



# DiffFit: Visually-Guided Differentiable Fitting of Molecule Structures to a Cryo-EM Map

**Deng Luo**<sup>1</sup>, Zainab Alsuwaykit<sup>1</sup>, Dawar Khan<sup>1</sup>,  
Ondřej Strnad<sup>1</sup>, Tobias Isenberg<sup>2</sup>, and Ivan Viola<sup>1</sup>

1: King Abdullah University of Science and Technology

2: Inria



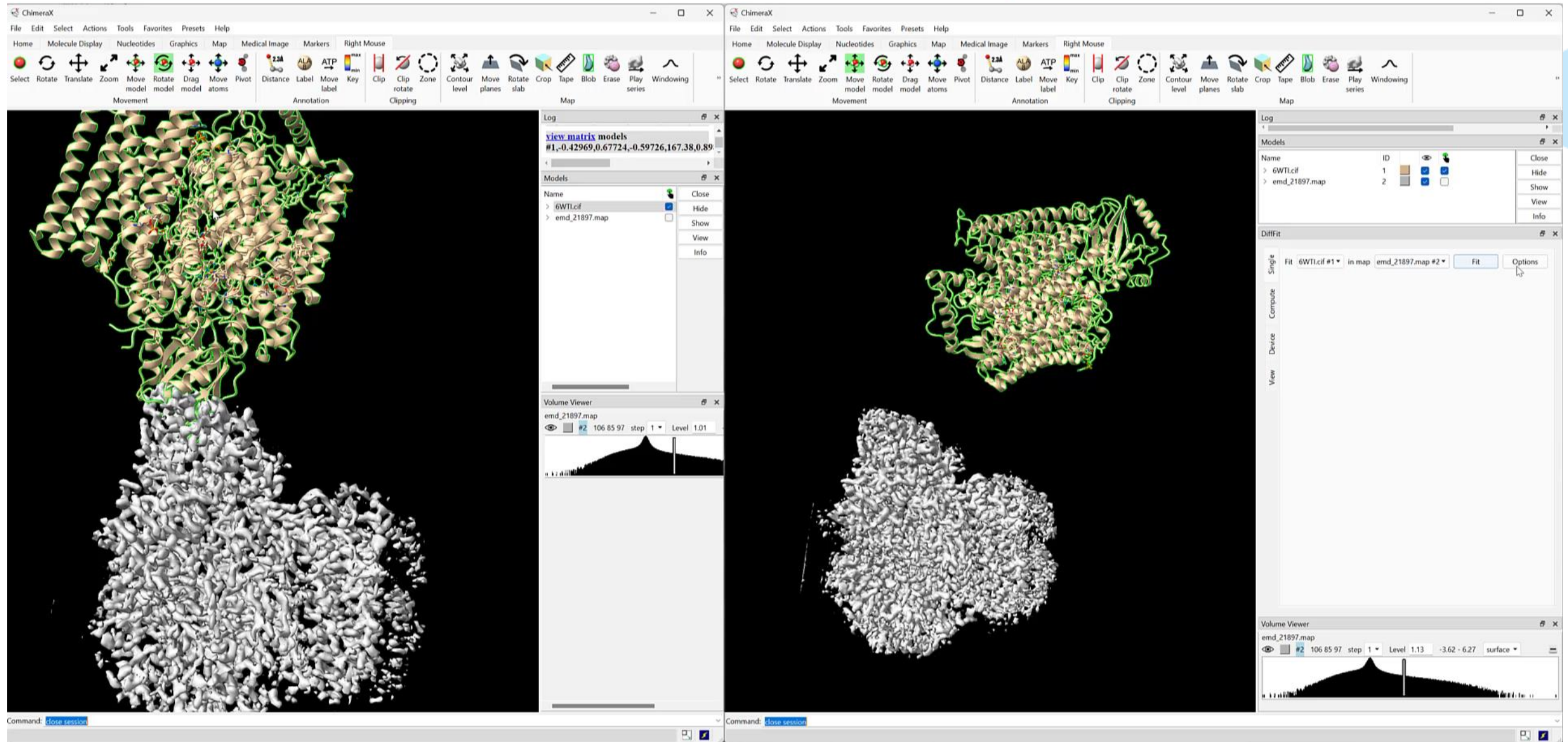
*Inria*



# Introduction

ChimeraX Fit in Map

DiffFit



# Introduction

The screenshot displays the DiffFit software interface. The main window shows a protein structure (stick representation) fitted into a cryo-EM density map (grey surface). The protein is colored by chain, and the density map is shown in grey. The interface includes a menu bar (Home, Molecule Display, Nucleotides, Graphics, Map, Medical Image, Markers, Right Mouse) and a toolbar with various icons for showing/hiding atoms, surfaces, and analysis tools.

The DiffFit panel on the left contains the following settings:

- Input mode: disk file
- Target Volume: posite\transformed\emd\_40589.map
- Structures Folder: composite\transformed\subunits\_cif
- Data Folder: res\_VIS-reproducibility\_search\_1K\_r0
- In contour threshold: 0.42
- Correlation threshold: 0.75
- Clustering - Shift Tolerance: 3.00
- Clustering - Angle Tolerance: 6.00

The Load table below shows the results of the fit:

Id	Mol Id	Hits	Density	Correlation
1	0	25	699.437	0.335
2	0	26	699.425	0.335
3	1	8	230.7	0.294
4	1	5	230.686	0.294
5	2	3	171.862	0.246
6	2	1	171.703	0.245
7	1	4	117.983	0.281
8	1	1	117.623	0.283
9	1	1	116.858	0.282
10	1	1	113.78	0.274
11	1	2	113.355	0.235
12	1	1	111.738	0.234
13	2	3	108.051	0.266
14	2	2	106.713	0.266
15	2	1	105.355	0.261
16	2	2	104.155	0.264
17	2	1	103.047	0.261
18	2	1	96.7633	0.243
19	2	3	93.521	0.227

The Log panel on the right shows the following information:

- Highest metric reached at iter : 12
- select #2.1
- 1 model selected
- open
- D:\GIT\DiffFit\case\_study\_data\8SMK\_composite\transformed\subunits\_cif\8sm
- Chain information for 8smk\_transformed\_chain\_A.cif #1

Chain	Description
A	Protein-arginine deiminase type-4

- Showing cluster ID: 1
- Cluster size: 25
- Highest metric reached at iter : 12

The Models panel shows the following models:

- 8smk\_transformed\_chain\_A.cif (ID 1)
- clusterSpheres (ID 2)
- emd\_40589.map (ID 10)

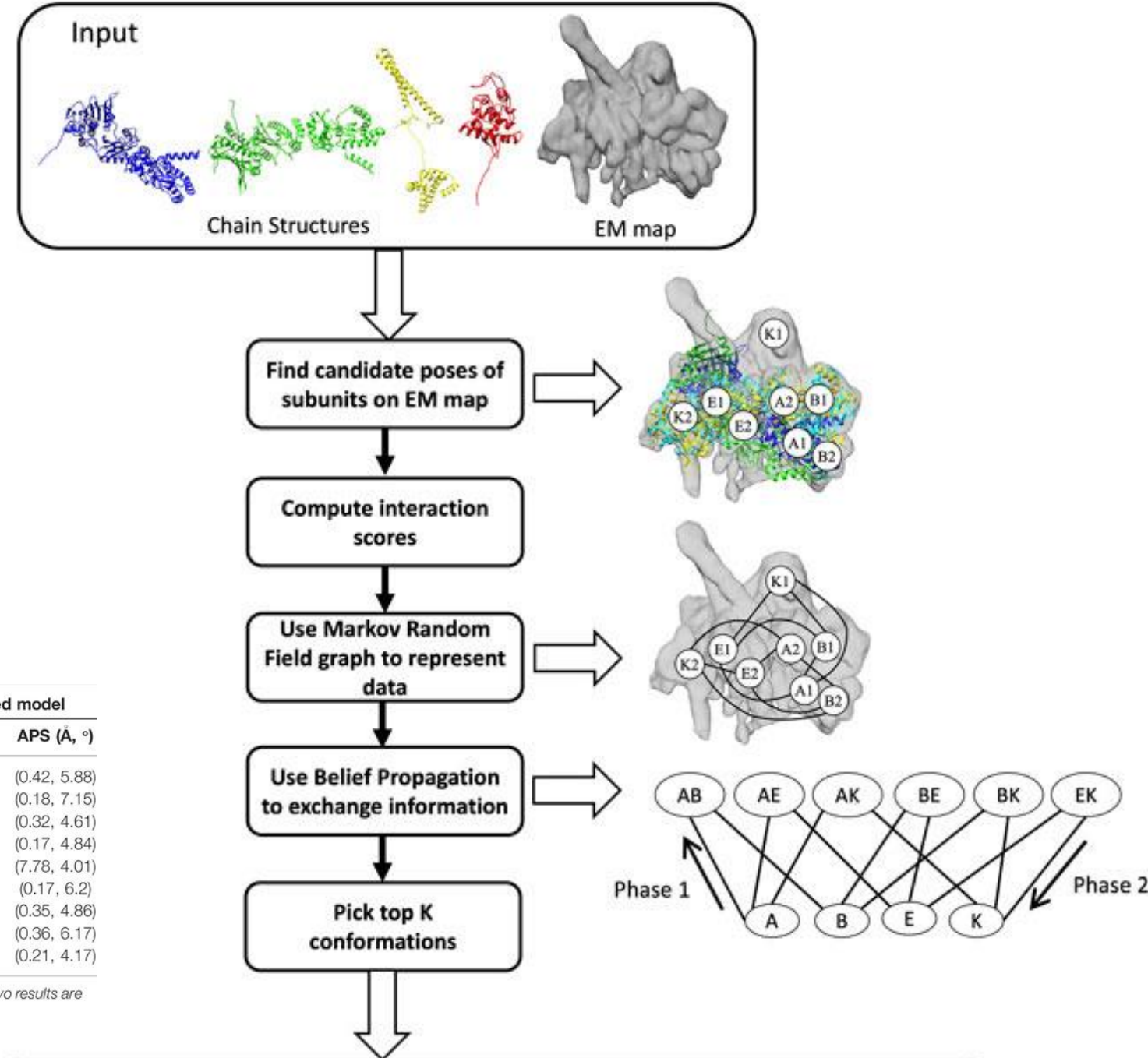
The Volume Viewer panel shows the density map (emd\_40589.map) with a resolution of 4.00 Å and a threshold of 0.0000. The map is shown as a surface, and the protein is shown as a stick model.

The Fit in Map panel shows the fit of the protein into the density map. The correlation is 0.75, and the average map value is 4. The fit is shown as a stick model, and the density map is shown as a surface. The fit is shown as a stick model, and the density map is shown as a surface. The fit is shown as a stick model, and the density map is shown as a surface.



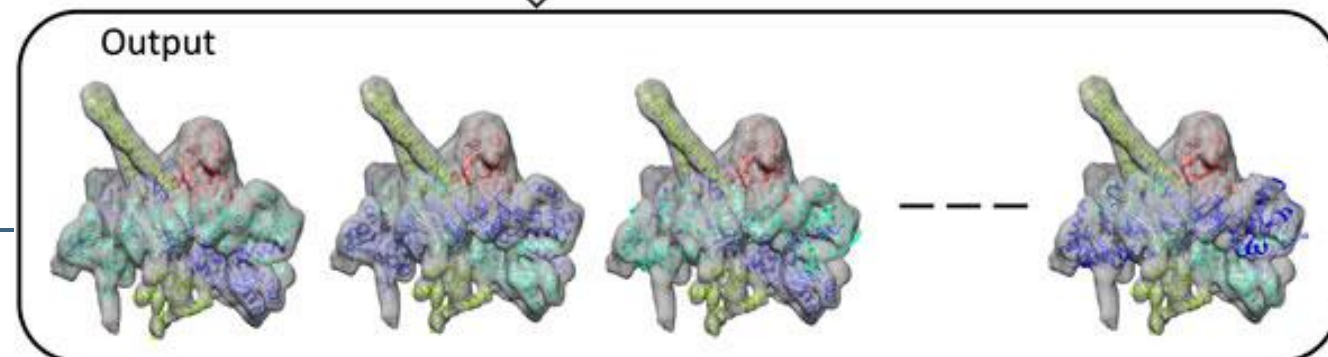
# Related Work

- Alnabati et al.: **MarkovFit**  
*Frontiers in Molecular Biosciences* (2022)
  - Benchmark table



