bioRxiv preprint doi: https://doi.org/10.1101/2021.07.25.453685; this version posted July 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

SegmA: Residue Segmentation of cryo-EM density maps

Mark Rozanov
Blavatnik School of Computer Science, Tel Aviv University

Haim J. Wolfson
Blavatnik School of Computer Science, Tel Aviv University

July 25, 2021

1 Abstract

The cryo-EM resolution revolution enables the development of algorithms for direct de-novo modelling of protein structures from given cryo-EM density maps. Deep Learning tools have been applied to locate structure patterns, such as rotamers, secondary structures and Cα atoms. We present a deep neural network (nicknamed SegmA) for the residue type segmentation of a cryo-EM density map. The network labels voxels in a cryo-EM map by the residue type (amino acid type or nucleic acid) of the sampled macromolecular structure. It also provides a visual representation of the density map by coloring the different types of voxels by their assgned colors. SegmA's algorithm combines a rotation equivariant group convolutional network with a traditional U-net network in a cascade. In addition SegmA estimates the labeling accuracy and reports only labels assigned with high confidence. At resolution of 3$\AA$ SegmAs accuracy is 80% for nucleotides. Amino acids which can be seen by eye, such as LEU, ARG and PHE, are detected by Segma with about 70% accuracy.

A web server of the application is under development at `https://dev.dcsh7cbr3o89e.amplifyapp.com`. The SegmA open code is available at `https://github.com/Mark-Rozanov/SegmA_3A/tree/master`

2 Introduction

Knowledge of the three dimensional structure of a protein is a major tool in elucidating its functionality. Cryo-electron microscopy (cryo-EM) (Callaway (2017)) has become a key experimental method for obtaining 3D visualization maps of large molecular structures. The recent "resolution revolution" (Kühlbrandt (2014)) in cryo-EM has led to an increasing number of near atomic resolution density maps deposited in the EM databank EMDB (Lawson et al. (2016)), making cryo-EM one of the leading experimental methods for determining protein structure. A microscope samples tens of thousands of images which are sorted, aligned and averaged to produce a 3-D density map volume. Afterwards a modelling process is applied to calculate atomic coordinates from this 3D map. The typical modelling process is mainly manual aided by computational modelling tools.

De-novo modelling methods process a cryo-EM map and obtain an atomic structure when homologous structures are not available. Traditional de-novo cryo-EM modelling algorithms utilize methods like template matching, 3D image analysis and energy minimization. Lasker et al. (2007), Lindert et al. (2012), Baker et al. (2011), Terashi & Kihara (2018), Chen et al. (2016), Liebschner et al. (2019), DiMaio et al. (2011) are examples of such methods. RENNSH (Ma et al. (2012)), Pathwalking (Chen et al. (2016)),

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.25.453685; this version posted July 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

SSELEarner (Si et al. (2012)) utilize traditional machine learning (ML) methods in structure determination. Those include among others kNN - k nearest neighbor clustering (RENNSH), SVM - support vector machines (SSE learner) and k-means (Pathwalking).

Recently Deep Learning (DL) tools started to play an important role in cryo-EM modeling. The Convolutional Neural Networks (CNN) and their variations perform well on the task of secondary structure detection (Maddhuri Venkata Subramaniya et al. (2019), Moritz et al. (2019)). One of the first attempts to get structural insight at the amino acid level was AAnchor(Rozanov & Wolfson (2018)). AAnchor estimates amino acid centers and types using a 3D CNN. Only residues recovered with high confidence (nicknamed anchors) are reported. A^2-Net (Xu et al. (2019)) is a two-stage modeling algorithm: a CNN for pattern recognition is followed by a Monte Carlo search to determine the protein structure. DeepTracer (Pfab et al. (2021)) uses a cascaded convolutional neural network to predict secondary structures, amino acid types, and backbone atom positions. Structure Generator (Li et al. (2020)) uses CNN to estimate amino acids with their poses, which are then fed to a graph convolutional network and a recurrent network to obtain the 3D protein structure. For a comprehensive review of Deep Learning methods in cryo-EM modelling we recommend Si et al. (2021) and Li et al. (2020).

Despite recent advances, fully automated protein structure determination from a cryo-EM map is still challenging. There are natural factors that limit the precision of an atomic structure estimation. Among those is the existence of regions of degraded quality in an experimental cryo-EM map caused by molecular flexibility or resolution degradation at the periphery of a molecular complex (Joseph et al. (2020), Si et al. (2021)). Another factor is the variation in residue size and conformation for different amino acid types. As a result residue recognition precision strongly depends on the amino acid type (Rozanov & Wolfson (2018)). The third factor to be mentioned is protein-RNA and protein-DNA complexes. Modelling of such a complex requires separation between amino acids and nucleic acids.

We address the three above mentioned challenges by a segmentation approach. In this segmentation task the output is a map of the same dimension as the input map where each voxel in the output map is coloured according to the amino acid type it belongs to (see Figure 6). Namely, the proposed DL algorithm produces a coloured visualization of an input cryo-EM map. White regions correspond to areas where the exact amino acid types cannot by detected, those are marked as "uncovered". In SegmA's output a residue conformation and boundary are seen in addition to the residue position and type. Voxels belonging to nucleic acids are marked by a different color, which enables separation of RNA/DNA from the proteins and the background.

Previous studies have highlighted some of the best practices in DL processing of a cryo-EM map:

- Use of cascade structures has advantages for unbalanced data. In this approach the first stage classifier rejects most false negatives and the successive classifiers perform exact labeling.
- Fully Convolutional U-Net architecture is a key to fast processing of a 3D image.
- A reliable data normalization technique is crucial due-to large data distribution variations (Figure 3).

SegmA incorporates the approaches mentioned above. In addition SegmA incorporates some novel approaches:

- SegmAs segmentation is based on the Group Convolutional Equivariant Network (G-CNN), (Cohen & Welling (2016)), to leverage rotation equivariance of a protein pose in a cryo-EM map.
- SegmA has the ability to estimate the regions where correct prediction is available and report results only in those regions.
- SegmA utilizes a simple and fully automatic input data normalization strategy.

