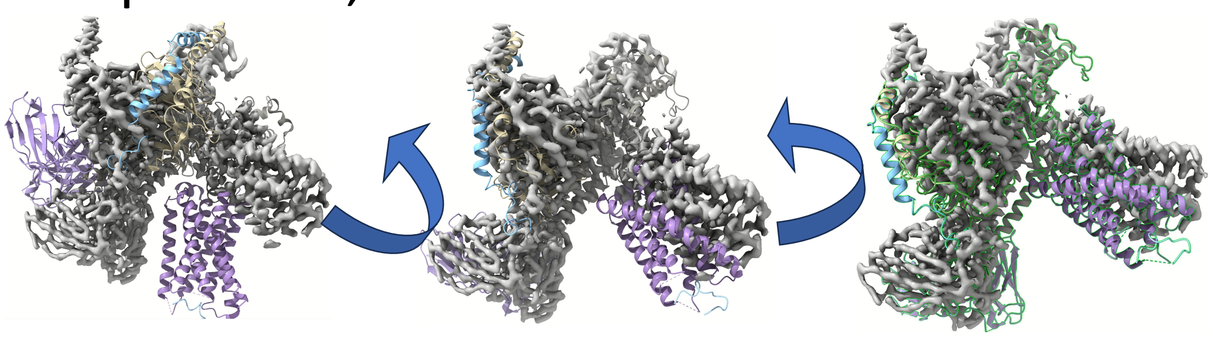
DiffFit: Differentiable Fitting of Molecule Structures to a Cryo-EM Map

Deng Luo, Zainab Alsuwaykit, Dawar Khan, Ondřej Strnad, Tobias Isenberg, Ivan Viola

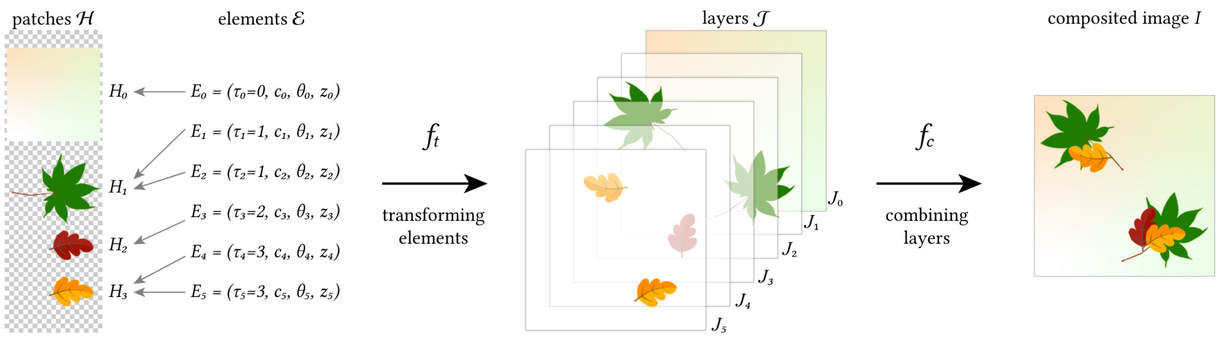
Introduction

A common practice in structural biology, often performed in the first step once a Cryo-EM map is ready

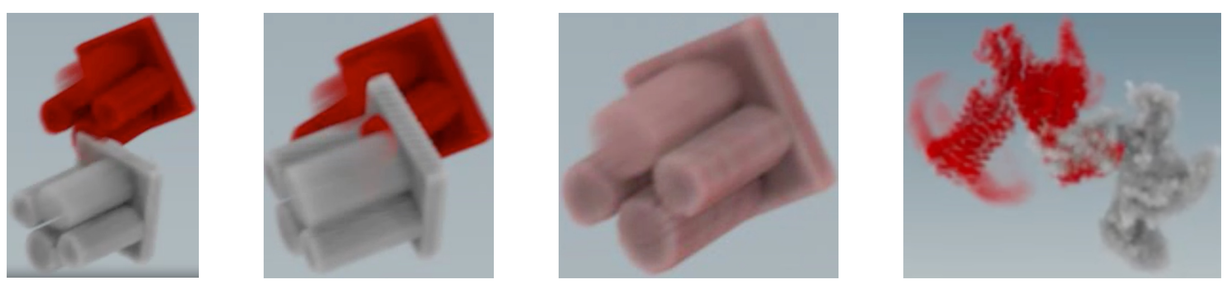
Existing solutions: Manual placing, slow computation, low hit rate