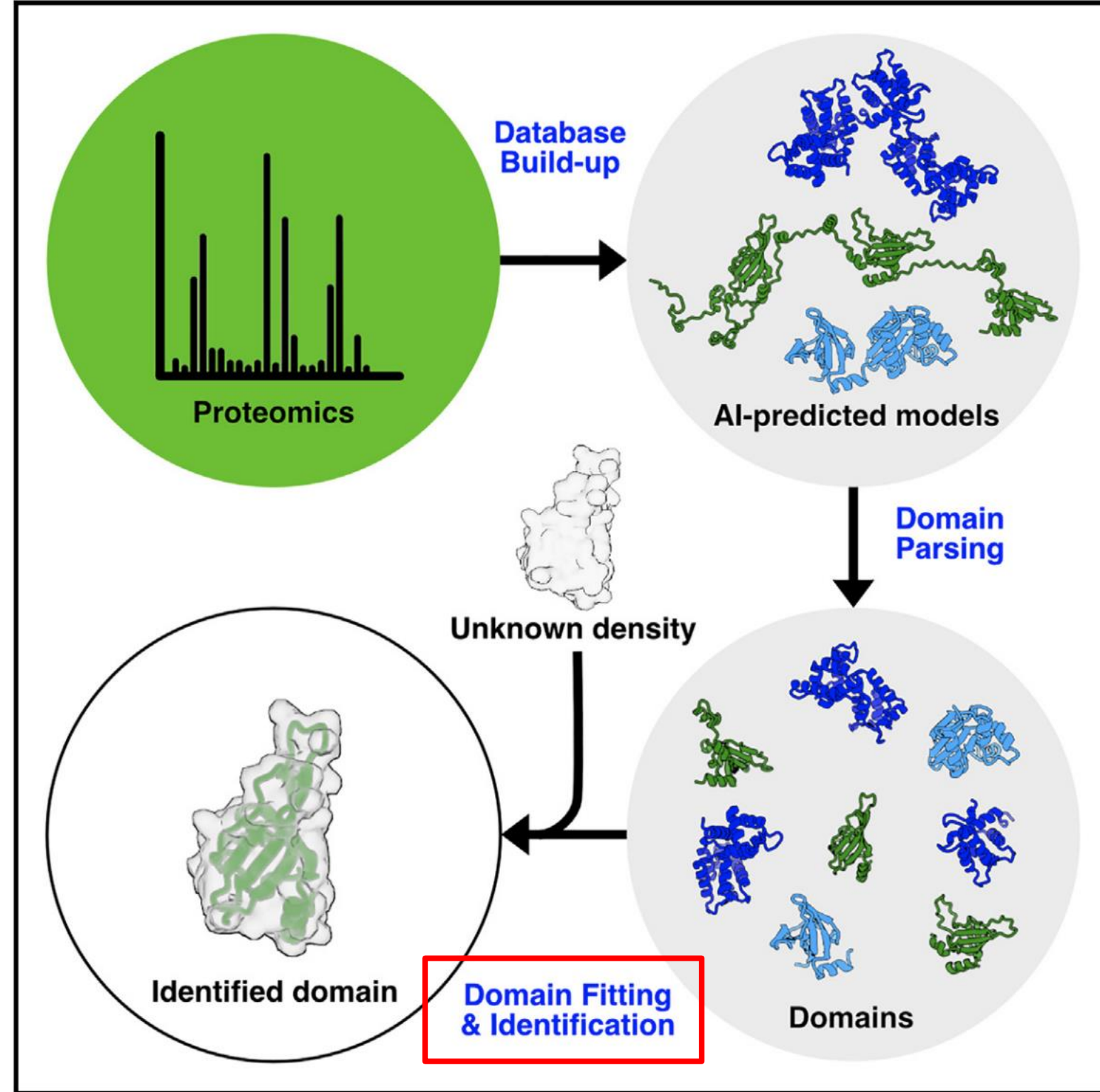
EMDB ID	PDB ID	No. subunits	Res. (Å)	Voxel size (Å)	Best model by RMSD among top 10			Top-scored model	
					Rank	RMSD (Å)	APS (Å, °)	RMSD (Å)	APS (Å, °)
9108	6MEO	3	3.9	1.059	1	1.94	(0.42, 5.88)	1.94	(0.42, 5.88)
13508	7PM0	3	3.6	1.1	1	1.64	(0.18, 7.15)	1.64	(0.18, 7.15)
25368	7SP8	3	2.7	1.08	1	1.29	(0.32, 4.61)	1.29	(0.32, 4.61)
30093	6M5U	3	3.8	1.062	1	2.36	(0.17, 4.84)	2.36	(0.17, 4.84)
21897	6WTI	4	2.38	1.08	1	1.31	(7.78, 4.01)	1.31	(7.78, 4.01)
23827	7MGE	4	3.94	0.94	1	1.87	(0.17, 6.2)	1.87	(0.17, 6.2)
30614	7D8X	4	2.6	1.0825	1	1.96	(0.35, 4.86)	1.96	(0.35, 4.86)
22417	7JPO	5	3.2	1.07	1	2.54	(0.36, 6.17)	2.54	(0.36, 6.17)
25426	7STE	5	2.73	0.826	1	1.74	(0.21, 4.17)	1.74	(0.21, 4.17)

No. subunits, the number of subunits in the structure; Res., reported resolution of the maps; voxel size, the size of the grid voxels of the maps. As we showed in **Table 1**, two results are shown: the best RMSD, model among the top 10-scored models and the top-scored models.



# Related Work

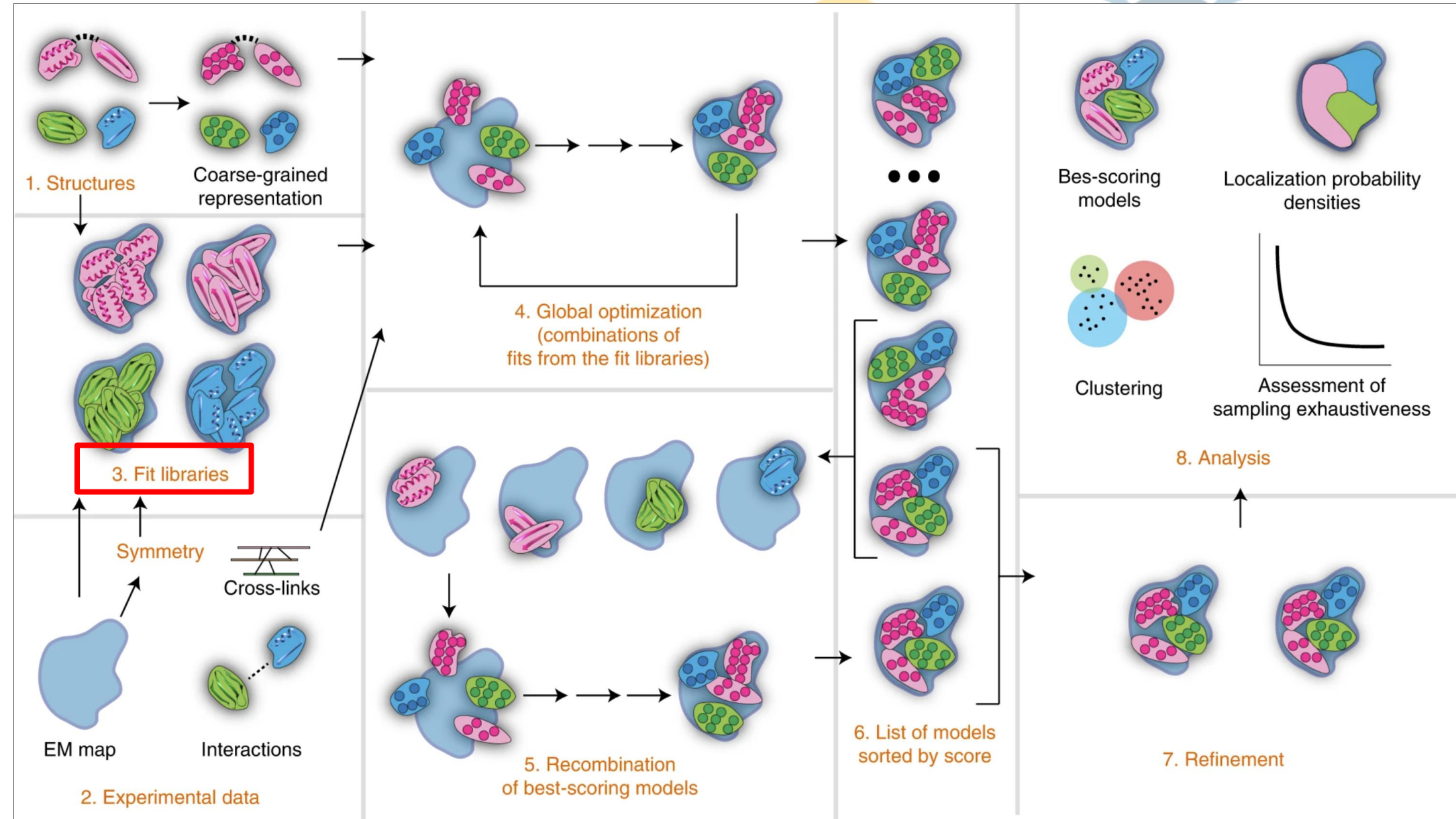
- ChimeraX “**Fit in Map**” functionality
  - Gao et al.:  
**DomainFit**  
*Structure* (2024)



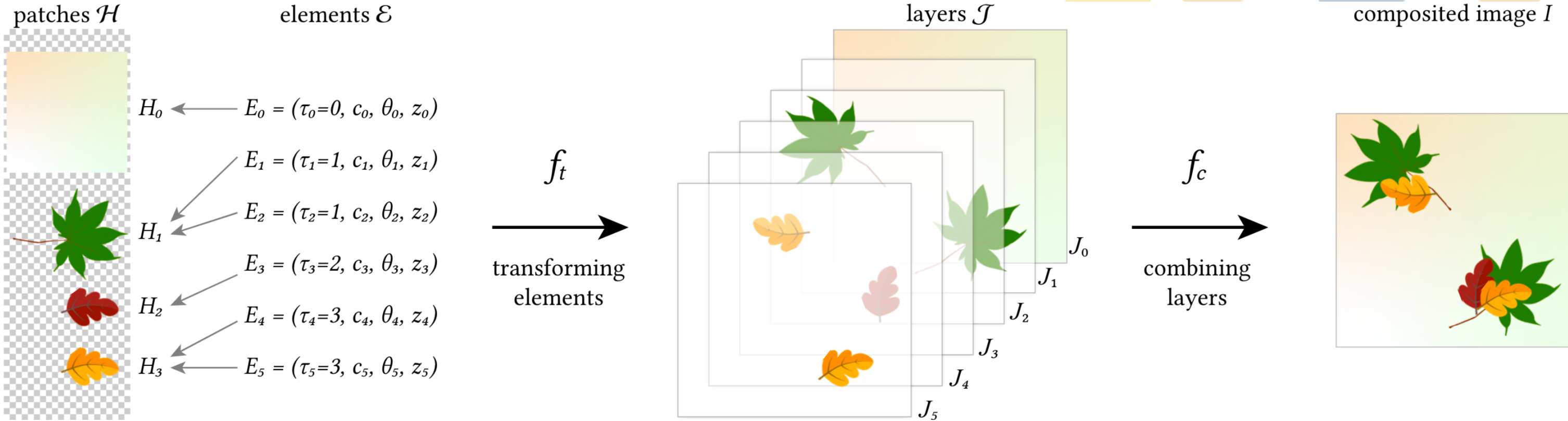


# Related Work

- ChimeraX “**Fit in Map**” functionality
  - Gao et al.: **DomainFit Structure** (2024)
  - Rantos et al.: **Assembleline Nature Protocols** (2022)

