Fast and effective protein model refinement using deep graph neural networks

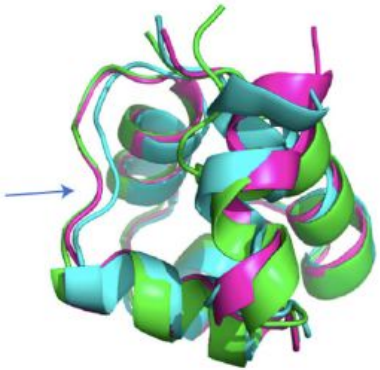
Xiaoyang Jing[1], Jinbo Xu[1,*]

[1]Toyota Technological Institute at Chicago, Chicago, IL 60637, USA

Presented by:
Sumit Tarafder

Refinement

- One of the last steps in prediction pipeline.
- Improve the predicted model quality
- Challenging prospect
    - Space of worse models is larger than better models



R0974s1 (GDT-HA: 65.58→79.35)

SOTA methods

- Large-Scale Conformational Sampling
  - Molecular Dynamics (MD) Simulations
  - Fragment assembly
- Deep learning based methods

Conformational sampling methods

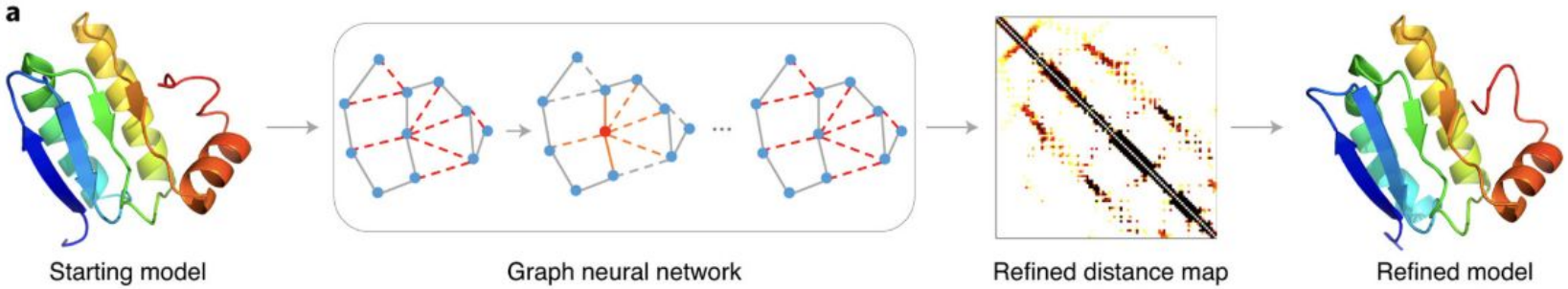
- FEIGLAB/ FEIG (Heo et al., 2019)
  - Iterative MD simulations (flat bottom harmonic)
- BAKER (Part et al. 2019, Hiranuma et al., 2021)
  - Local error estimation, fragment assembly
  - Recombine 2D motifs, replace torsion angles
  - Deep learning, estogram
- GalaxyRefine (Lee et al., 2019)
  - MD simulation, side chain re-packing
- YASARA server
  - Template modeling, MD simulation, YASARA force field
- MUFOLD server
  - Distance distribution for all pair, potential function, maximize using L-BFGS

Deep learning methods

- DeepAccNet (Hiranuma et al., 2021)
  - 2D and 3D CNN, estimates error
- refineD (Bhattacharya et al., 2019)
  - DeepCNF
- DeepRefiner (Shuvo et al, 2021)
  - Bhattacharya Server (CASP13, CASP14)
- ATOMRefine (Wu et al., 2023)
  - All atom graph (SE (3) - graph transformer)