At resolution of $3\mathring{A}$ SegmAs accuracy is 80% for nucleotides. Visually distinct amino acids, such as LEU, ARG and PHE, are detected by Segma with 70%

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.25.453685; this version posted July 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

accuracy. A web server of the application is under development at `https://dev.dcsh7cbr3o89e.amplifyapp.com`. The SegmAs open code is available at `https://github.com/Mark-Rozanov/SegmA_3A/tree/master`

3 Methods

3.1 Map Processing

Given a 3D electron density map of a protein/macromolecular assembly, our task is to assign a label to each voxel within the map. The assigned label should specify that the voxel belongs to one of the following 23 types: one of the 20 amino acids $(1-20)$, nucleotide (21), background (0), or uncovered (22), where "background" denotes voxels which belong neither to proteins nor to amino acids and "uncovered" denotes voxels which have not been labeled confidently enough to be considered. The algorithm is executed by a cascade of three neural networks - see Figure 2.

The initial labeling of the processed volume is performed by the Classification Net (CLF-NET). The CLF-NET is a group convolutional network, which has the rotation equivariance property (Cohen & Welling (2016)), i.e. its features are invariant to rotation of the input cryo-EM map. The Classification Net receives as input a normalized volumetric patch of the cryo-EM map and outputs the labels of each voxel within this patch. CLF-NET has a receptive field of $10\mathring{A}^3$.

The results of CLF-NET are fed to the Segmentation Net (SEG-NET), which performs the final labeling of the voxels. SEG-NET is a U-net CNN (Ronneberger et al. (2015)) which combines the cryo-EM values of the input voxels with the results of CLF-NET on the neighboring voxels. The Segmentation Net has two inputs: a normalized patch of the cryo-EM map and the result of CLF-NET for this patch. The output is the final label of each voxel, which takes into account the neighborhood information. SEG-NET has a receptive field of $15\mathring{A}^3$.

Finally, the Confidence Net (CNF-NET) evaluates the results of SEG-NET. It assigns a binary label (1 -correct, 0 incorrect) to each voxel. Only "correct" voxels are reported, the rest are marked with label 22 (uncovered). The Confidence Net is a rotation equivariant convolutional network. It has a receptive field of $15\mathring{A}^3$ and 24 channels of input (the source cryo-EM map and 23 output channels of the SEG-NET).

3.2 Input Normalisation

Experimental cryo-EM maps suffer from variation in their density value distributions. While voxels with higher density should mark the locations of atoms, density values of similar molecules may differ significantly from one map to another (Lawson et al. (2020)). This is due to lack of standardization combined with multiple factors affecting the density values, such as the microscope type and reconstruction software. Online databases such as EMDataResource (Lawson et al. (2016)) assign different threshold values for each map for the interpretation of atom locations. Figure 3 shows density values for three test maps on a logarithmic scale. Each of the three shown cases has a unique density distribution which is different from one map to another. Moreover the density distribution varies as function of the distance of a voxel to the map center. This variation does not have any recognisable pattern.

Machine learning algorithms usually apply a data normalisation procedure which ensures that the training data and the test data are of similar distribution. Cascaded-CNN (Moritz et al. (2019)) has a five stage normalisation algorithm based on ordered statistics and manually selected thresholds. Emap2Sec (Maddhuri Venkata Subramaniya et al. (2019)) normalises each voxel in the map to the range $0-1$ by a linear transformation. AAnchor (Rozanov & Wolfson (2018)) adjusts the mean and standard deviation of each sample box of size $10\mathring{A}^3$. Normalising each sample box prevents using a fully convolutional network on the whole map and therefore is time consuming. Having said that, note that normalising a whole map is not efficient, since density distribution changes within a map. Moreover, a typical map has a large "background" volume which does not contain any atoms. SegmA applies a trade-off approach. A

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.25.453685; this version posted July 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

map is divided into patches of size N^3_{patch}, where each patch is normalised to mean value of 0.5 and standard deviation of 0.15. A fully convolutional neural network is applied separately on each patch. Large values of N_{patch} result in fast running time at the cost of density distribution variability. Based on empirical estimations, SegmA uses $N_{patch} = 45\mathring{A}$.

3.2.1 Group CNN for the Amino Acids Classification

A function $f(x)$ is said to be **equivariant** to a function g if $f(g(x)) = g(f(x))$. A function $f(x)$ is said to be **invariant** to a function g if $f(g(x)) = f(x)$. Note that the matrix convolution operation is equivariant to translation: a shift in the input leads to a corresponding shift in the output. Due to the translation equivariance, conventional CNNs classify structural patterns regardless of their location without additional effort.

However many classification tasks, especially in 3D, should also accommodate the *rotational symmetry* property: a rotation of an object should not change its classification. Moreover, sometimes rotational symmetry holds for local volume fragments, for example, a rotation of an amino acid by 90° should not alter its classification. The most popular approach to deal with angular rotations is an augmentation of the training dataset with rotated versions of the input. In this case a CNN "duplicates" learning features for "all" possible rotations. The obvious disadvantages of such an approach are more critical in 3D, due to the large input data size and the 3D space of rotations.

Group convolutional neural networks (G-CNNs), introduced in Cohen & Welling (2016), have an architecture which exploits the rotational symmetry. A layer in a G-CNN performs operations which are equivariant to rotations, i.e. $f^i_{G-CNN}(g_{rot}(x)) = g_{rot}\left(f^i_{G-CNN}(x)\right)$, where g_{rot} is a rotation function and f^i_{G-CNN} is i layer of the G-CNN network. A 3D classification G-CNN processes a 3D input matrix and outputs a scalar label, therefore G-CNN is **rotation invariant**: $f_{G-CNN}(g_{rot}(x)) = f_{G-CNN}(x)$ where $f_{G-CNN} : R^{N\times N\times N} \rightarrow \{1, \ldots, N_{labels}\}$ is a G-CNN classification function.

A G-CNN layer is built on the basis of a symmetry group of a discrete set of rotations. Conventionally, a discrete set of rotations by 90° in each of the 3 dimensions is used, because those are actually permutations of 3D matrix elements and no interpolation and matrix multiplication are required. Table 1 lists the rotation groups used in this work.

From the implementation point of view a group convolution layer outputs N_{rot} 4D matrices of dimension $N_{chan\ out} \times L \times W \times H$, where $N_{chan\ out}$ is the number of output channels and L, W, H are the length, width and height of a filter. Cohen & Welling (2016) provide extended explanation of group CNNs, including the theoretical basis and implementation issues. For understanding group CNNs in three dimension we recommend Worrall & Brostow (2018) and Winkels & Cohen (2019). SegmAs G-CNN implementation was based on the open source code of Romero et al. (2020) and Worrall & Brostow (2018).

3.3 Networks Architecture

The architecture of the Classification Net is presented in Figure 1a. CLF-NET assigns a label to the center of an $11\mathring{A}^3$ cube. CLF-NET has 22 label types: 20 amino acids, nucleotide and background.