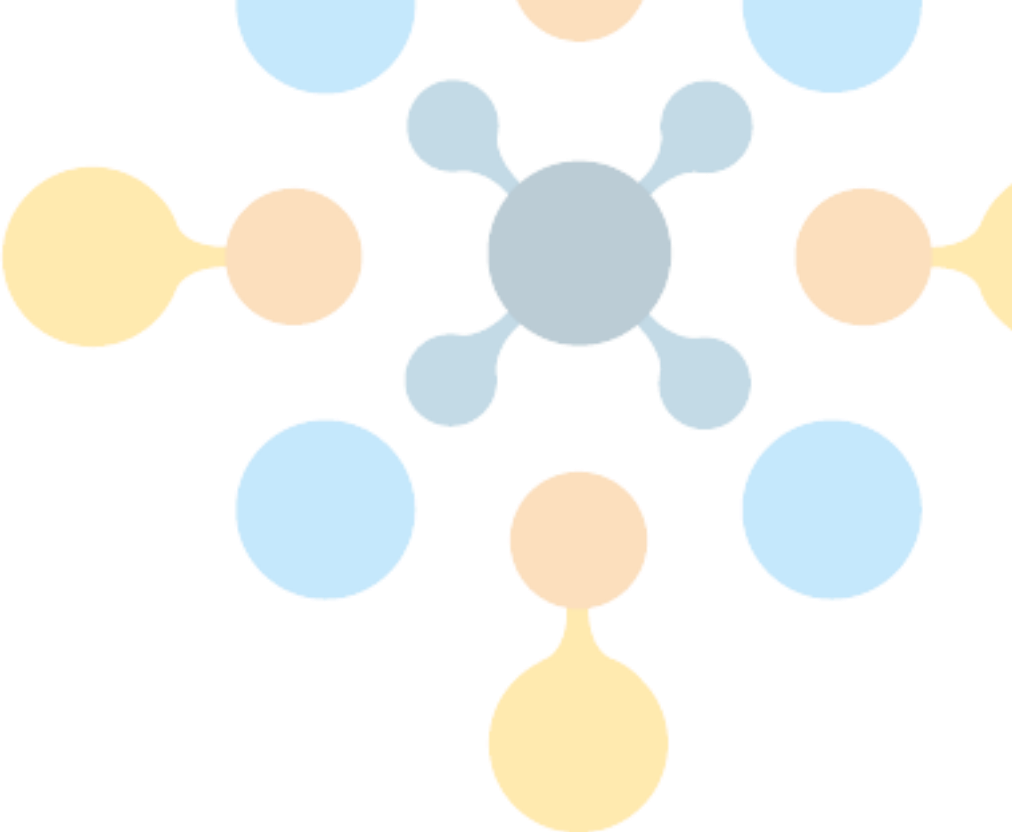


# Inspiration



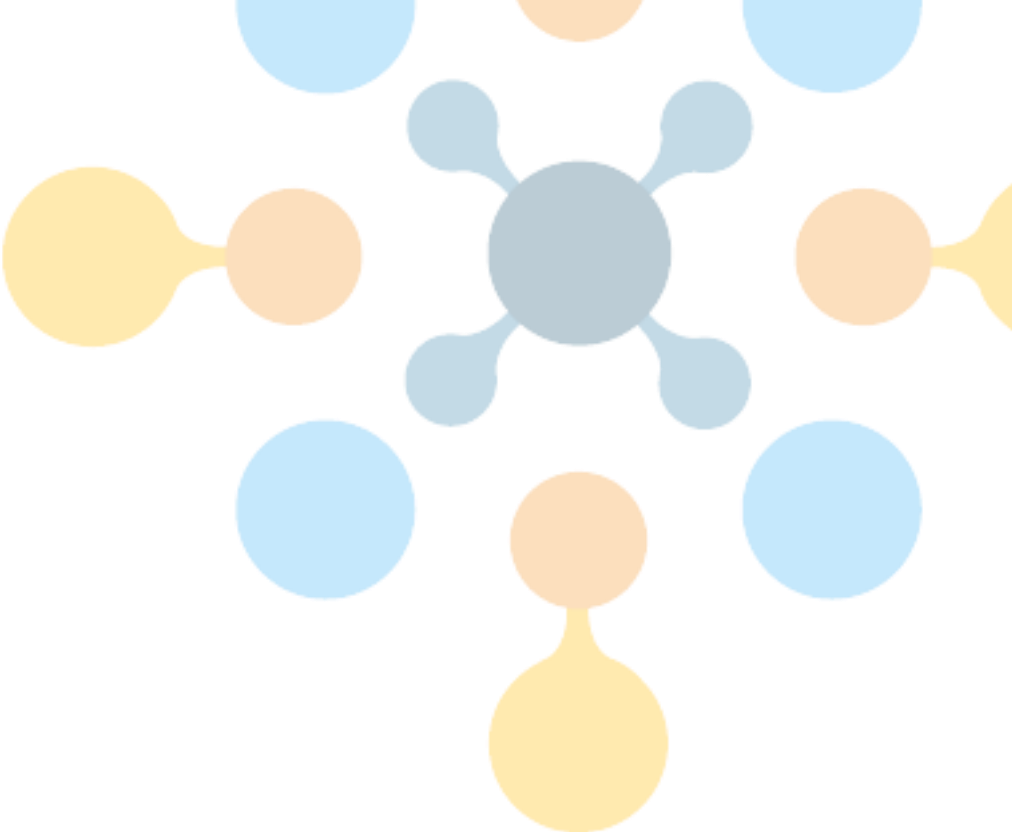
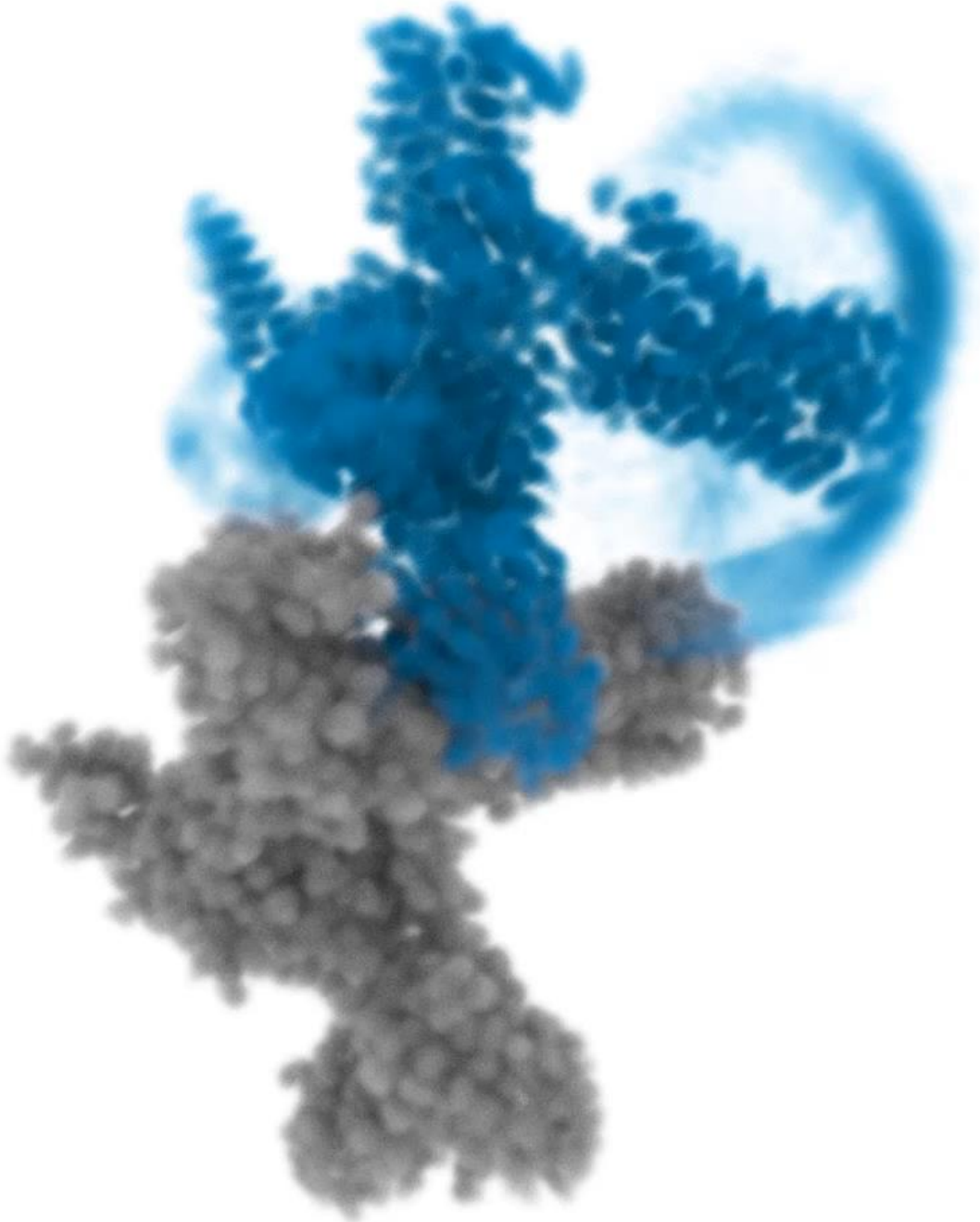
Reddy et al.: Differentiable compositing, *TOG* 2020

# Method

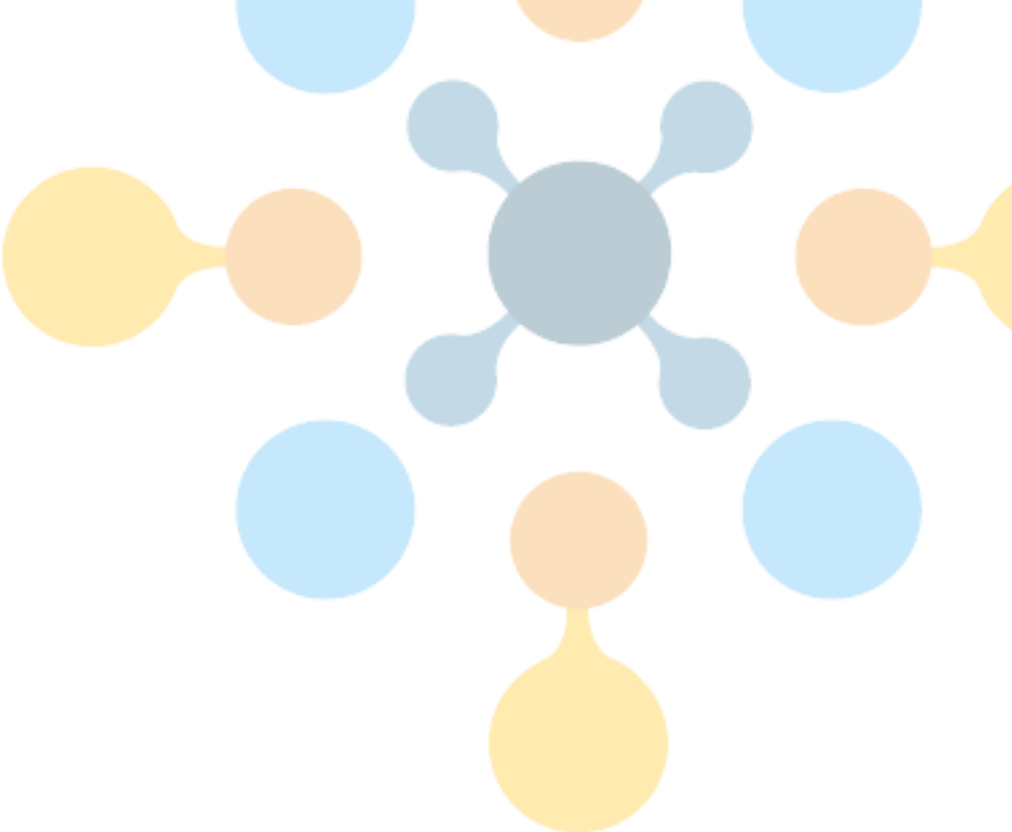
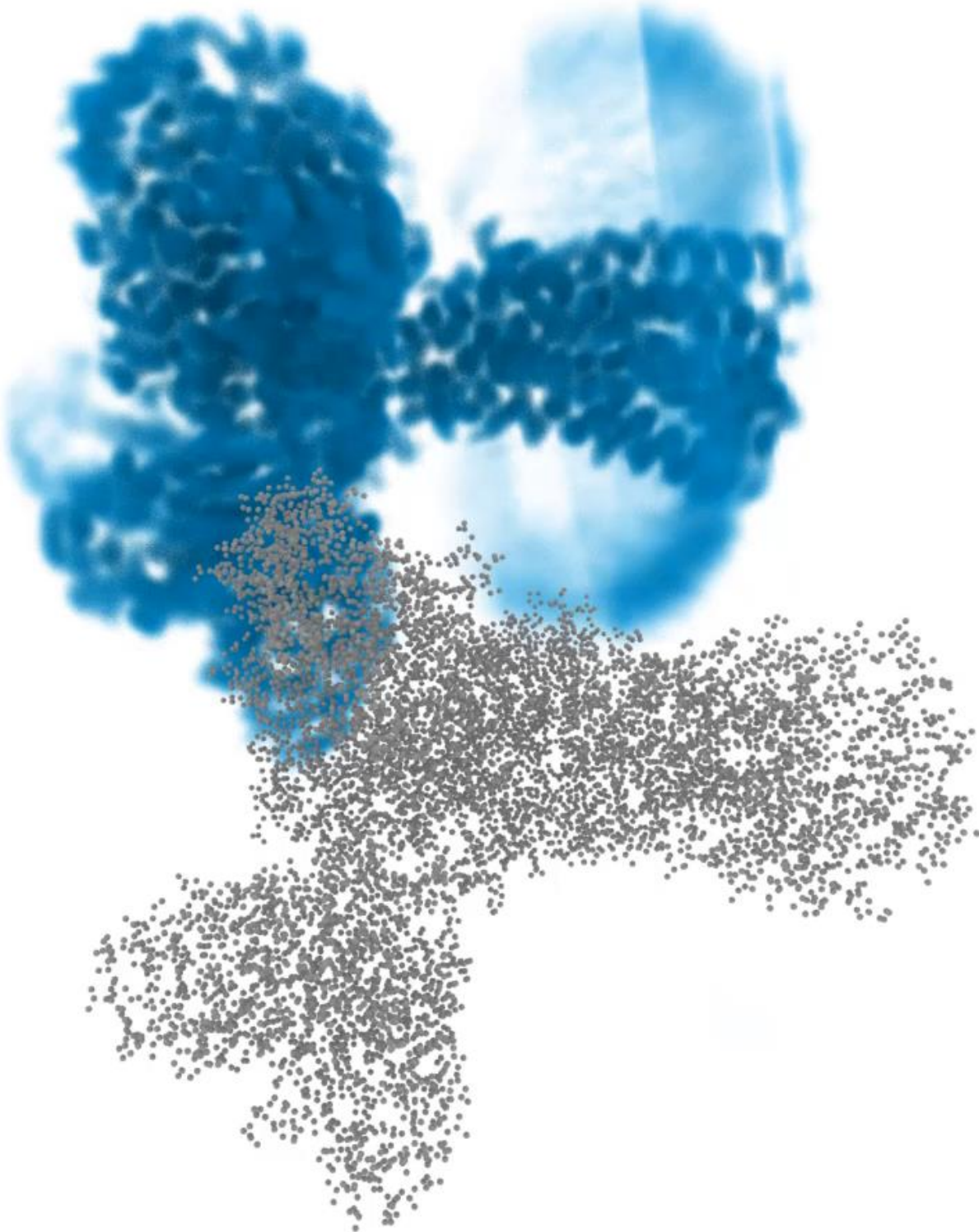




