. *SIGGRAPH Comput. Graph.*, 21(4):163–169, [1987].
58. T. Akiba, S. Sano, T. Yanase, T. Ohta and M. Koyama. Optuna: A Next-generation Hyper-parameter Optimization Framework, [2019]. Number: arXiv:1907.10902 arXiv:1907.10902 [cs].

Supplementary material

Interactive segmentation of membrane and membrane mimic densities in cryo-EM maps

Alok Bharadwaj[a], Lotte Veerbeek[a], and Arjen J. Jakobi[a]

[a] Department of Bionanoscience, Kavli Institute of Nanoscience, Delft University of Technology, 2629HZ, The Netherlands

Supplementary Figures

S1 **Shape variety of segmented membrane mimic densities** S3
S2 **Network architecture, training behaviour and confidence characteristics of SURFER predictions.** S4

Supplementary Tables

S1 EMDB and PDB identifiers for maps and models used in SURFER training and validation set. S2
S2 EMDB and PDB identifiers for maps and models used in SURFER test set. S2
S3 SCUNet architecture summary S2

Table S1. EMDB and PDB identifiers for maps and models used in SURFER training and validation set.

#	EMDB	PDB	#	EMDB	PDB	#	EMDB	PDB	#	EMDB	PDB	#	EMDB	PDB	#	EMDB	PDB	#	EMDB	PDB
1	23276	7lck	2	31399	7f0j	3	20244	6p46	4	10933	6yup	5	13882	7qb9	6	22316	7jga	7	24958	7sax
8	9932	6k7h	9	10830	6ykr	10	9112	6mgv	11	11799	7ain	12	22311	7jg5	13	24825	7s3i	14	32110	7vsg
15	31433	7f3f	16	10995	6yxr	17	13035	7oqz	18	10919	6ytl	19	22979	7koo	20	8354	5t4d	21	0976	6lto
22	21453	6vxz	23	30904	7dxc	24	11720	7ad3	25	22120	6xbm	26	22376	7jlp	27	23425	7lll	28	9935	6k7j
29	21972	6wzg	30	12142	7bcq	31	0200	6hd1	32	4270	6fkf	33	21705	6wkn	34	0926	6ln7	35	23436	7lly
36	13008	7ooa	37	21234	6vm0	38	0924	6ln5	39	10828	6ykm	40	30713	7dlu	41	9941	6k7m	42	25492	7sx3
43	25691	7t4x	44	30535	7d0i	45	30907	7dxf	46	13155	7p14	47	9695	6iok	48	0827	6l3v	49	8118	5irz
50	11733	7adp	51	22983	7kox	52	30017	6lyp	53	10184	6sgt	54	13985	7qia	55	0925	6ln6	56	12939	7oj8
57	22689	7k65	58	11782	7ahc	59	22882	7ki0	60	0928	6ln9	61	4789	6rb9	62	23476	7lpc	63	4733	6r69
64	20278	6p9y	65	11208	6zgj	66	24956	7sat	67	13095	7ow8	68	4611	6qp6	69	13662	7puy	70	0919	6lmt
71	30219	7bvg	72	21512	6w1m	73	4587	6qm4	74	20194	6ot5	75	11588	6zzx	76	12026	7b5d	77	22264	6xmu
78	20537	6pzt	79	31010	7e84	80	10049	6rx4	81	4411	6i53	82	20488	6pv8	83	7099	6bhu	84	30903	7dxb
85	22771	7kai	86	30216	7bvc	87	25800	7tbw	88	21236	6vm2	89	4594	6qmb	90	10094	6s3q	91	9937	6k7k
92	0947	6lqj	93	21672	6whg	94	25803	7tc0	95	23093	7kzz	96	20508	6pwn	97	10895	6ysf	98	22772	7kaj
99	9696	6iol	100	9119	6mhv	101	0469	6np0	102	21511	6w1j	103	0275	6hug	104	9113	6mgw	105	0975	6ltn
106	13201	7p5m	107	0921	6lmv	108	13969	7qhb	109	0501	6nt4	110	10735	6y9b	111	0094	6gyo	112	0280	6huk
113	9942	6k7n	114	4645	6qv6	115	26188	7tyn	116	9851	6jnf	117	9931	6k7g	118	24957	7sau	119	10182	6sgr
120	0915	6lly	121	0488	6nr3	122	21518	6w1y	123	0487	6nr2	124	8117	5irx	125	21671	6whc	126	21237	6vm3
127	12095	7b93	128	4919	6rld	129	30906	7dxe	130	0004	6giq	131	20509	6pwo	132	3860	5oyb	133	0923	6lmx
134	0199	6hcy	135	10185	6sgu	136	0716	6kls	137	22375	7jlo	138	30217	7bve	139	0498	6nsj	140	10829	6ykp
141	23473	7lp9	142	0922	6lmw	143	24310	7r87	144	7039	6b3j	145	4592	6qm9	146	30575	7d4p	147	31135	7egk
148	30576	7d4q	149	0499	6nsk	150	13880	7qa8	151	11629	7a46	152	30908	7dxg	153	12296	7nf6	154	30218	7bvf
155	6822	6jpf	156	4972	6roh	157	0981	6lum	158	4593	6qma	159	12996	7onj	160	0093	6gyn	161	4612	6qpb
162	0912	6lle	163	0927	6ln8	164	9610	6adq	165	31003	7e7s	166	20145	6oo4	167	20715	6ubt	168	20143	6oo3
169	6770	5xsy	170	9115	6mho															