The architecture of the Segmentation Net (SEG-NET) is presented in Figure 1b. SEG-NET utilizes the U-net network architecture. First proposed for the analysis of 2D biomedical images by Ronneberger et al. (2015), nowadays U-net is one of the most powerful architectures in image segmentation tasks, both in 2D and 3D (Siddique et al. (2020)). Typical U-net architecture consists of a contraction path and an expansion path. The contraction path, also known as the encoder resembles a convolution network and provides classification information. SEG-NET has two contraction paths: one is a conventional sequence of convolutional layers, and the other is the pretrained CLF-NET. The outputs of the two paths are concatenated and fed into the last layer of the encoder. The expansion path or the decoder consists of a sequence of deconvolution (or inverse convolution) layers. An input to a deconvolution is the concatenation of output from previous layer with the output from one of the convolution layers from the encoder.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.25.453685; this version posted July 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

The Confidence Network (CNF-NET) is a typical group CNN and its architecture is like that of CLF-NET. The input is a 3D map with 23 channels, which is the normalized patch of a cryo-EM map concatenated with the result of SEG-NET (22 pseudo-probabilities). The output is a binary label which asserts the SEG-NET labeling: 1 -correct, 0 incorrect.

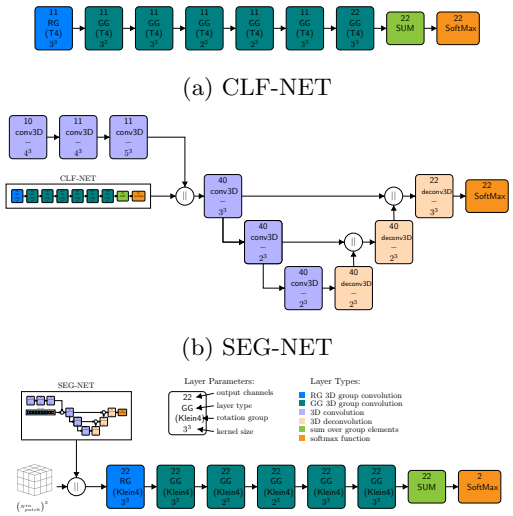
All networks use the rectified linear (ReLU) activation function. The networks are implemented as Fully Convolutional Networks (FCNs, Long et al. (2015)), so they run on a variable input size. FCN architecture reduces the running time of SegmaA: training and testing are done on a large map patch and no sampling or sliding window is required.

We first train the rotation equivariant CLF-NET, afterwards SEG-NET training starts from the pretrained values of CLF-NET. The CNF-NET is trained last on the output of SEG-NET.

3.4 Data Processing

The **preprocessing** of the training data consists of cropping, augmentation, interpolation, normalization and labeling. Test data preprocessing consists of interpolation and normalization. Given a cryo-EM map and a fitted atomic model, the cropping procedure crops out the cryo-EM map edges which are far from the fitted model. Afterwards each map-structure tuple is copied 10 times with a random rotation. This data augmentation is necessary, since rotation equivariance of a group-CNN is only for a discrete set of rotations. The interpolation phase modifies the voxels size of a map to $1\mathring{A}^3$. The labeling procedure assigns to each voxel one of the 22 labels (see Figure 4 for details). An amino acid type (or nucleotide) is assigned to a voxel which is in the proximity of an atom of the residue of this type. A voxel with no residue atom in its proximity gets the label 0 - background.

In the final stage the SEG-NET labeling of the map is modified according to the CNF-NET output. Namely, voxels marked by CNF-NET as "correct" are reported with the SEG-NET labeling, the rest receive the label "uncovered".



(a) CLF-NET

(b) SEG-NET

(c) CNF-NET

Figure 1: (a) Classification Net architecture based on T4 group. This is a typical group CNN network. A GG convolution layer performs convolution on a group of N_{rot} rotations. The very first layer is a RG convolution which inputs a 3D map and outputs a group. The sequence of GG layers is followed by a summation layer. (b) Segmentation Net architecture based on U-net approach. (c) Confidence Net: equivariant network based on the Klein4 group

Klein	4 Rotations	Set of all 180° rotations around x, y, z axes
T4	12 Rotations	Tetrahedral group. Set of all two-axes rotations by 90^0
S4	24 Rotations	Set of all 90° rotations around x, y, z axes

Table 1: Rotation Groups Implemented in CONF-NET

4 Experimental Results

SegmA was trained on 15 experimental cryo-EM maps of resolutions $2.9\mathring{A}$ -$3.1\mathring{A}$ and tested on three cryo-EM maps - see Table 2.

The Classification Net was trained and tested in 4 different configurations. Three group convolution networks for each of the symmetry groups (Klein,

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.25.453685; this version posted July 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

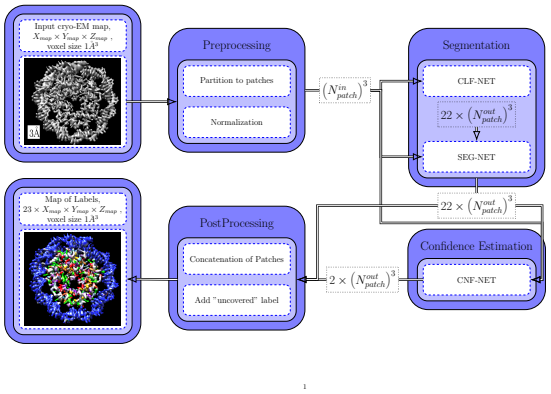


Figure 2: SegmA's processing scheme. In the **preprocessing** phase a input map is partitioned to overlapping patches and each patch is normalized. **Classification Net** assigns primary labels to each voxel, except for the boundary regions of a patch. **Segmentation Net** assigns final labels based on the primary labeling of nearby voxels. **Confidence Net** analyses the result of SEG-NET and marks voxels whose labeling is of low confidence . In the **postprocessing** phase all patches are concatenated in a full size map. Voxels marked as "uncorrect" by CNF-NET are reported with the label "uncovered" (22). The rest are reported with the labels assigned by SEG-NET.

T4 and S4) as well as for the AAnchor configuration without rotation equvariance were tested. All four configurations had a similar amount of learnable parameters. Table 3 summarizes the performance of CLF-NET and SEG-NET for the four different configurations of CLF-NET. Precision and recall values were calculated for four amino acid types and nucleic acids. One can see that the performance of group convolutional networks is better then performance of the regular CNN network. The performance degradation of the S4 group may be explained by training difficulties. In this configuration each layer has 576 matrices: 24 sets of 24 rotated matrices. The contribution of SEG-NET is notable mostly in the recall values, i.e., more voxels of given amino acid types are detected. The rest of this section shows the results for the T4 group used in CLF-NET.