# Method



# Method



# Method - Iterative Convolution

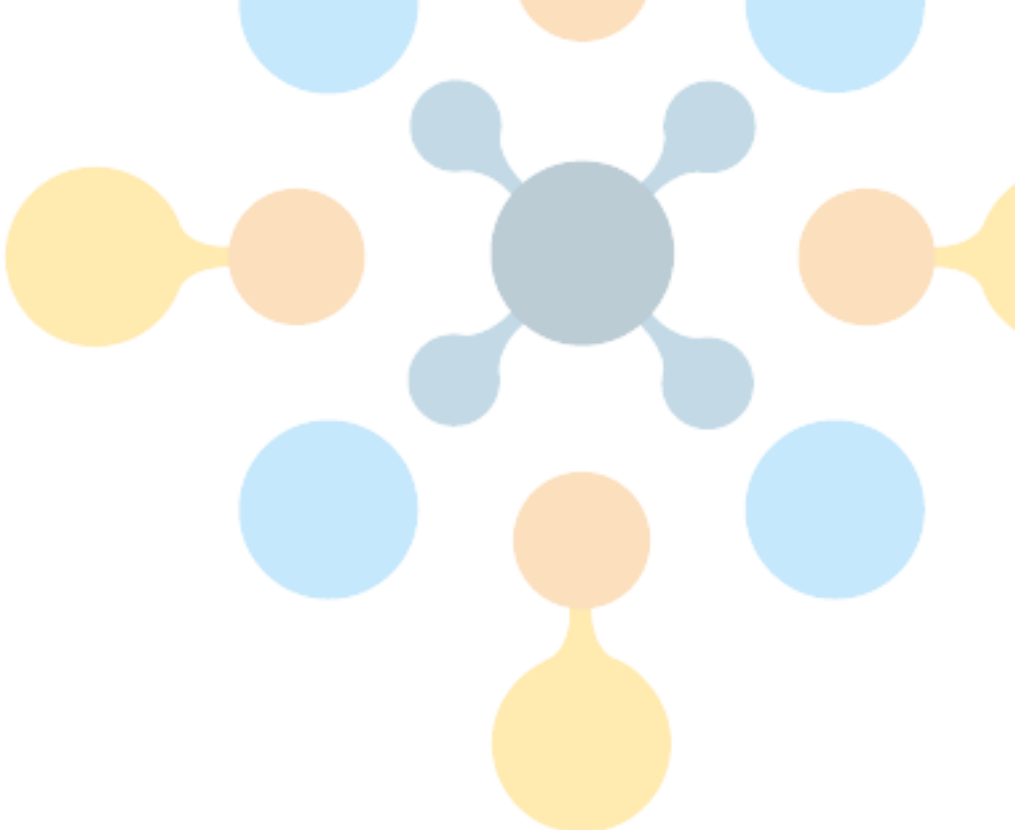
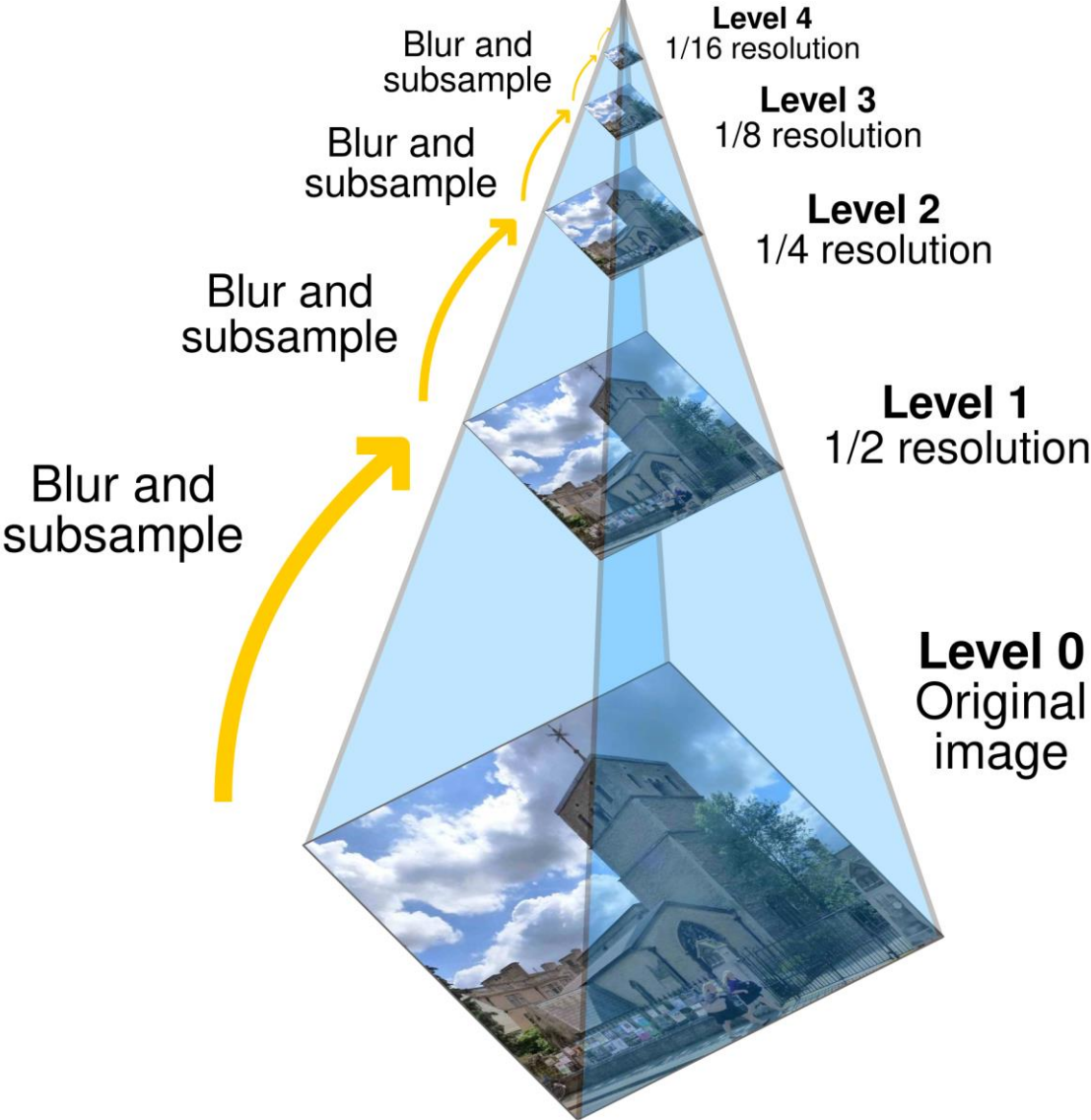
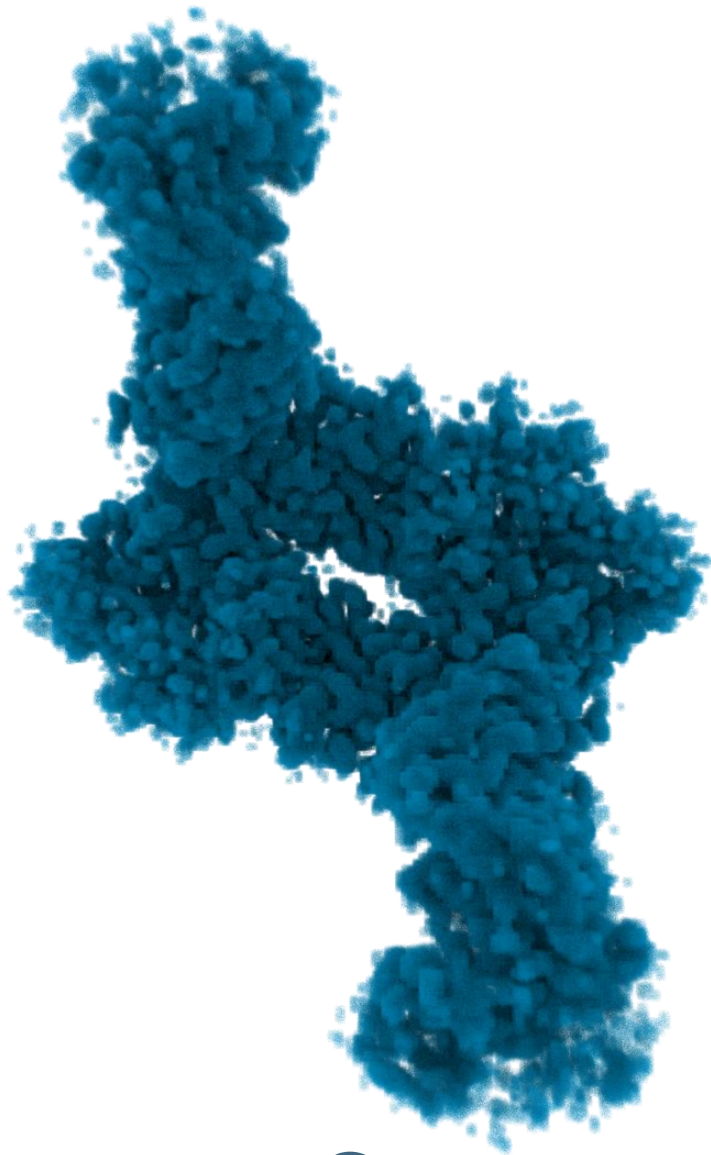


Image Source: Wikipedia - Pyramid (image processing)

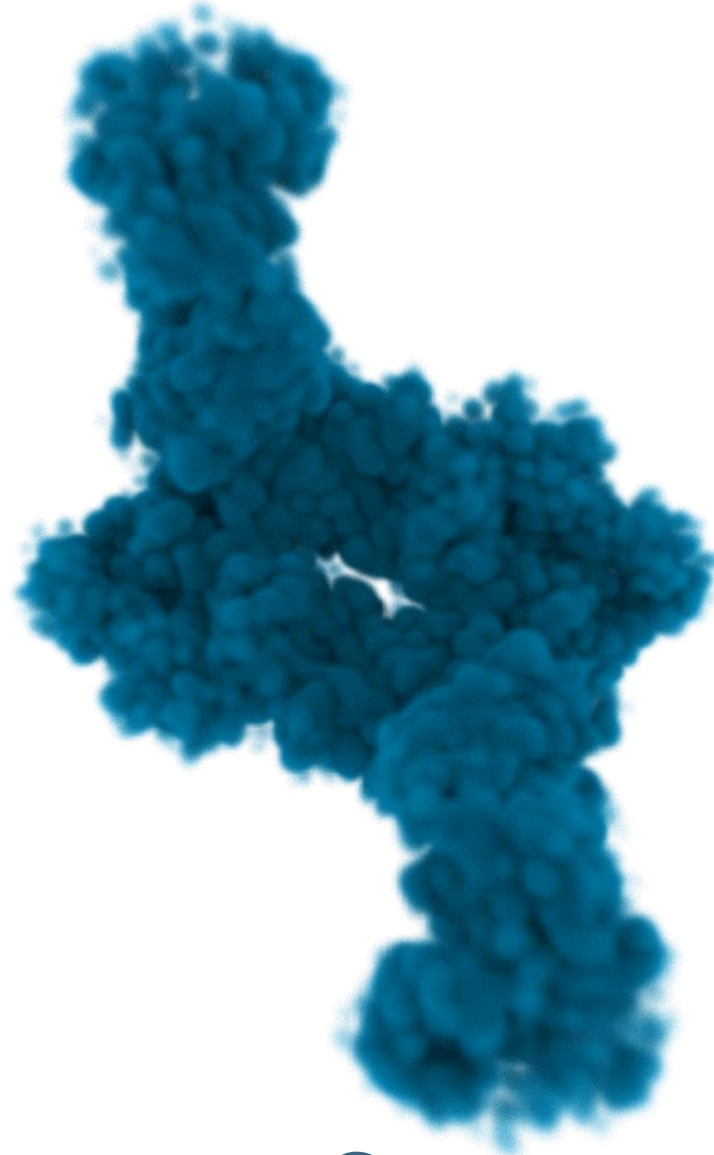


# Method - Iterative Convolution

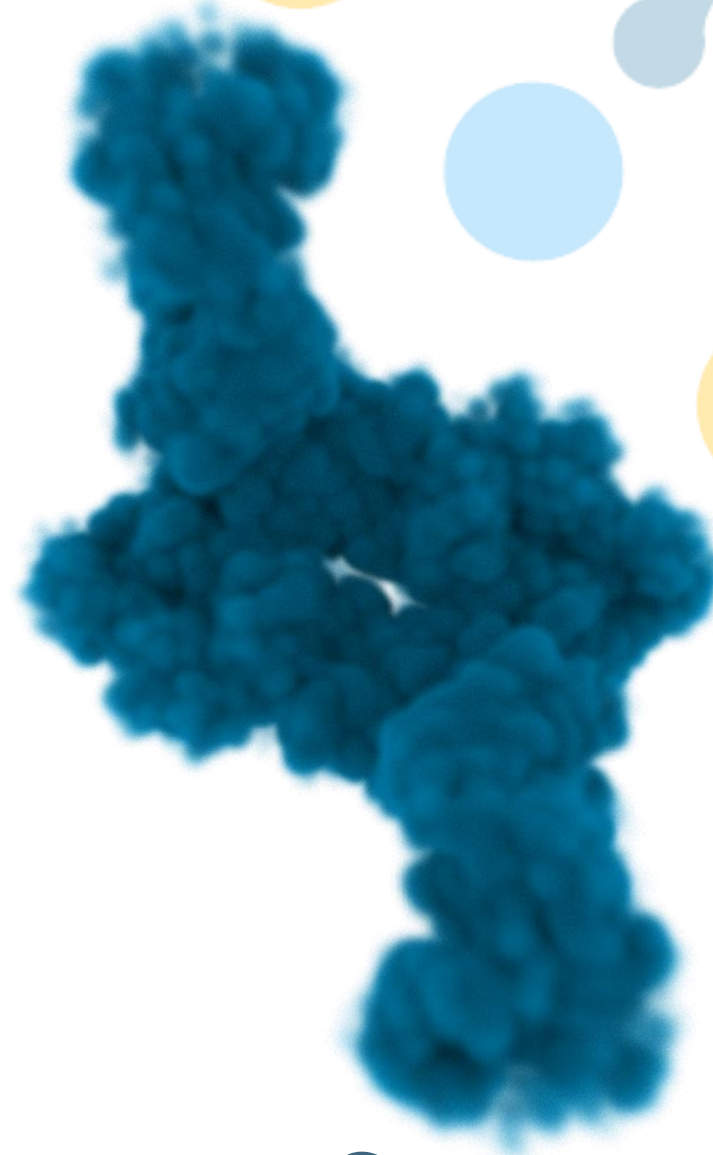
$$\hat{V}_F^{G_n} = \hat{V}_F^{G_{n-1}} * G_n$$