Table S2. EMDB and PDB identifiers for maps and models used in SURFER test set.

#	EMDB	PDB	#	EMDB	PDB	#	EMDB	PDB	#	EMDB	PDB	#	EMDB	PDB	#	EMDB	PDB	#	EMDB	PDB
1	0257	6hra	2	4032	5lc5	3	4272	6fki	4	4288	6fo2	5	4588	6qm5	6	4589	6qm6	7	4646	6qvb
8	4746	6r7x	9	7009	6ave	10	7133	6bqv	11	7882	6dg7	12	8702	5vkq	13	8958	6e1n	14	8960	6e1p
15	9934	6k7i	16	9939	6k7l															

Table S3. SCUNet architecture summary

Layer (name)	Input	Output	InC	OutC	Params	Kernel
SCUNet	48^3	48^3	1	1	–	–
Head	48^3	48^3	1	32	–	–
Conv3d	48^3	48^3	1	32	864	3^3
Down-sampling block 1	48^3	24^3	32	64	–	–
ConvTransBlock (0)	48^3	48^3	32	32	19,469	–
ConvTransBlock (1)	48^3	48^3	32	32	19,469	–
Conv3d (2)	48^3	24^3	32	64	16,384	2^3
Down-sampling block 2	24^3	12^3	64	128	–	–
ConvTransBlock (0)	24^3	24^3	64	64	76,826	–
ConvTransBlock (1)	24^3	24^3	64	64	76,826	–
Conv3d (2)	24^3	12^3	64	128	65,536	2^3
Down-sampling block 3	12^3	6^3	128	256	–	–
ConvTransBlock (0)	12^3	12^3	128	128	305,204	–
ConvTransBlock (1)	12^3	12^3	128	128	305,204	–
Conv3d (2)	12^3	6^3	128	256	262,144	2^3
Bottleneck	6^3	6^3	256	256	–	–
ConvTransBlock (0)	6^3	6^3	256	256	1,216,616	–
ConvTransBlock (1)	6^3	6^3	256	256	1,216,616	–
Up-sampling block 3	6^3	12^3	256	128	–	–
ConvTranspose3d (0)	6^3	12^3	256	128	262,144	2^3
ConvTransBlock (1)	12^3	12^3	128	128	305,204	–
ConvTransBlock (2)	12^3	12^3	128	128	305,204	–
Up-sampling block 2	12^3	24^3	128	64	–	–
ConvTranspose3d (0)	12^3	24^3	128	64	65,536	2^3
ConvTransBlock (1)	24^3	24^3	64	64	76,826	–
ConvTransBlock (2)	24^3	24^3	64	64	76,826	–
Up-sampling block 1	24^3	48^3	64	32	–	–
ConvTranspose3d (0)	24^3	48^3	64	32	16,384	2^3
ConvTransBlock (1)	48^3	48^3	32	32	19,469	–
ConvTransBlock (2)	48^3	48^3	32	32	19,469	–
Tail	48^3	48^3	32	1	–	–
Conv3d (0)	48^3	48^3	32	1	864	3^3