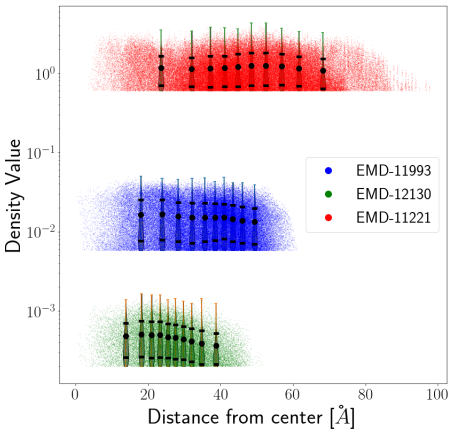


Figure 3: Density value distribution for three test cryo-EM maps. Errorbars show mean and standard deviation value

Segmentation results of the three test map are shown in Figures 8, 9, 10. The segmentation accuracy was calculated for every voxel of value above the recommended threshold. Multilabel classification performance is evaluated by a **confusion matrix**. An entry $a_{i,j}$ of a **confusion matrix** A is defined as the ratio $a_{i,j} = \frac{T^j_j}{N_j}$, where N_j is the number of input cubes labeled j and T^i_j is the number of inputs labeled j with predicted label i. An ideal classification algorithm will result in $a_{i,i} = 1$ and $a_{i,j} = 0$ if $i \neq j$. We also refer to $a_{i,i}$ as the **total accuracy** of label i. SEG-NET achieves performance of about 70% for nucleotides and several amino acids: ALA, ARG, LEU, VAL, TYR, ARG.

We observed that labeling errors often occur between amino acid types of similar physico-chemical properties. 22 labels where grouped into seven *classes* (see Table 4): five classes of amino acids with similar physico-chemical, nucleic acids class and a class for background and uncovered voxels. As expected, SegmAs precision is considerably higher on a segmentation to classes - Figures 8f, 9f, 10f. For example a voxel of an aliphatic amino acid can be estimated with accuracy of about 80%.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.25.453685; this version posted July 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

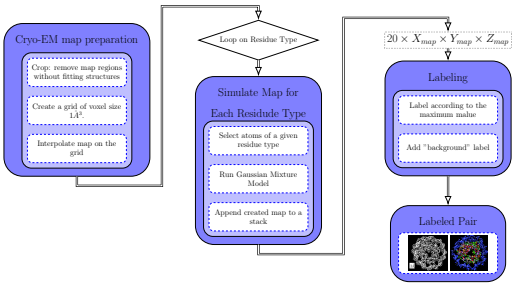


Figure 4: Labeling Procedure: First, the map is cropped and interpolated to voxel size of $1\mathring{A}^3$. Afterwards the pdb2mrc routine (Tang et al. (2007)) is applied only for atoms of a specified residue. pdb2mrc creates a 3D density map which is Gaussian Mixture, where a Gaussian centers are the atom centers and amplitudes are proportional to the atomic number. A residue type whose GMM map has a maximum value at an (i, j, k) voxel is the voxel label. Voxels with a maximal GMM value below a threshold are labeled as background (0).

It has been known from years of human eye examination of electron density maps that some amino acids can be misinterpreted. For example Branden & Tooze (2012) mention some "indistinguishable" pairs: THR-VAL, ASP-ASN, GLU-GLN, HIS-PHE and others. SegmA experience on cryo-EM maps is slightly different: most of the mislabeling occurs between amino acids of similar physico-chemical properties. Only LEU-ILE and THR-VAL have a considerably high of diagonal values in the confusion matrix.

The role of CNF-NET is to filter out false labeled voxels. Figure 5 shows the segmentation accuracy values for different map density levels. While SEG-NET shows degrading performance at intermediate density values, CNF-NET filters out most of the false labeled voxels and the accuracy of reported voxels remains high. An isosurface display of Chimera (Pettersen et al. (2004)) is given in Figure 6. At high density values, which represent protein backbone and atom centres, SegmA reaches nearly 80% precision (see Figure 6 for an isosurface display of Chimera). 3D visualization programs such as Chimera, Pymol or Coot enable a user to control the displayed surface by selecting and appropriate density threshold value. By increasing the shown map threshold, a viewer can obtain more precise segmentation at the cost of shown structure details (Figure 6).

Another insight into the SegmA net accuracy can be obtained by clipping an isosurface and showing the obtained cut (Figure 7). Uncovered regions appear mostly in a residue periphery, i.e., boundary between residues or a boundary between a residue and the background. The degraded performance of SegmA in a residue periphery is caused mainly by molecular flexibility and insufficient precision of the training database. Since a cryo-EM map is actually an average of a large number of flexible molecules, the residue periphery often belongs to more than one amino acid or nucleotide. Moreover, due to errors in atoms position, the periphery regions may have ambiguous labels in the training database.

EMD 12130	Res. $2.9\mathring{A}$	Coronavirus Spike from Smuggled Guangdong Pangolin Wrobel et al. (2021)
EMD 11221	Res. $3\mathring{A}$	Nucleosome with regulatory linker of ALC1 Lehmann et al. (2020)
EMD 11993	Res. $3.1\mathring{A}$	Elongating SARS-CoV-2 RNA-dependent RNA polymerase Kokic et al. (2021)

Table 2: Experimental maps used in SegmA performance evaluation

5 Conclusions

SegmA is a instance segmentation approach to obtain structural insight into a cryo-EM map. SegmA is a cascade of three convolution networks. A fully convolutional U-net architecture enables rapid (within

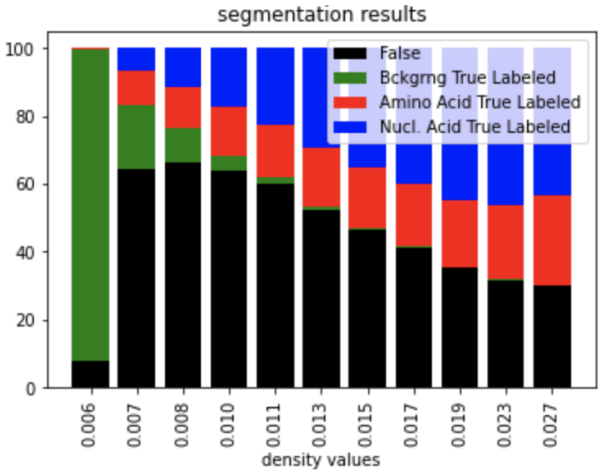
bioRxiv preprint doi: https://doi.org/10.1101/2021.07.25.453685; this version posted July 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