$G_0$



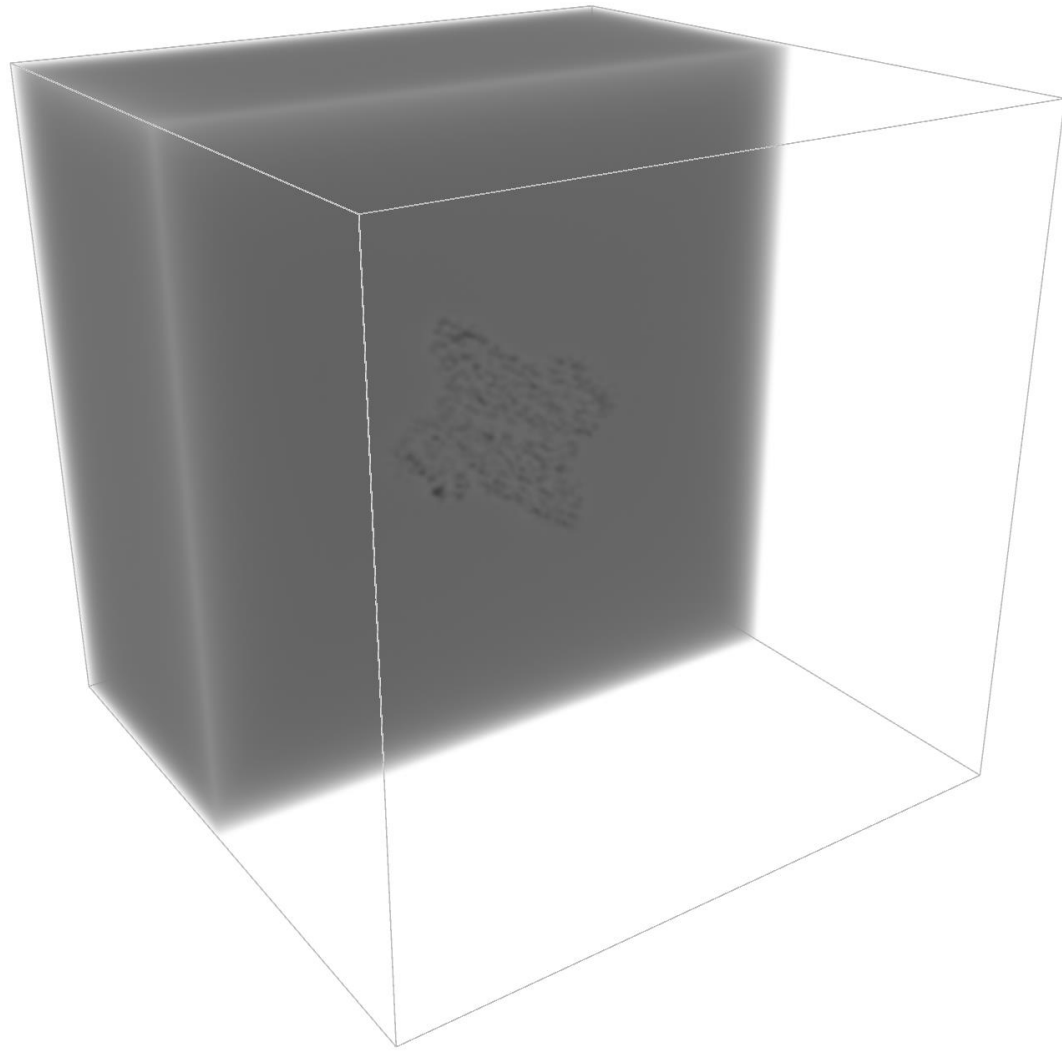
$G_1$



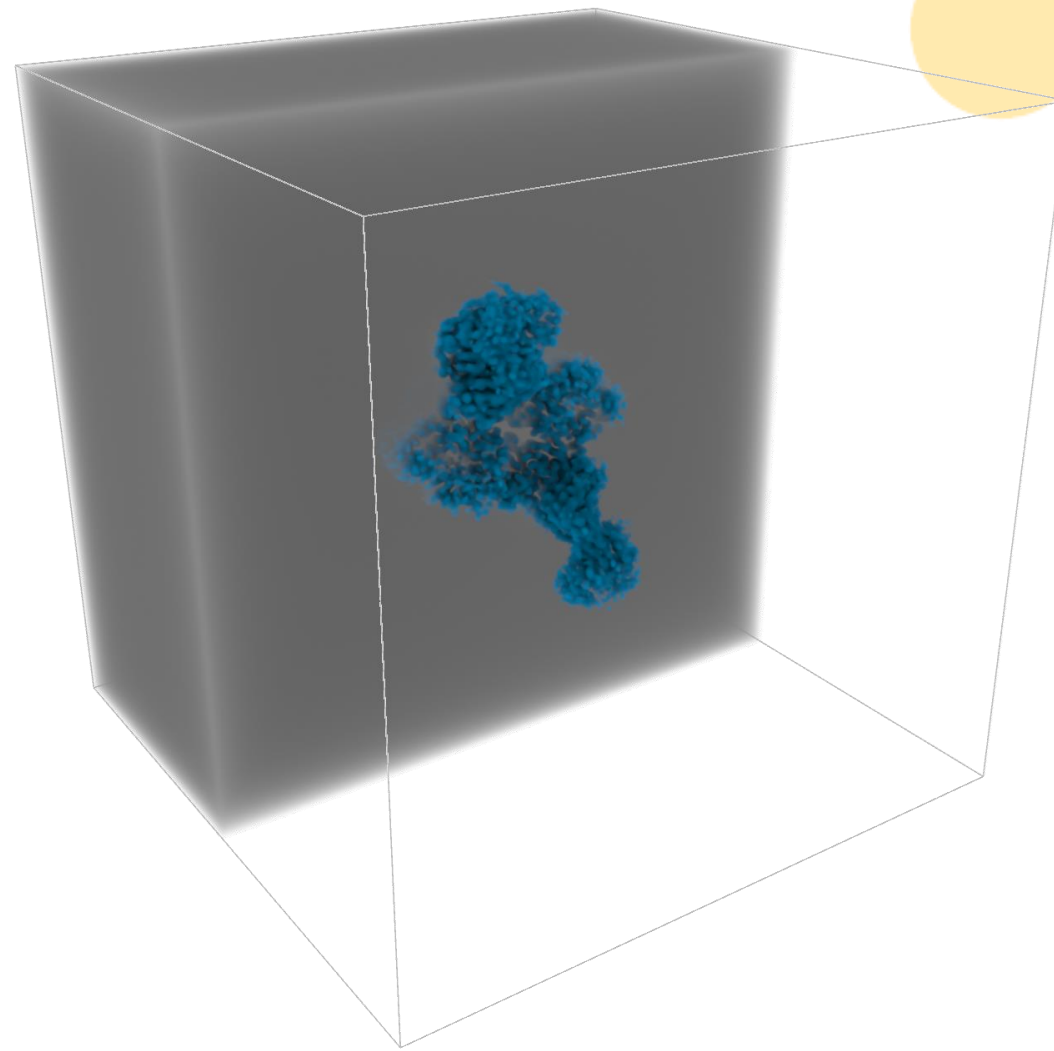
$G_2$

# Method - Negative Space

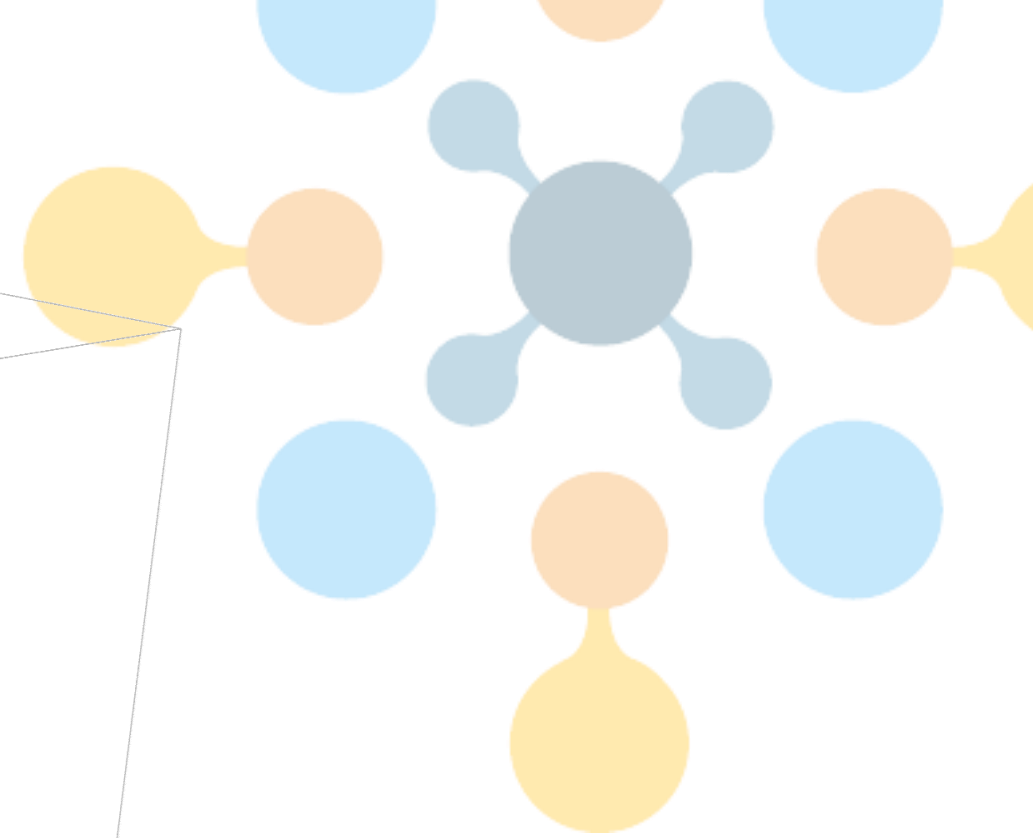
$$\hat{V}_{F-c}^{G_j}$$



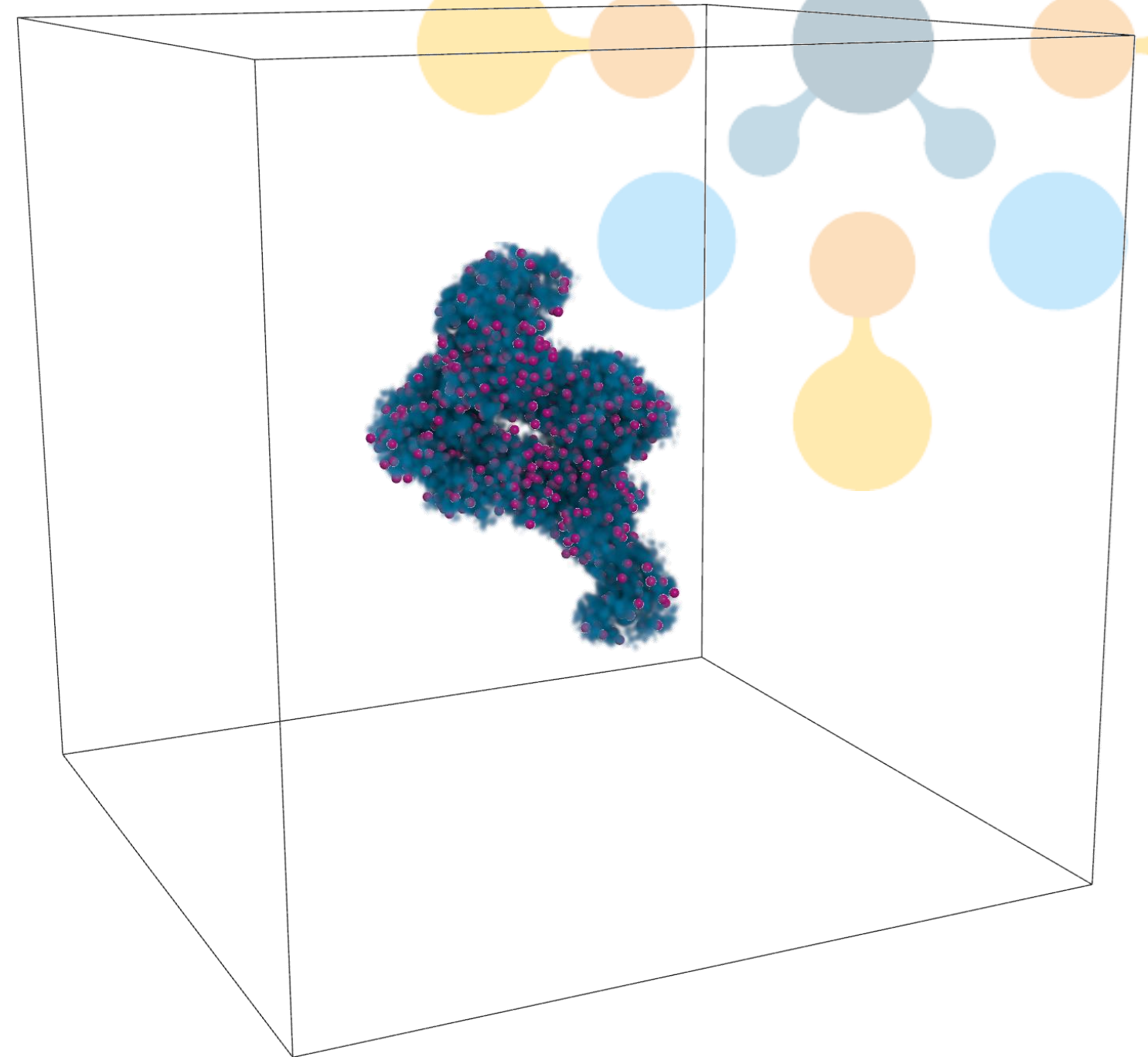
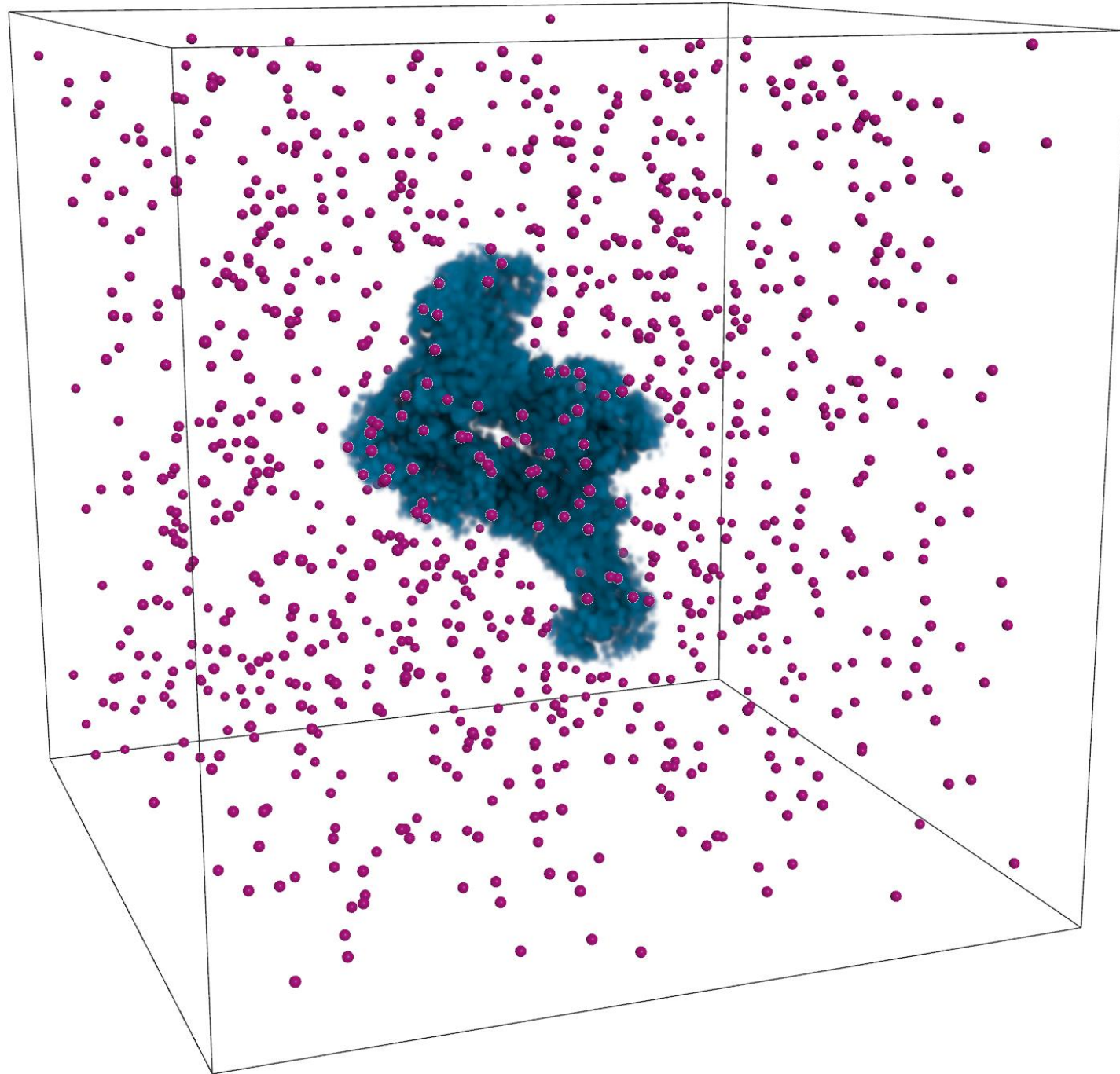
Gray: Negative space  
in cutaway view