Total parameters: 4,729,084 **Trainable:** 4,727,676 **Non-trainable:** 1,408

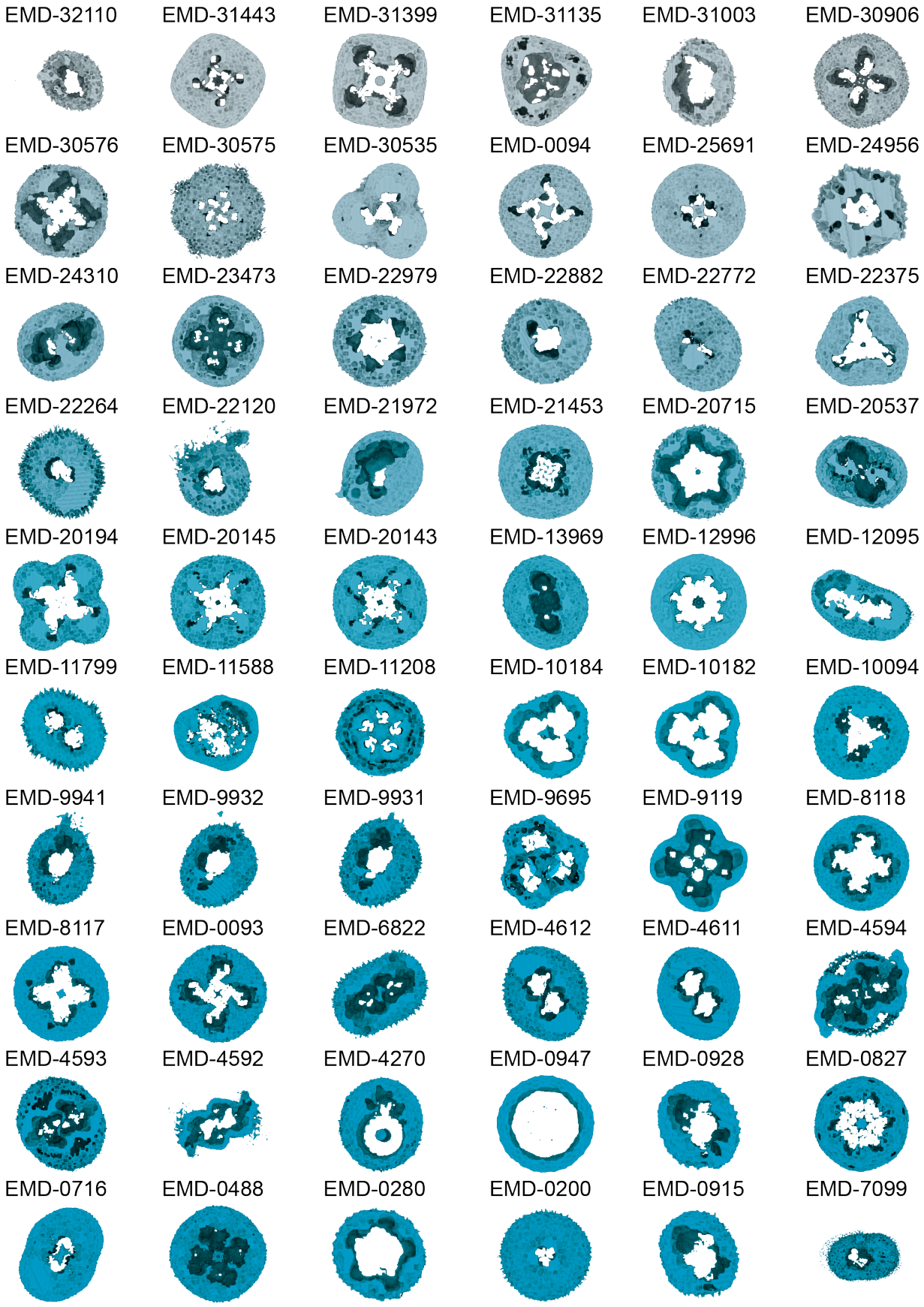

Fig. S1. | **Shape variety of segmented membrane mimic densities.** | Randomly selected examples of segmented membrane mimic densities in the training data.