		Precision (Recall) %				
		LEU	PRO	VAL	ARG	NCL
CNN	CLF-NET	39(71)	39(44)	50(46)	24(53)	56(82)
	SEG-NET	49(66)	40(53)	46(42)	30(56)	57(68)
Klein	CLF-NET	58(58)	39(46)	44(56)	35(54)	53(91)
	SEG-NET	62(69)	49(53)	42(69)	42(68)	54(98)
T4	CLF-NET	62(64)	58(52)	56(46)	32(67)	57(87)
	SEG-NET	60(66)	54(65)	58(47)	40(70)	55(94)
S4	CLF-NET	42(43)	41(41)	50(21)	24(44)	55(68)
	SEG-NET	44(74)	60(55)	65(42)	25(78)	59(77)

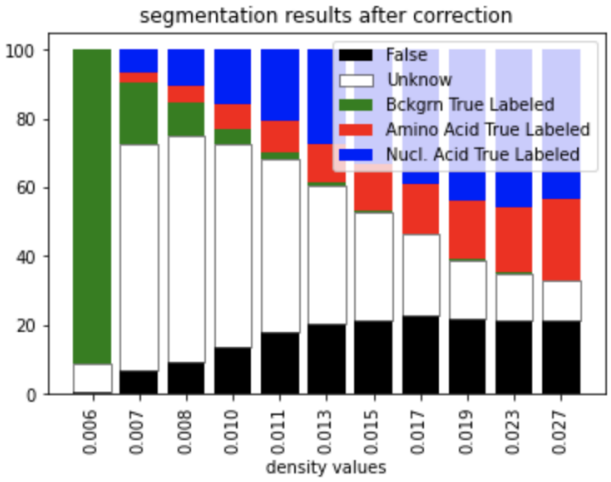
Table 3: CLF-NET performance for various 3D groups. $Precision = \frac{TP}{TP+FP}$, $Precision = \frac{TP}{TP+FN}$. TP is a number of true positives. FP is a number of false positives. FN is a number of false negatives. CNN denotes convolution neural network without rotations. Klein - the group of four 180° rotations. S4 is the group of all possible 90° rotations, 24 members. T4 is the tetrahedral group which consists of 12 two axis rotation by 90°

0. INLBL	Background, Unconfident
1. NCL	Nucleic acids.
2. POS	Basic amino acids: LYS, ARG, HIS.
3. NEG	Acidic amino acids: ASP,GLU.
4. PLR	Polar, uncharged amino acids: SER, THR, CYS, MET, ASN, GLN.
5. ARM	Nonpolar, aromatic amino acids: PHE, TYR, TRP.
6. ALP	Nonpolar, aliphatic amino acids: GLY , ALA, VAL, LEU, ILE, PRO".

Table 4: Segmentation Classes



(a) Segmentation Net Accuracy



(b) Effect of Confidence Net

Figure 5: Segmentation Accuracy for EMD-11221 for different density values. (a) The SQ-NET accuracy is better for voxels for of high density values, except for very low densities, where most voxels are labeled as background. (b) CNF-NET correction. White bar shows the share of the uncovered region.

seconds) processing of a cryo-EM 3D map. Considerable improvement in the segmentation precision was achieved by using a group convolutional network which is the heart of the algorithm. This type of CNN has the rotational equivariance property, which leverages the native rotation invariance of a molecule. An instance segmentation task has natural performance limits when applied to a cryo EM map. Some amino acids types cannot be revealed at a given resolution. Moreover, an experimental cryo-EM map has regions of degraded resolution. SegmA detects those regions and marks them as "uncovered", which

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.25.453685; this version posted July 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

results in a considerably higher performance in the "covered" regions. SegmA estimates a residue type at the voxel level, presenting the user a coloured 3D image. SegmA also posses an advantage of a segmetation approach: a residue pose and boundary are seen in addition to the residue position and type. SegmA also has the ability to distinguish between protein residues and nucleotides.

References

Baker, M. L., Abeysinghe, S. S., Schuh, S., Coleman, R. A., Abrams, A., Marsh, M. P., Hryc, C. F., Ruths, T., Chiu, W. & Ju, T. (2011), 'Modeling protein structure at near atomic resolutions with Gorgon.', *Journal of structural biology* **174**(2), 360–373.

Branden, C. I. & Tooze, J. (2012), *Introduction to protein structure*, Garland Science.

Callaway, E. (2017), 'Molecular-imaging pioneers scoop Nobel', *Nature* **550**(7675), 167.

Chen, M., Baldwin, P. R., Ludtke, S. J. & Baker, M. L. (2016), 'De Novo modeling in cryo-EM density maps with Pathwalking', *Journal of Structural Biology* .

Cohen, T. S. & Welling, M. (2016), 'Group equivariant convolutional networks', *33rd International Conference on Machine Learning, ICML 2016* **6**, 4375–4386.

DiMaio, F., Leaver-Fay, A., Bradley, P., Baker, D. & André, I. (2011), 'Modeling Symmetric Macromolecular Structures in Rosetta3', *PLOS ONE* **6**(6), 1–13.
URL: *https://doi.org/10.1371/journal.pone.0020450*

Joseph, A. P., Lagerstedt, I., Jakobi, A., Burnley, T., Patwardhan, A., Topf, M. & Winn, M. (2020), 'Comparing Cryo-EM Reconstructions and Validating Atomic Model Fit Using Difference Maps', *Journal of Chemical Information and Modeling* **60**(5), 2552–2560.

Kokic, G., Hillen, H. S., Tegunov, D., Dienemann, C., Seitz, F., Schmitzova, J., Farnung, L., Siewert, A., Höbartner, C. & Cramer, P. (2021), 'Mechanism of SARS-CoV-2 polymerase stalling by remdesivir', *Nature communications* **12**(1), 1–7.

Kühlbrandt, W. (2014), 'The resolution revolution', *Science* **343**(6178), 1443–1444.

Lasker, K., Dror, O., Shatsky, M., Nussinov, R. & Wolfson, H. J. (2007), 'EMatch: discovery of high resolution structural homologues of protein domains in intermediate resolution cryo-EM maps.', *IEEE/ACM transactions on computational biology and bioinformatics* **4**(1), 28–39.

Lawson, C. L., Berman, H. M. & Chiu, W. (2020), 'Evolving data standards for cryo-EM structures', *Structural Dynamics* **7**(1).