Blue: Positive space



# Method - Enveloped Initialization





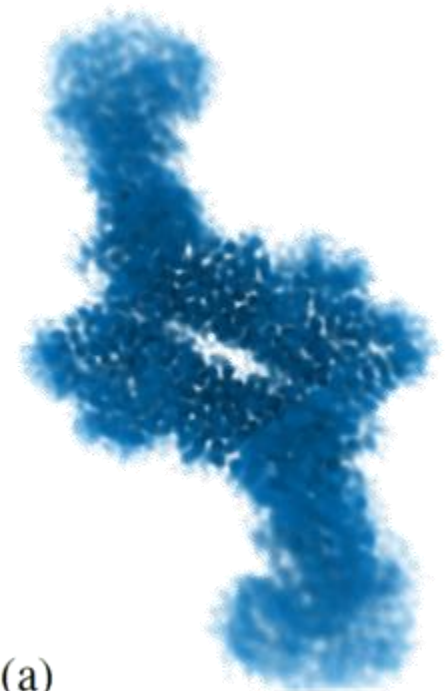
## Method - Optimization Loss Function

$$\underline{L}(\underline{\mathbf{p}}, \underline{\mathbf{q}}, \underline{\mathbf{X}}_m, \underline{V}) = - \left( \frac{1}{N} \sum_i^N \underline{D}(\underline{T}(\underline{\mathbf{x}}_i)) \right) = - \frac{1}{N} \sum_i^N \underline{S}(\underline{M}_{\mathbf{q}} \cdot \underline{\mathbf{x}}_i + \underline{\mathbf{p}}, \underline{V})$$

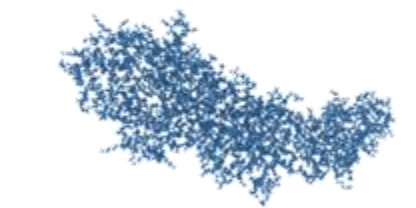
$$\underline{L}_m([\underline{\mathbf{p}}, \underline{\mathbf{q}}]) = \sum_{j=1}^n w_j \cdot \underline{L}(\underline{\mathbf{p}}, \underline{\mathbf{q}}, \underline{\mathbf{X}}_m, \underline{\hat{V}}_{F-c}^{G_j})$$

$$\underline{L}_{par}(m) = \sum_{k=1}^{\underline{N}_{\mathbf{q}} \cdot \underline{N}_{\mathbf{p}}} \underline{L}_m([\underline{\mathbf{p}}_k, \underline{\mathbf{q}}_k]) \quad \underline{L}_{all} = \sum_{l=1}^M \underline{L}_{par}(l)$$

# Method - Fit Results Clustering and Filtering



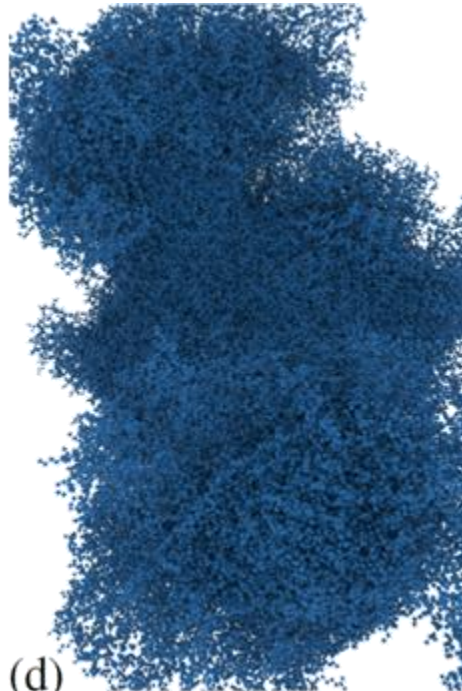
(a) Target map



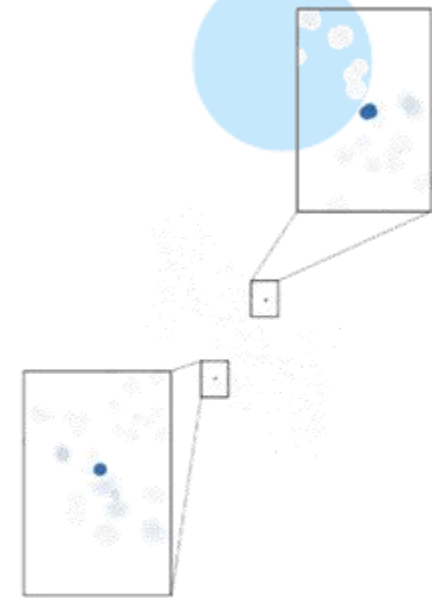
(b) Source structure



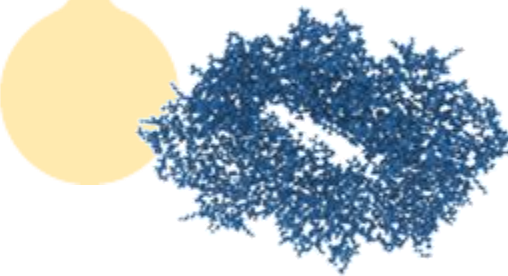
(c) Centers of 1000 fits' final placements



(d) All instances of the structure



(e) Clustering and filtering



(f) Instances in 2 clusters standing out

# Method - Tabular Browser for Fit Result Clusters

The screenshot displays the ChimeraX interface with a central 3D visualization of a protein structure. On the left, the DiffFit panel shows a table of fit result clusters. On the right, a Log window provides details for selected clusters, and a Models window shows the loaded files. A Volume Viewer at the bottom right shows a density map.

id	Mol Id	Hits	Density	C
1	0	25	699.437	0.335
2	0	26	699.425	0.335
3	1	8	230.7	0.294
4	1	5	230.686	0.294
5	0	2	218.936	0.264
6	0	5	217.253	0.264
7	0	6	213.968	0.267
8	0	3	209.273	0.261
9	0	8	206.489	0.258
10	0	14	206	0.257
11	0	14	205.738	0.263
12	0	1	203.75	0.271
13	0	6	203.692	0.261
14	0	5	203.049	0.269
15	0	4	202.195	0.259
16	0	8	200.243	0.261
17	0	1	200.051	0.259
18	0	2	199.833	0.259
19	0	1	199.745	0.257

Log window details:

- Showing cluster ID: 2  
Cluster size: 26  
Highest metric reached at iter : 19
- Showing cluster ID: 3  
Cluster size: 8  
Highest metric reached at iter : 10
- Showing cluster ID: 4  
Cluster size: 5  
Highest metric reached at iter : 12
- Showing cluster ID: 1  
Cluster size: 25  
Highest metric reached at iter : 12

Models window:

Name	ID	Visible	Selected
emd_40589.map	1	<input checked="" type="checkbox"/>	<input type="checkbox"/>
8smk_transformed_chain_A.cif	2	<input checked="" type="checkbox"/>	<input type="checkbox"/>

Volume Viewer:

emd\_40589.map  
#1 400<sup>3</sup> step 2 Level 2 -12.6 -20.8 surface



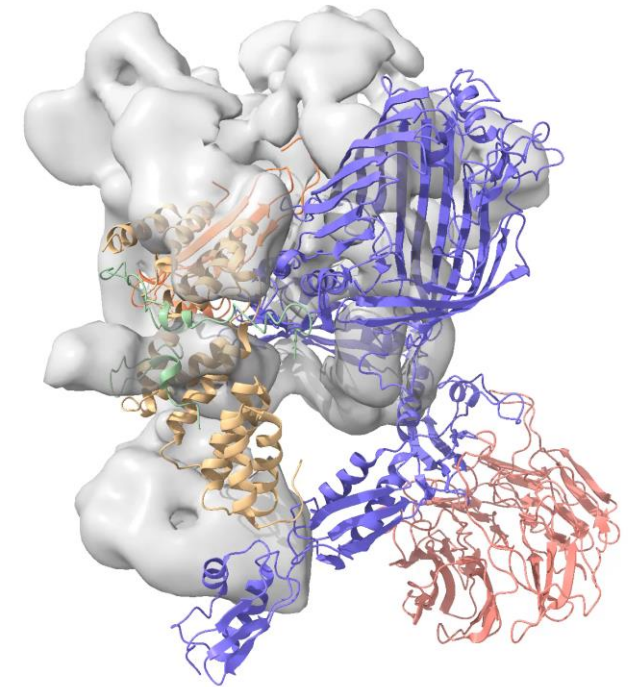
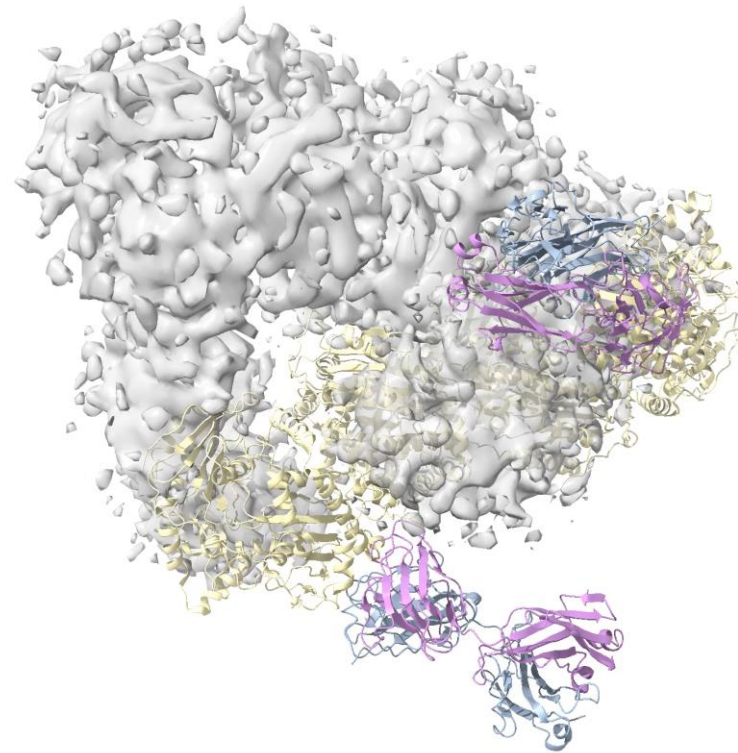
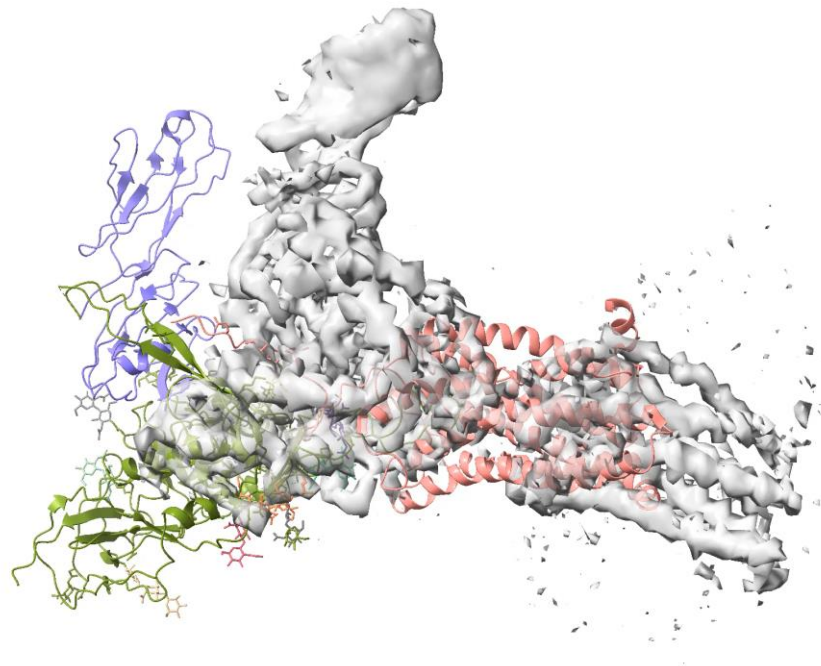
# Method - Spatial Browser for Fit Result Clusters

The screenshot displays the ChimeraX interface with a 3D protein structure and a DiffFit panel. The DiffFit panel includes a table with the following data:

id	Mol Id	Hits	Density
1	0	25	699.437
2	0	26	699.425
3	1	8	230.7
4	1	5	230.686
5	0	2	218.936
6	0	5	217.253
7	0	6	213.968
8	0	3	209.273
9	0	8	206.489
10	0	14	206
11	0	14	205.738
12	0	1	203.75
13	0	6	203.692
14	0	5	203.049
15	0	4	202.195
16	0	8	200.243
17	0	1	200.051
18	0	2	199.833
19	0	1	199.745

Additional interface elements include a Log window showing cluster information for 'Activating Fab 362 heavy chain' and 'Protein-arginine deiminase type-4', and a Models window listing 'emd\_40589.map', '8smk\_transformed\_chain\_A.cif', and 'clusterSpheres'.