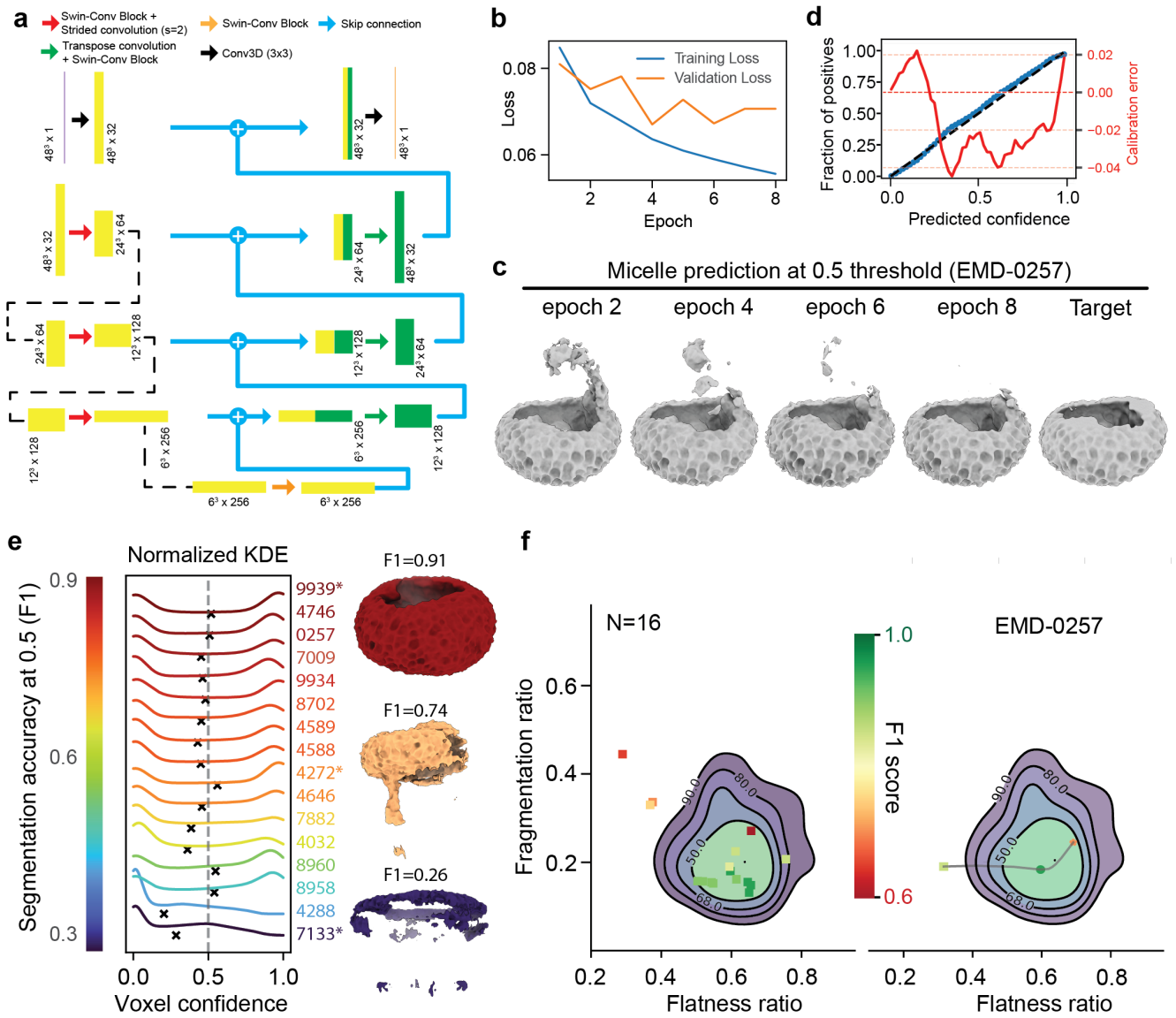

Fig. S2. | **Network architecture, training behaviour and confidence characteristics of SURFER predictions..** | (a) Network architecture of the SURFER SCU-Net. (b) Training and validation loss curves over eight epochs, showing stable optimisation without divergence between training and validation loss. (c) Predicted detergent micelle segmentation for a held-out test map (EMD-0257) using model checkpoints from successive epochs, binarised at a fixed threshold of 0.5. The overall micelle geometry is recovered early during training, while continued optimisation primarily reduces false-positive regions and sharpens the micelle boundary at the same threshold. (d) Reliability diagram aggregating voxel-wise predictions across all test maps, showing predicted confidence versus the observed fraction of true positives. The red curve indicates calibration error (predicted confidence minus observed fraction), with negative values corresponding to under-confidence. (e) Kernel density estimates of voxel-wise predicted confidence for individual test maps, weighted by the relative micelle volume within the molecular boundary. High-accuracy segmentations (top) show clearly bimodal confidence distributions with peaks near 0 and 1, whereas lower-accuracy cases (bottom) exhibit a skew toward low-confidence values, reflecting bias toward the majority class. Representative segmentations binarised at a threshold of 0.5 are shown alongside each distribution, with F1 scores indicated. (f) Left: Fragmentation and flatness ratios of the 16 test maps, thresholded at their respective optimal F1 scores, superimposed on the flatness–fragmentation distribution of the SURFER training set (see Figure 3c). Right: Trajectory of one representative test map (EMD-0257) across different binarisation thresholds (square: 0.1; circle: 0.5; asterisk: 0.9). Data points are coloured according to F1 score.