Lawson, C. L., Patwardhan, A., Baker, M. L., Hryc, C., Garcia, E. S., Hudson, B. P., Lagerstedt, I., Ludtke, S. J., Pintilie, G., Sala, R., Westbrook, J. D., Berman, H. M., Kleywegt, G. J. & Chiu, W. (2016), 'EMDataBank unified data resource for 3DEM', *Nucleic Acids Research* .

Lehmann, L. C., Bacic, L., Hewitt, G., Brackmann, K., Sabantsev, A., Gaullier, G., Pytharopoulou, S., Degliesposti, G., Okkenhaug, H., Tan, S. & Others (2020), 'Mechanistic Insights into Regulation of the ALC1 Remodeler by the Nucleosome Acidic Patch', *Cell reports* **33**(12), 108529.

Li, P. N., De Oliveira, S. H., Wakatsuki, S. & Van Den Bedem, H. (2020), 'Sequence-guided protein structure determination using graph convolutional and recurrent networks', *Proceedings - IEEE 20th International Conference on Bioinformatics and Bioengineering, BIBE 2020* pp. 122–127.

Liebschner, D., Afonine, P. V., Baker, M. L., Bunkóczi, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L. W., Jain, S., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Poon, B. K., Prisant, M. G., Read, R. J., Richardson, J. S., Richardson, D. C., Sammito, M. D., Sobolev, O. V., Stockwell, D. H., Terwilliger, T. C., Urzhumtsev, A. G.,

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.25.453685; this version posted July 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Videau, L. L., Williams, C. J. & Adams, P. D. (2019), 'Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix.', *Acta crystallographica. Section D, Structural biology* **75**(Pt 10), 861–877.

Lindert, S., Alexander, N., Wötzel, N., Karakas, M., Stewart, P. L. & Meiler, J. (2012), 'EM-Fold: De Novo Atomic-Detail Protein Structure Determination from Medium-Resolution Density Maps', *Structure* **20**(3), 464–478.

Long, J., Shelhamer, E. & Darrell, T. (2015), 'Fully convolutional networks for semantic segmentation', *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* **07-12-June**, 3431–3440.

Ma, L., Reisert, M. & Burkhardt, H. (2012), 'RENNSH: A Novel α-Helix Identification Approach for Intermediate Resolution Electron Density Maps', *IEEE/ACM Trans. Comput. Biol. Bioinformatics* **9**(1), 228–239.
URL: *https://doi.org/10.1109/TCBB.2011.52*

Maddhuri Venkata Subramaniya, S. R., Terashi, G. & Kihara, D. (2019), 'Protein secondary structure detection in intermediate-resolution cryo-EM maps using deep learning', *Nature Methods* **16**(9), 911–917.

Moritz, S. A., Pfab, J., Wu, T., Hou, J., Cheng, J., Cao, R., Wang, L. & Si, D. (2019), 'Cascaded-CNN: Deep Learning to Predict Protein Backbone Structure from High-Resolution Cryo-EM Density Maps', *bioRxiv* p. 572990.
URL: *www.biorxiv.org/content/10.1101/572990v3*

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. (2004), 'UCSF Chimera - A visualization system for exploratory research and analysis', *Journal of Computational Chemistry* **25**(13), 1605–1612.

Pfab, J., Phan, N. M. & Si, D. (2021), 'DeepTracer for fast de novo cryo-EM protein structure modeling and special studies on cov-related complexes', *Proceedings of the National Academy of Sciences of the United States of America* **118**(2).

Romero, D. W., Bekkers, E. J., Tomczak, J. M. & Hoogendoorn, M. (2020), 'Attentive group equivariant convolutional networks', *arXiv* (2016).

Ronneberger, O., Fischer, P. & Brox, T. (2015), 'U-net: Convolutional networks for biomedical image segmentation', *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **9351**, 234–241.

Rozanov, M. & Wolfson, H. J. (2018), AAnchor: CNN guided detection of anchor amino acids in high resolution cryo-EM density maps, *in* '2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)', pp. 88–91.

Si, D., Ji, S., Nasr, K. A. & He, J. (2012), A machine learning approach for the identification of protein secondary structure elements from electron cryo-microscopy density maps, *in* 'Biopolymers', Vol. 97, pp. 698–708.

Si, D., Nakamura, A., Tang, R., Guan, H., Hou, J., Firozi, A., Cao, R., Hippe, K. & Zhao, M. (2021), 'Artificial intelligence advances for de novo molecular structure modeling in cryo-electron microscopy', *WIREs Computational Molecular Science* **n/a**(n/a), e1542.
URL: *https://doi.org/10.1002/wcms.1542*

Siddique, N., Sidike, P., Elkin, C. & Devabhaktuni, V. (2020), 'U-Net and its variants for medical image segmentation: theory and applications'.
URL: *http://arxiv.org/abs/2011.01118*

Tang, G., Peng, L., Baldwin, P. R., Mann, D. S., Jiang, W., Rees, I. & Ludtke, S. J. (2007), 'EMAN2: An extensible image processing suite for electron microscopy', *Journal of Structural Biology* **157**(1), 38–46.

Terashi, G. & Kihara, D. (2018), 'De novo main-chain modeling for em maps using MAINMAST', *Nature Communications* .

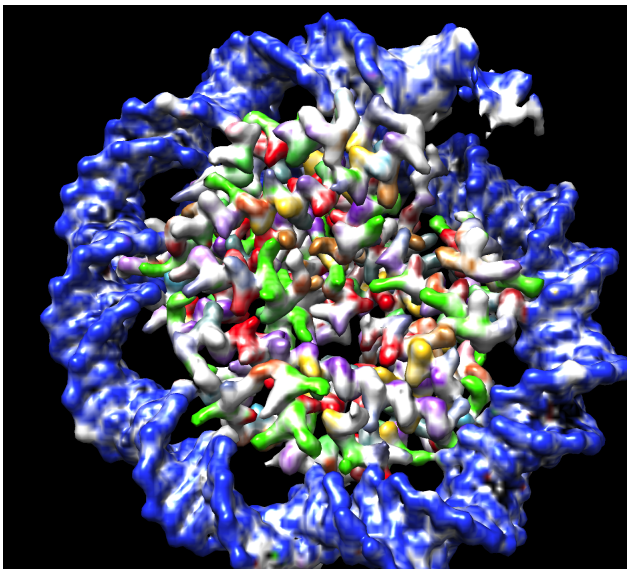
bioRxiv preprint doi: https://doi.org/10.1101/2021.07.25.453685; this version posted July 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Winkels, M. & Cohen, T. S. (2019), 'Pulmonary nodule detection in CT scans with equivariant CNNs', *Medical Image Analysis* **55**, 15–26. **URL:** *https://doi.org/10.1016/j.media.2019.03.010*