# Result - Use Case 1: Fit a single structure in sec





# Result - Use Case 1: Fit a single structure in sec

C: ChimeraX  
Fit in Map

D: DiffFit

G: Gain

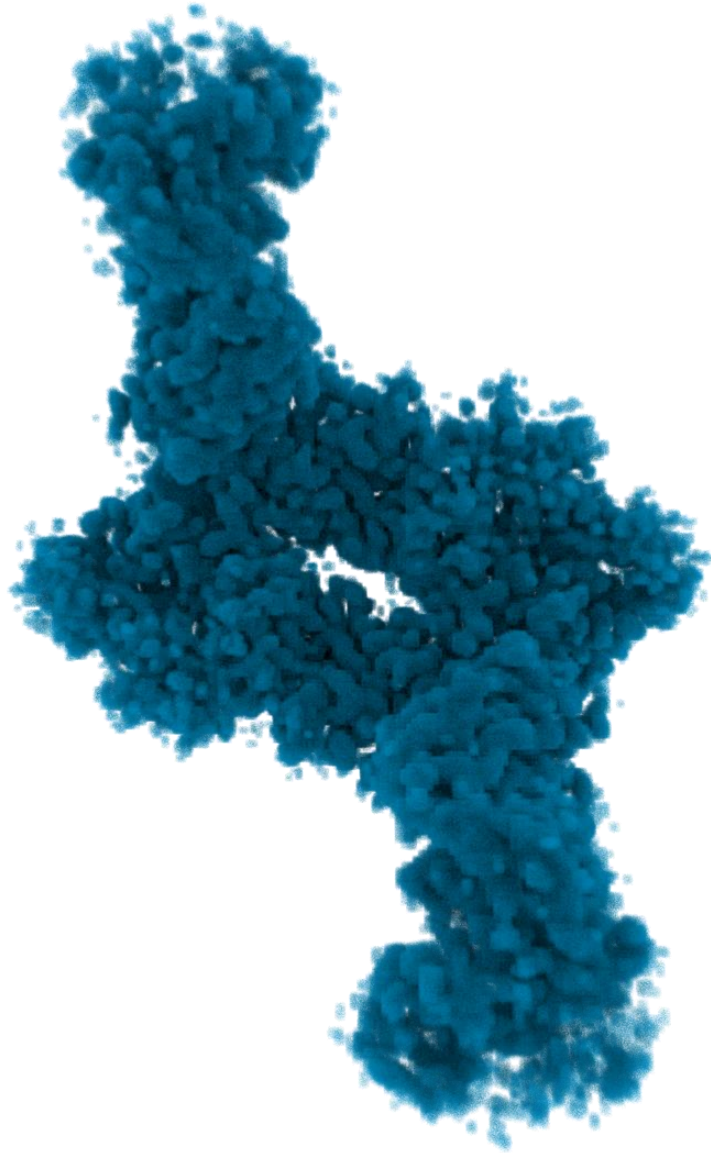
M: MarkovFit

DC: DiffFit  
refined by  
ChimeraX

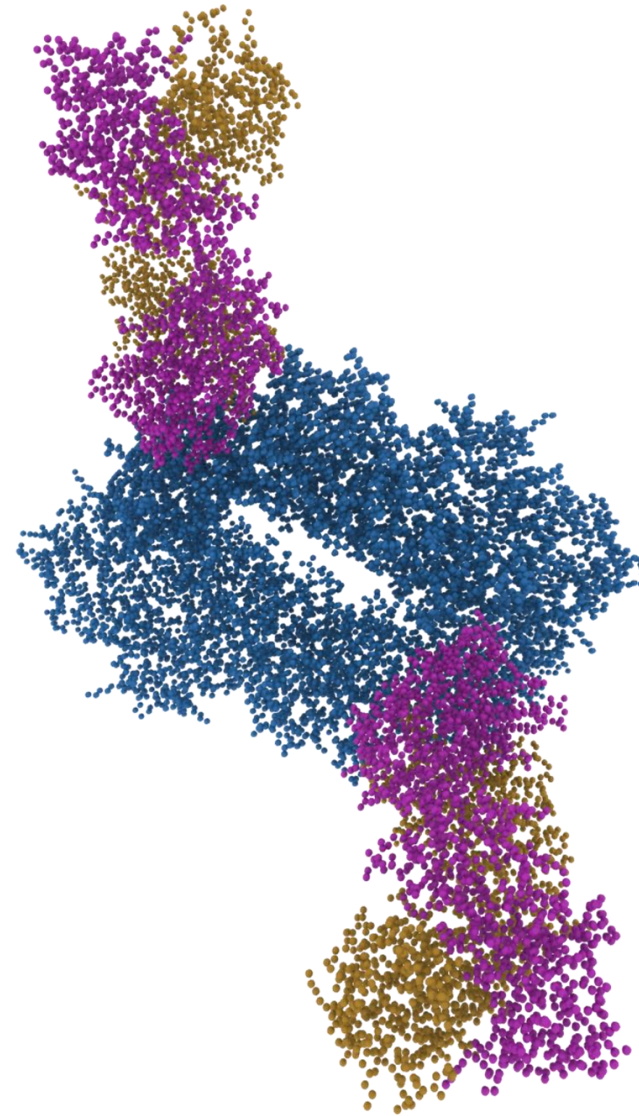
PDB	Res	Hit rate			Computing time			RMSD (Å)			
		C	D	G	C	D	G	M	C	D	DC
6WTI	2.38	0.0	<b>136.8</b>	n/a	150.3	<b>3.8</b>	39.7	1.310	n/a	0.942	<b>0.037</b>
7D8X	2.60	0.0	<b>202.0</b>	n/a	196.0	<b>5.2</b>	37.6	1.960	n/a	0.984	<b>0.014</b>
7SP8	2.70	4.6	<b>188</b>	40.9	130.6	<b>2.6</b>	50.5	1.290	0.996	0.969	<b>0.025</b>
7STE	2.73	14.0	<b>110.4</b>	7.9	806.1	<b>12.1</b>	66.6	1.740	0.062	0.662	<b>0.058</b>
7JPO	3.20	5.4	<b>191.8</b>	35.5	250.7	<b>6.7</b>	37.2	2.540	0.017	0.922	<b>0.015</b>
7PM0	3.60	44.0	<b>195.4</b>	4.4	352.4	<b>4.1</b>	86.7	1.640	0.030	0.907	<b>0.024</b>
6M5U	3.80	0.0	<b>105.0</b>	n/a	162.2	<b>4.1</b>	39.2	2.360	n/a	0.912	<b>0.018</b>
6MEO	3.90	7.4	<b>116.0</b>	15.7	128.2	<b>3.2</b>	40.1	1.940	0.489	0.786	<b>0.488</b>
7MGE	3.94	4.8	<b>123.6</b>	25.8	337.6	<b>4.3</b>	78.1	1.870	<b>0.017</b>	0.819	<b>0.017</b>
High-avg		8.9	<b>152.1</b>	21.7	279.3	<b>5.1</b>	52.8	1.850	0.268	0.878	<b>0.077</b>
5NL2	6.60	1.8	<b>163.2</b>	90.7	94.6	<b>2.0</b>	48.0	2.440	0.093	1.124	<b>0.056</b>
7K2V	6.60	49.0	<b>165.6</b>	3.4	240.6	<b>4.1</b>	58.2	25.290	<b>0.338</b>	1.323	<b>0.338</b>
7CA5	7.60	55.8	<b>72.4</b>	1.3	322.6	<b>2.9</b>	110.0	3.290	2.042	<b>1.207</b>	2.042
5VH9	7.70	68.6	<b>158.0</b>	2.3	1147.8	<b>14.1</b>	81.3	0.960	<b>0.085</b>	0.991	<b>0.085</b>
6AR6	9.00	78.0	<b>182.6</b>	2.3	74.9	<b>1.5</b>	49.3	2.200	0.123	2.617	<b>0.117</b>
3J1Z	13.00	138.6	<b>172.2</b>	1.2	64.4	<b>2.0</b>	33.0	32.330	0.396	2.612	<b>0.388</b>
Med-avg		65.3	<b>152.3</b>	16.9	324.1	<b>4.4</b>	63.3	11.085	0.513	1.646	<b>0.504</b>
All-avg		31.5	<b>152.2</b>	<b>19.8</b>	297.3	<b>4.9</b>	<b>57.0</b>	5.544	0.366	1.185	<b>0.248</b>



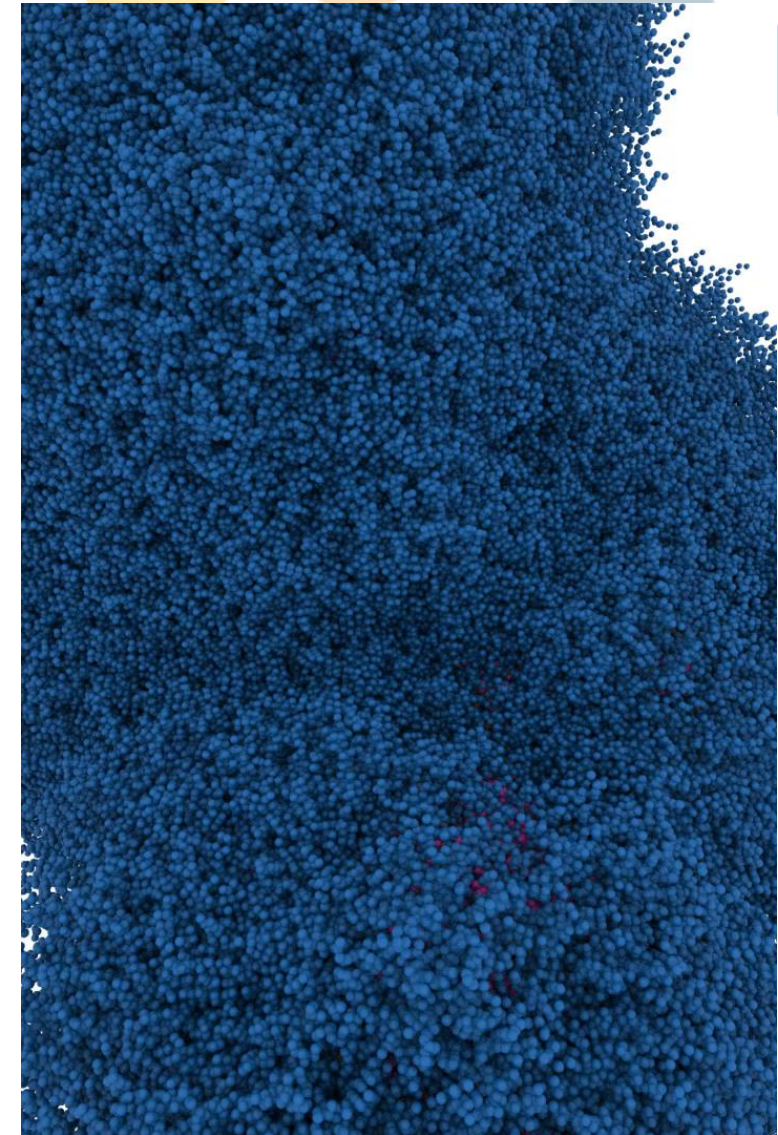
# Result - Use Case 2: Composite multiple structures in min