Inspiration – Reddy *et al.* DiffComp 2020

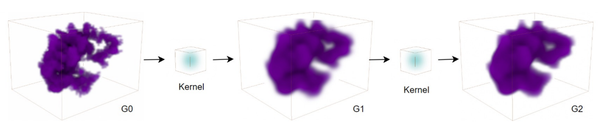


Method



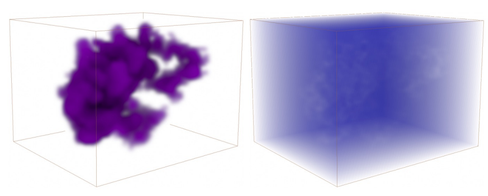
Iterative convolution

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

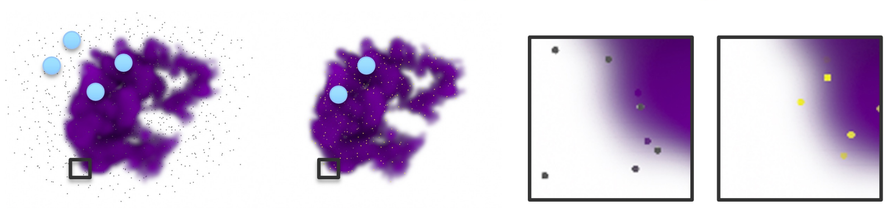


Negative space

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

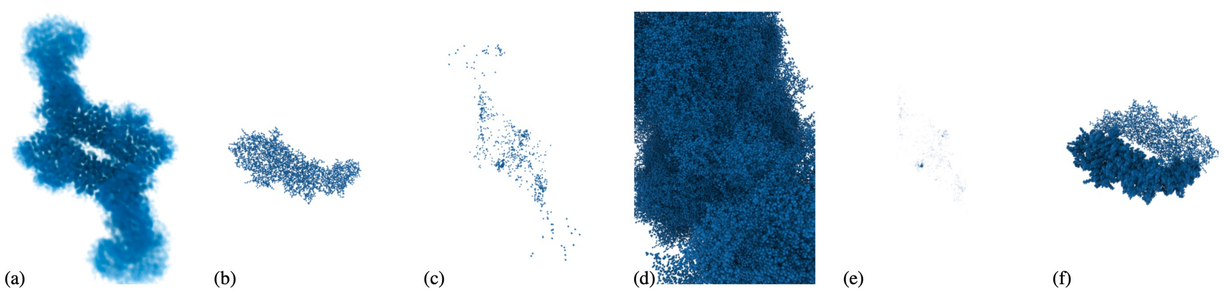


Enveloped initialization



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

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

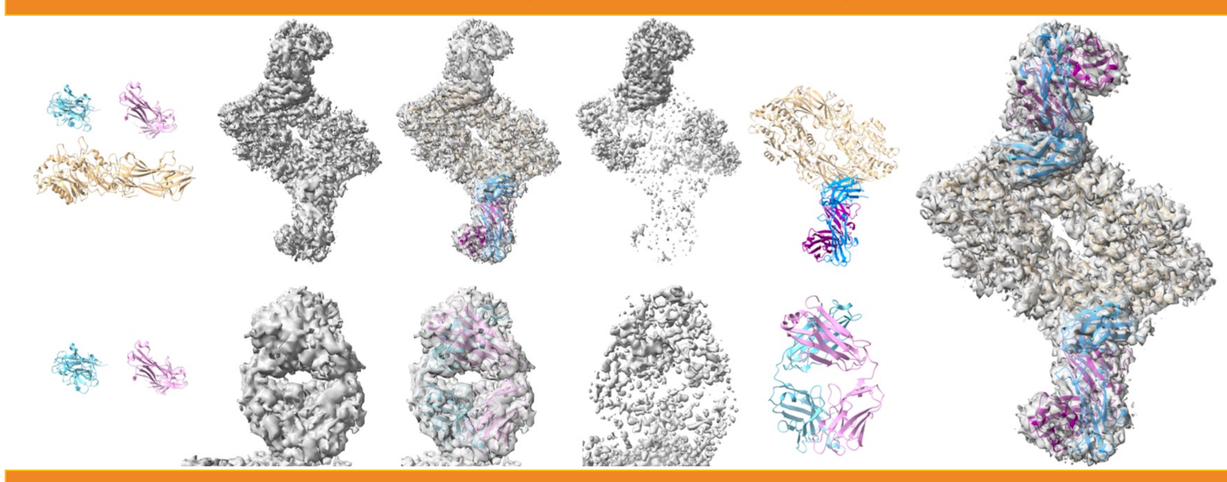


Use case 1: Fit a single structure

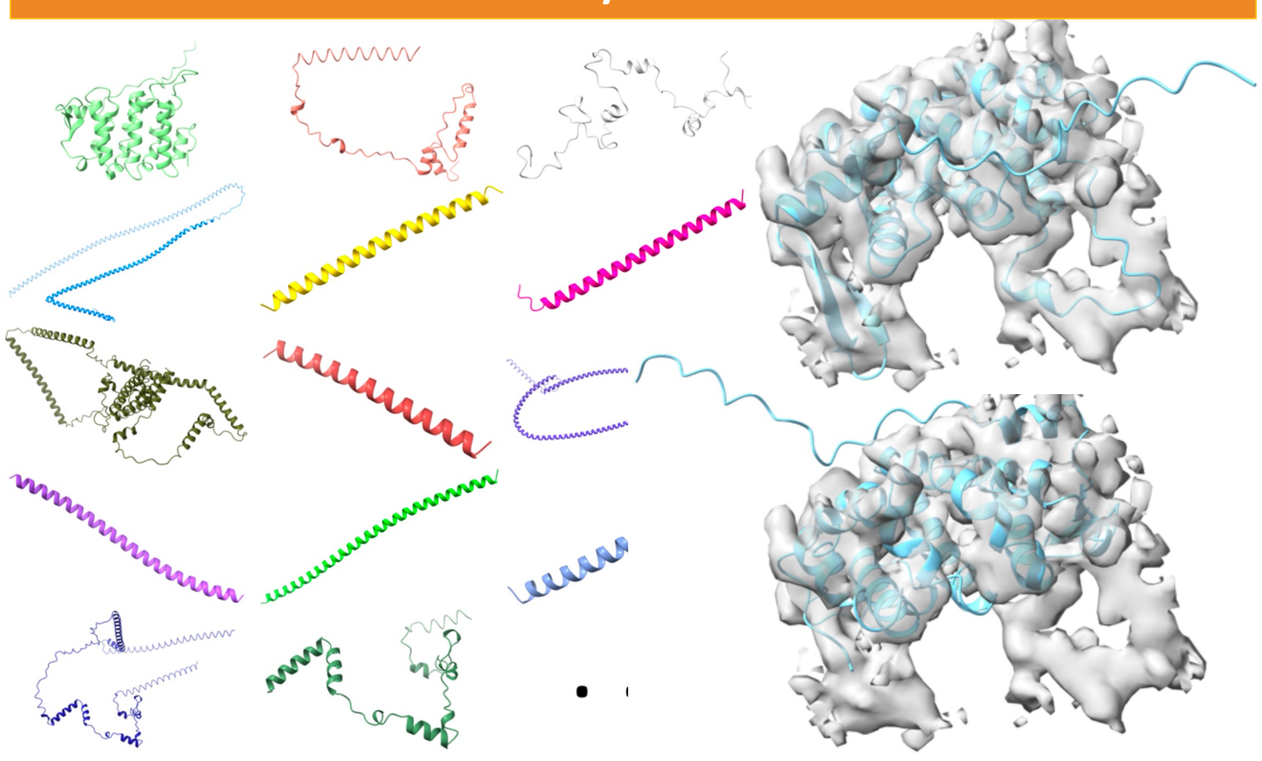
(a) Source structure. (b) Target map. (c) Fit result.

PDB	Res	Hit rate C	Hit rate D	Hit rate G	Time (sec) C	Time (sec) D	Time (sec) G	RMSD C	RMSD M	RMSD D	RMSD DC
6MEO	3.90	7.4	**214.6**	41.2	128.2	**9.4**	13.7	0.489	1.940	0.790	**0.483**
7PM0	3.60	44.0	**195.2**	4.5	352.4	**7.0**	50.3	0.029	1.640	0.976	**0.027**
7SP8	2.70	4.6	**238.8**	117.5	130.58	**12.2**	10.9	0.996	1.290	0.779	**0.024**
6M5U	3.80	0.0	**277.0**	INF	162.2	**18.8**	8.2	69.413	2.36	0.944	**0.014**
5NL2	6.60	0.6	**179.6**	200.0	92.8	**9.6**	10.4	23.110	2.440	1.903	**0.047**
7K2V	6.60	49	**170.4**	3.3	240.6	**8.56**	29.3	0.338	2.440	1.532	**0.337**
3J1Z	13.00	138.6	**441.6**	3.20	64.4	**3.0**	21.5	**0.396**	32.330	2.436	0.399

Use case 2: Composite multiple structures



Use case 3: Identify unknown densities



Future

- Comprehensive visualization
- Deformable transformations
- Collision handling
- Splitting and splicing structures

ICML International Conference On Machine Learning
2024 Workshop | Differentiable Almost Everything