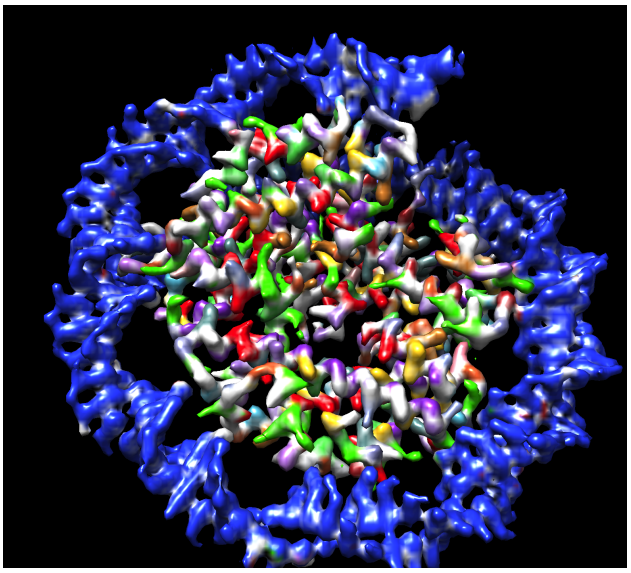
Worrall, D. & Brostow, G. (2018), 'CubeNet: Equivariance to 3D Rotation and Translation', *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **11209 LNCS**, 585–602.

Wrobel, A. G., Benton, D. J., Xu, P., Calder, L. J., Borg, A., Roustan, C., Martin, S. R., Rosenthal, P. B., Skehel, J. J. & Gamblin, S. J. (2021), 'Structure and binding properties of Pangolin-CoV Spike glycoprotein inform the evolution of SARS-CoV-2', *Nature communications* **12**(1), 1–6.

Xu, K., Wang, Z., Shi, J., Li, H. & Zhang, Q. C. (2019), 'A2-Net: Molecular structure estimation from cryo-EM density volumes', *33rd AAAI Conference on Artificial Intelligence, AAAI 2019, 31st Innovative Applications of Artificial Intelligence Conference, IAAI 2019 and the 9th AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019* pp. 1230–1237.



(a) EMD-11221 isosurface for threshold value 0.013. Segmentation precision 66%, coverage 66%.



(b) EMD-11221 isosurface for threshold value 0.019. Segmentation precision 76%, coverage 85%.

Figure 6: EMD-11221 isosurfaces for values 0.013 and 0.019 coloured by an amino acid type. Blue regions indicate nucleotides. Uncovered regions showed by White color. The segmentation precision improves and coverage increases as the threshold value grows at the cost of shown map.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.25.453685; this version posted July 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

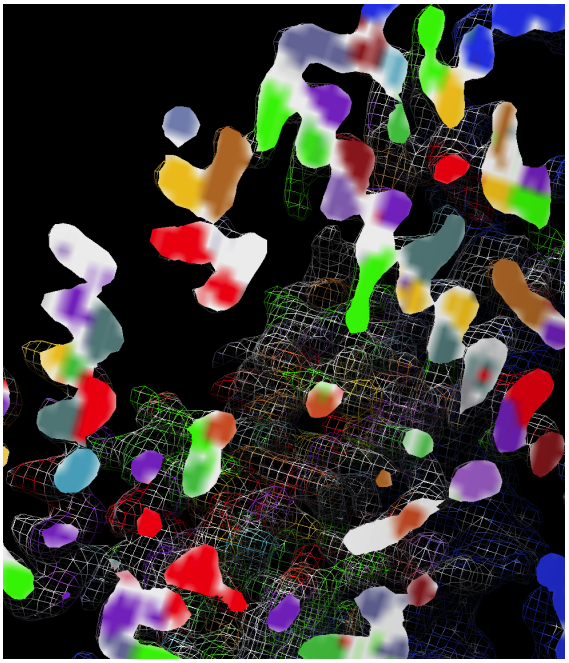
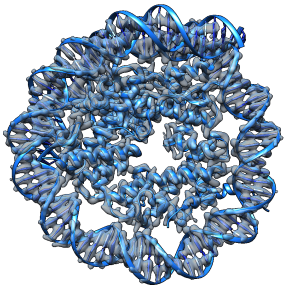
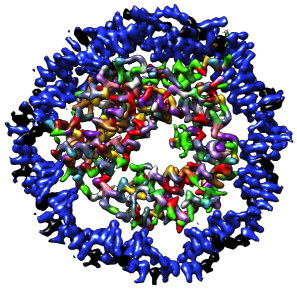


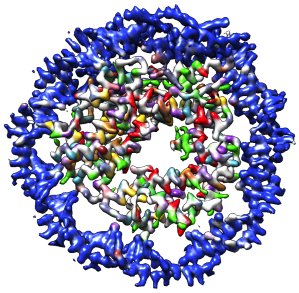
Figure 7: EMD-11221 Clipped Surface. The mesh shows isosurface at density value 0.013. The cutting plane is shown as continuously colored regions. Uncovered (white) regions occur mostly on the boundary between residues or a boundary between a residue and the background .



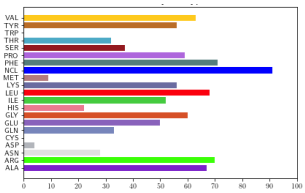
(a) Input cryo-EM map with fitted structure