Target map

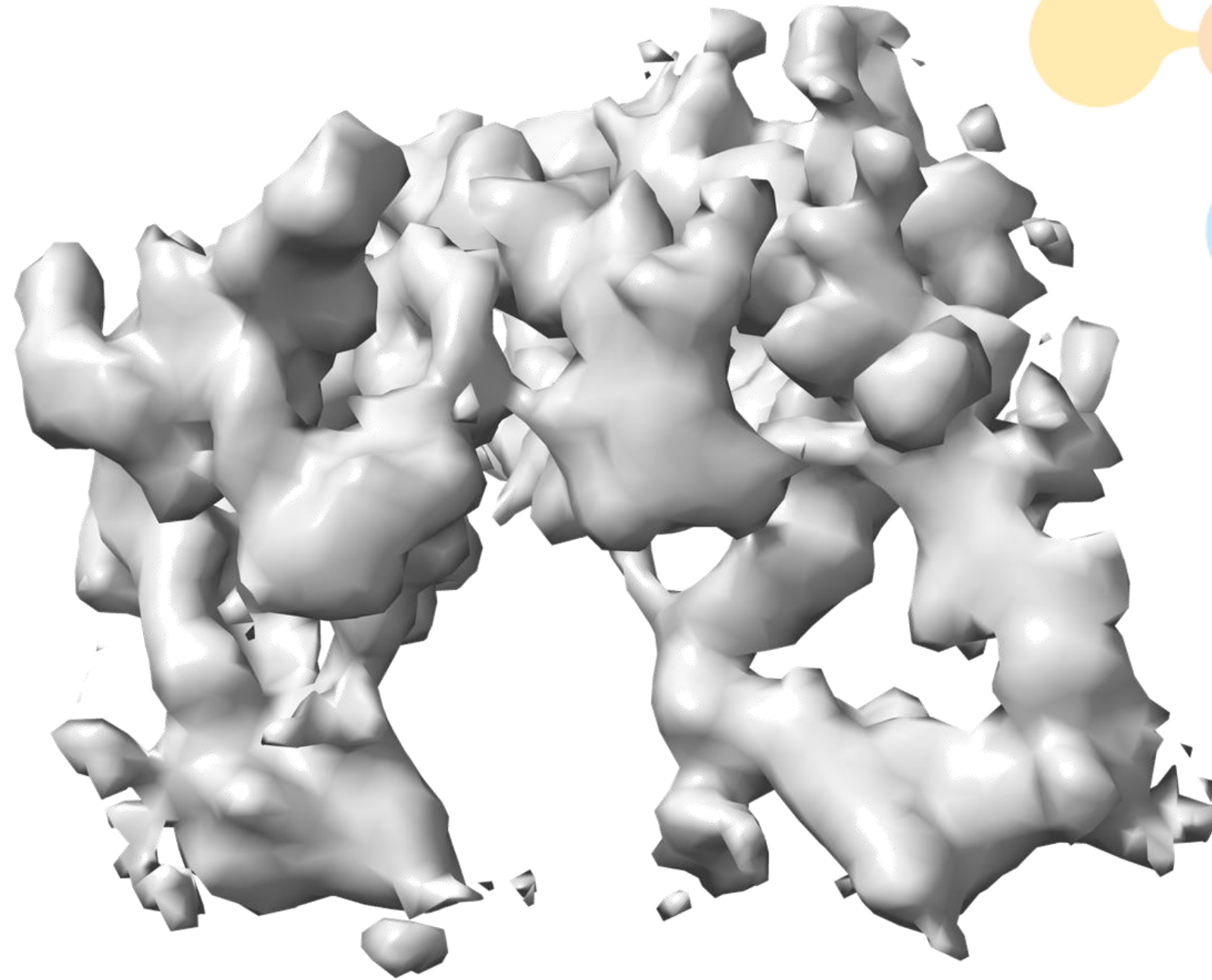


Ground truth



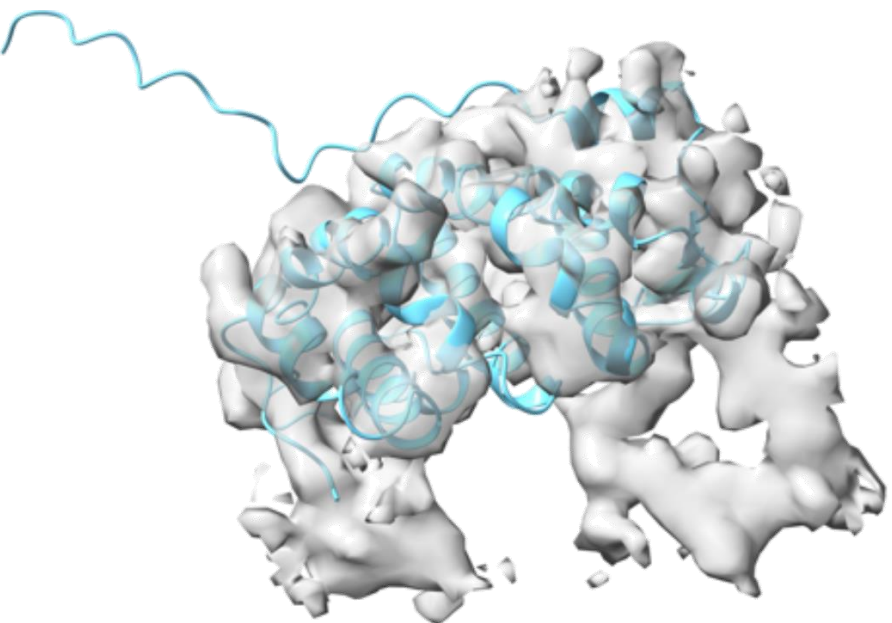
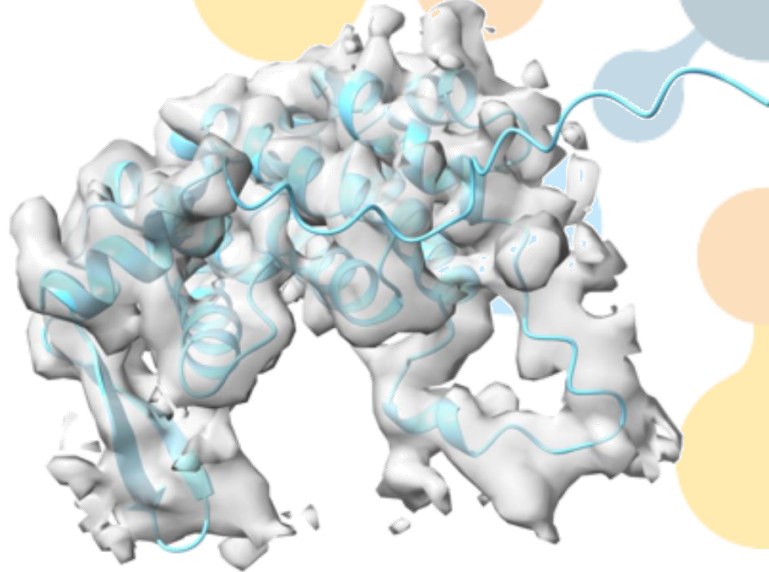
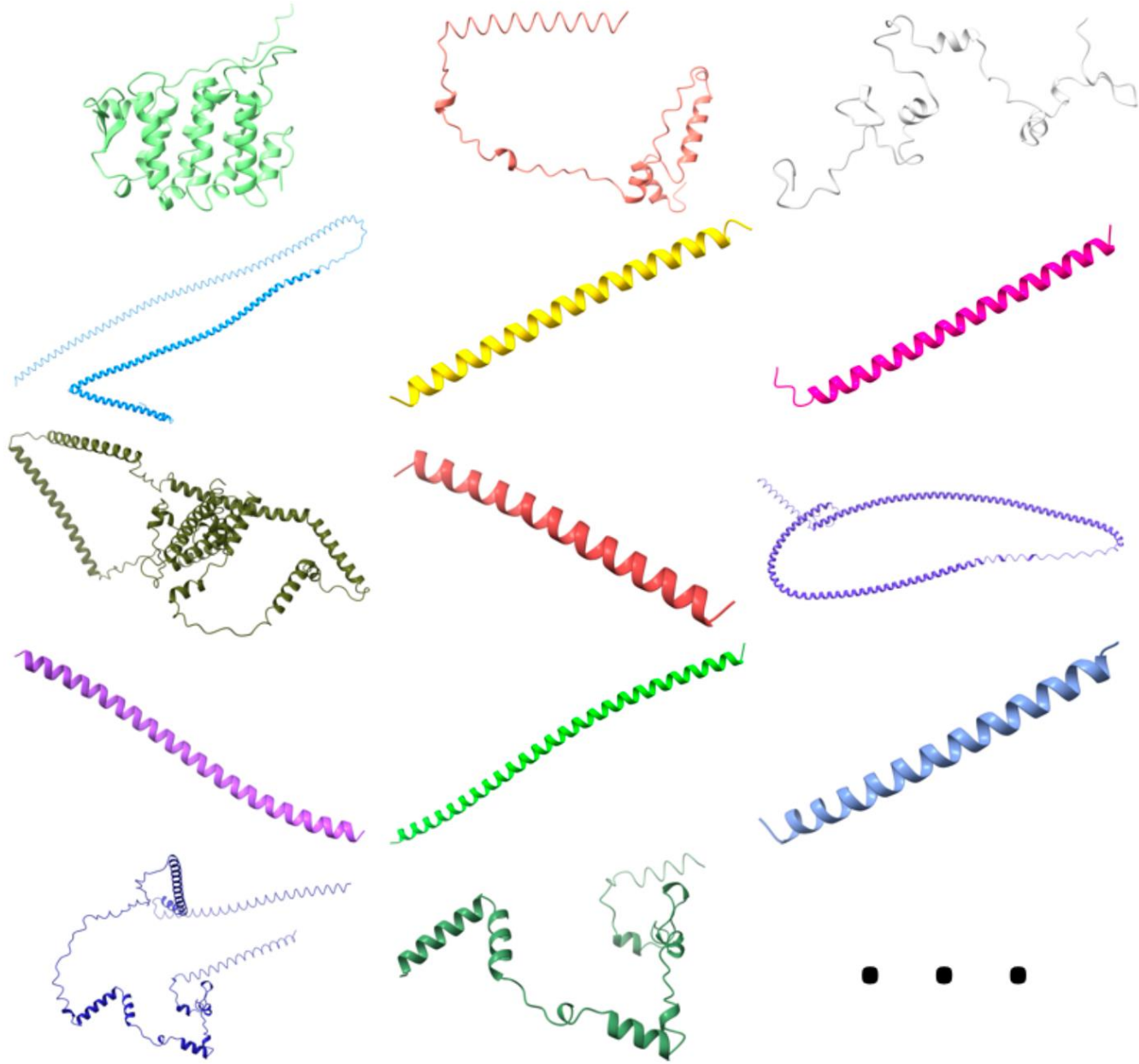
Optimization process

# Result - Use Case 3: Identify unknown densities in min-hour





# Result - Use Case 3: Identify unknown densities in hours



• • •



# Result - Use Case 3: Identify unknown densities in hours

Structure	C Hit	D Hit	Gain
I7MLV6_D3	108	254	2.4x
I7M317_D1	127	240	1.9x

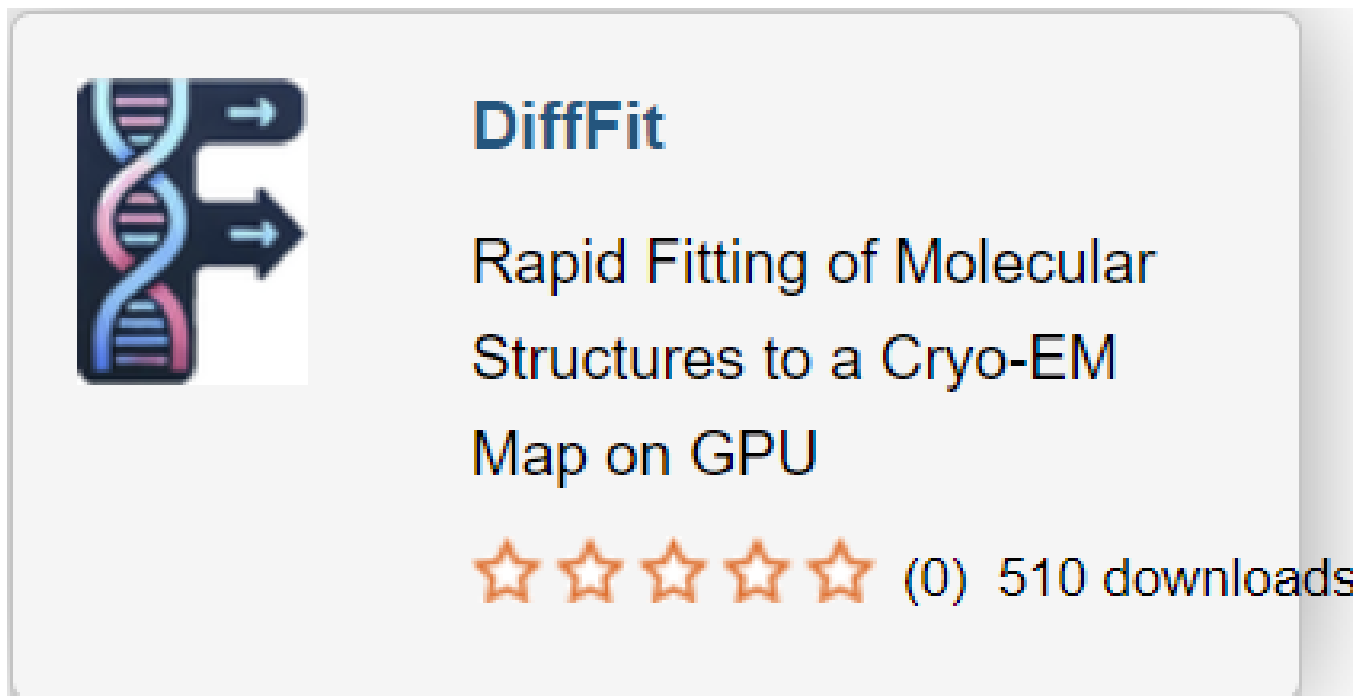
DomainFit: ~12 hours  
DiffFit: ~7 minutes (103x)


# Feedback

- “Quite intuitive and easy to use”
- The automatic fitting and visual inspection approach “could be a key feature in ChimeraX that could become a standard in many pipelines” as well as “a key implementation for a standard modeling workflow.”

# Tool & Code

- Python command line script
- <https://cxtoolshed.rbvi.ucsf.edu/>
  - (The server is down for maintenance now as announced on X @UCSFChimeraX.)
  - 510 downloads

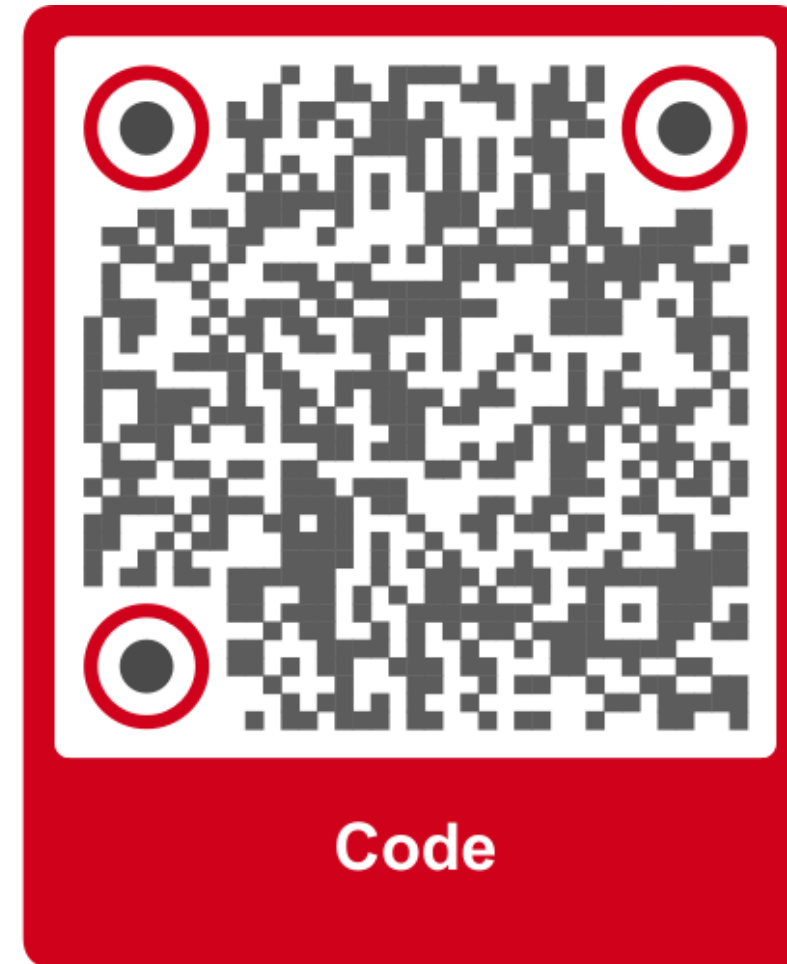




**DiffFit**

Rapid Fitting of Molecular Structures to a Cryo-EM Map on GPU

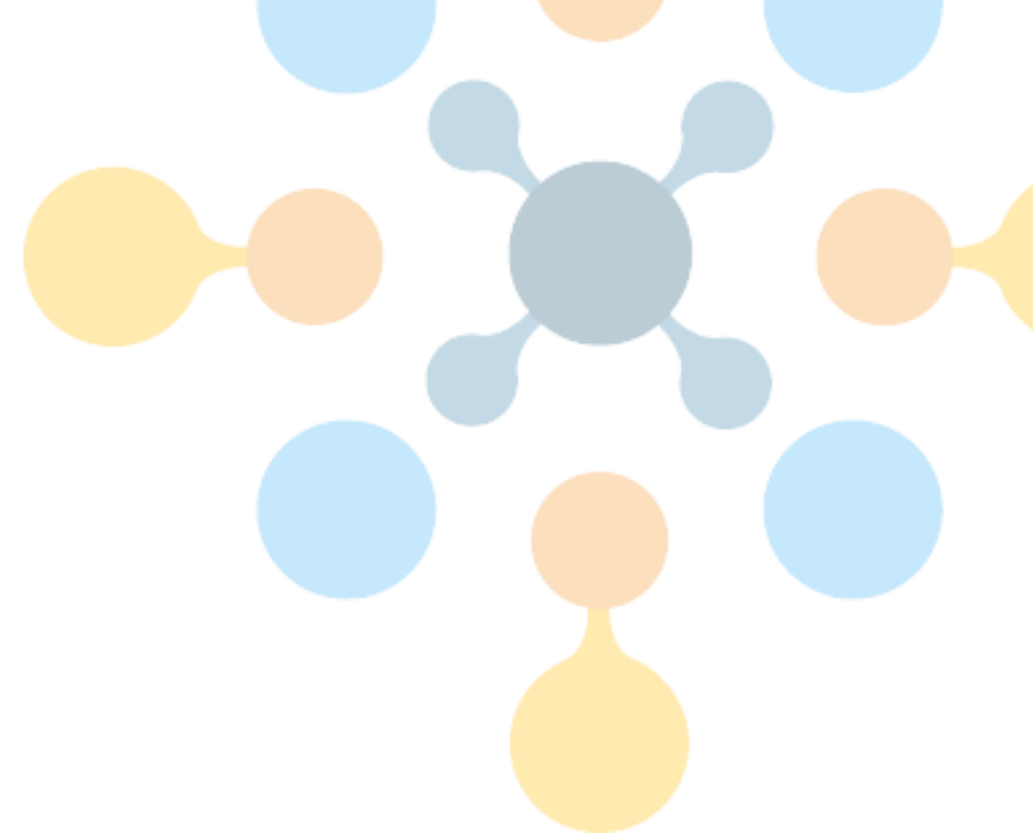
☆☆☆☆ (0) 510 downloads





# Next Steps

- Comprehensive visualization
- Deformable transformations
- Collision handling
- Splitting and splicing the Cryo-EM map and the structures



# Thank You!



Zainab Alsuwaykit



Dawar Khan



Ondřej Strnad



Tobias Isenberg



Ivan Viola



[nanovis.org](https://nanovis.org)