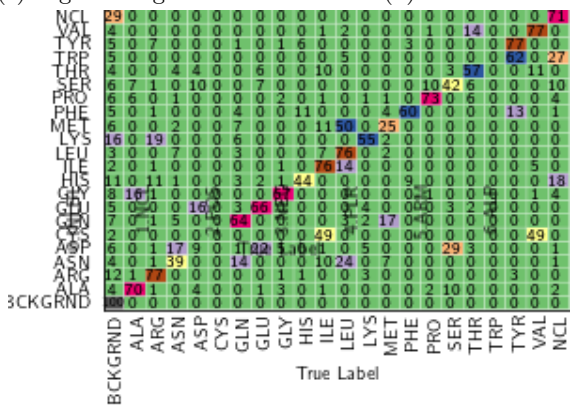
(b) True Segmentation



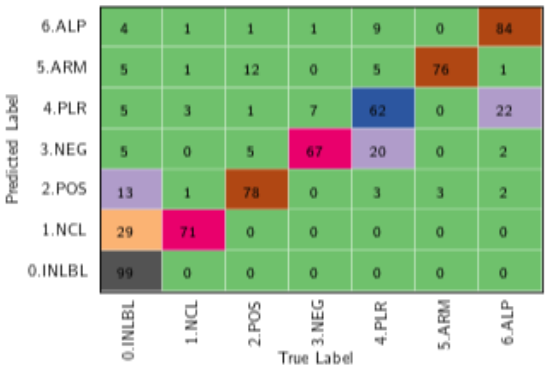
(c) SegmAs Segmentation



(d) Results



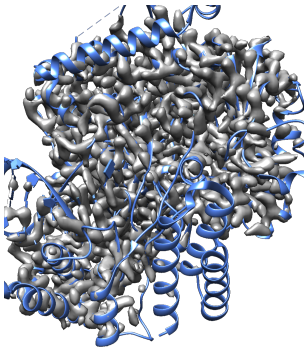
(e) Segmentation to AA



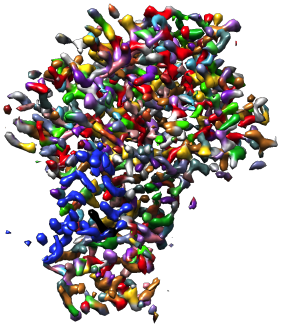
(f) Segmentation to Classes

Figure 8: EMD-11221 Segmentation Results

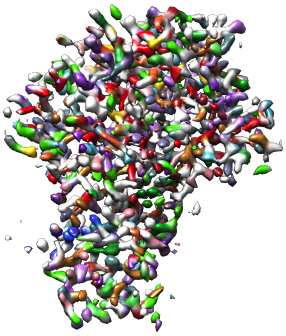
bioRxiv preprint doi: https://doi.org/10.1101/2021.07.25.453685; this version posted July 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.



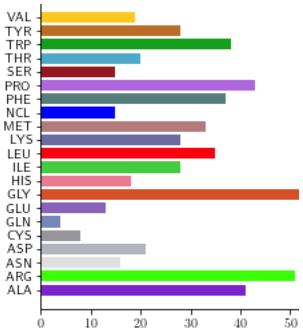
(a) Input cryo-EM map with fitted structure



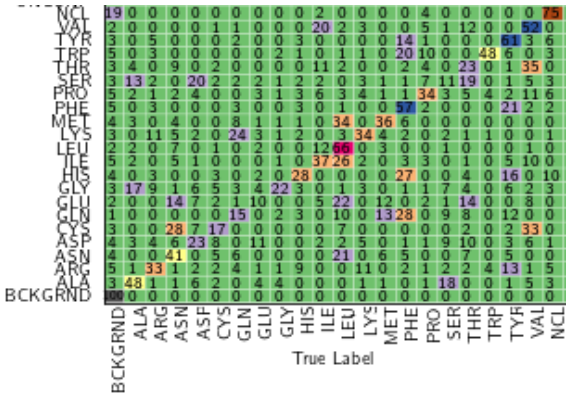
(b) True Segmentation



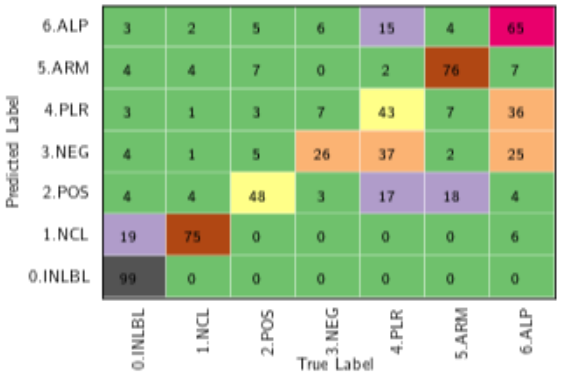
(c) SegmAs Segmentation



(d) Results

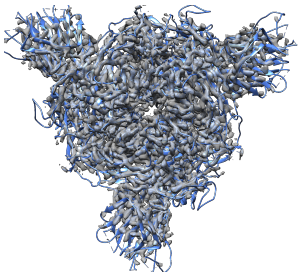


(e) Segmentation to AA

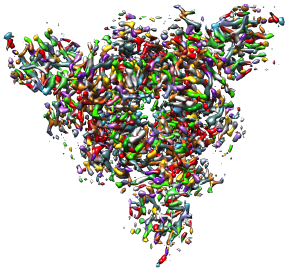


(f) Segmentation to Classes

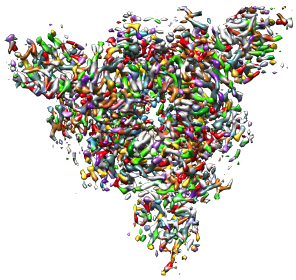
Figure 9: EMD-11993 Segmentation Results



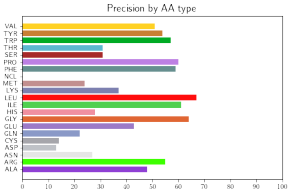
(a) Input cryo-EM map with fitted structure



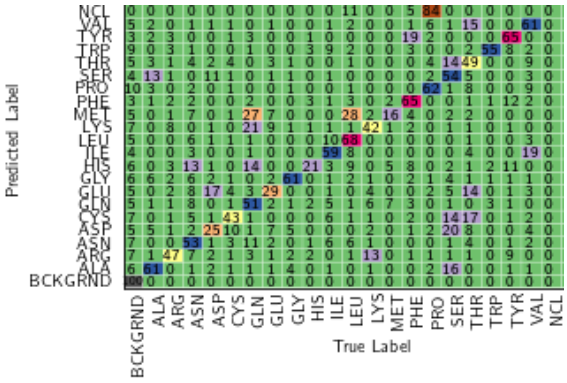
(b) True Segmentation



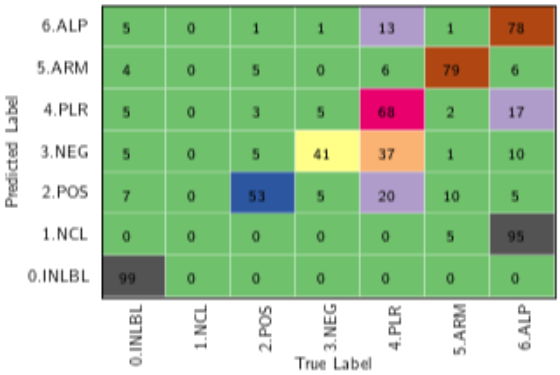
(c) SegmA Segmentation



(d) Results



(e) Segmentation to AA



(f) Segmentation to Classes

Figure 10: EMD-12130 Segmentation Results