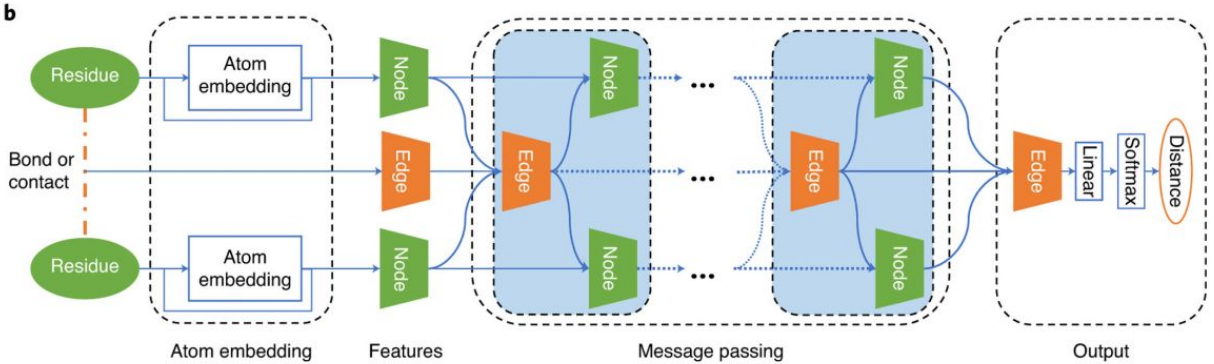
GNNRefine

- Deep learning instead of physics based sampling
- Graph neural network based approach
- Predict inter-residue distance distribution
  - Limited conformational sampling. Faster
  - Convert the distribution into potentials
    - Relax using PyRosetta FastRelax



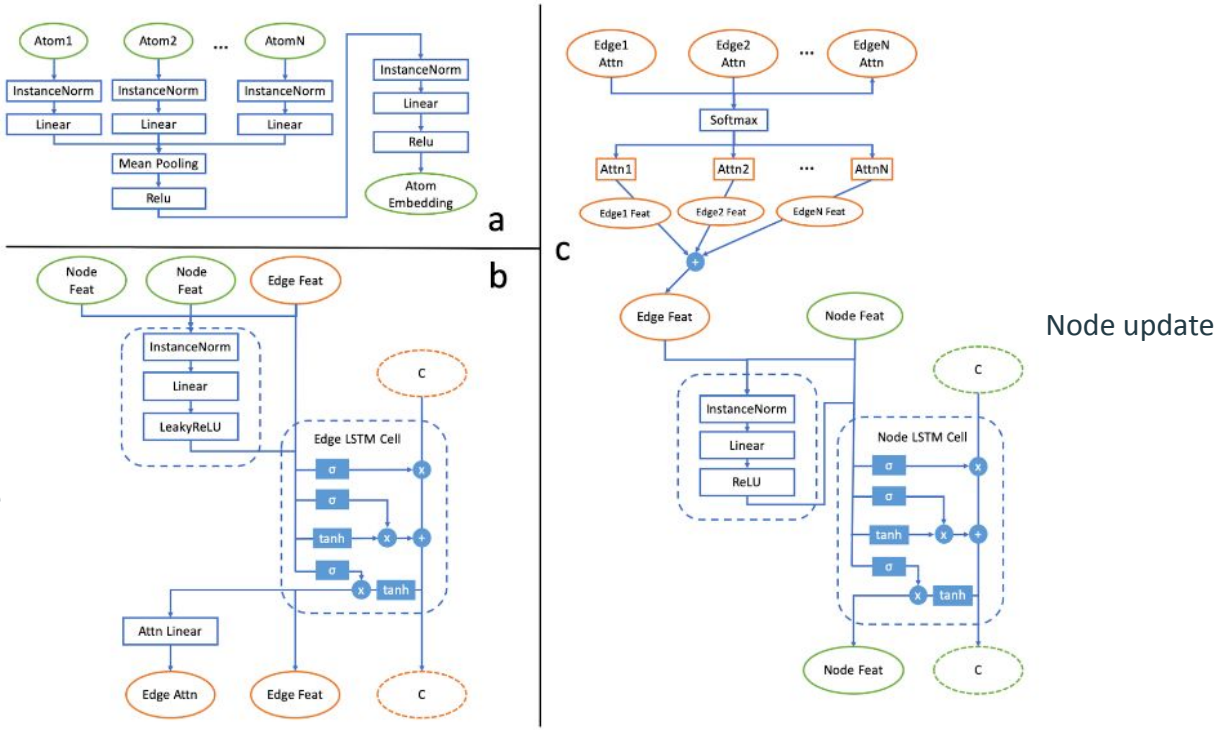
Architecture

- Atom embedding layer
  - 1
- Message passing layers
  - 10
- Output layer
  - 1



Architecture (2)

Atom embedder



Edge update

Node update

Features

- Node and Edge features
  - Sequential
  - Structural
- Atom features
  - One hot encoding of 4 atoms (C,N,O,S)
  - Relative coordinates
    - $x\text{-}x_{\alpha}$, $y\text{-}y_{\alpha}$,$z\text{-}z_{\alpha}$

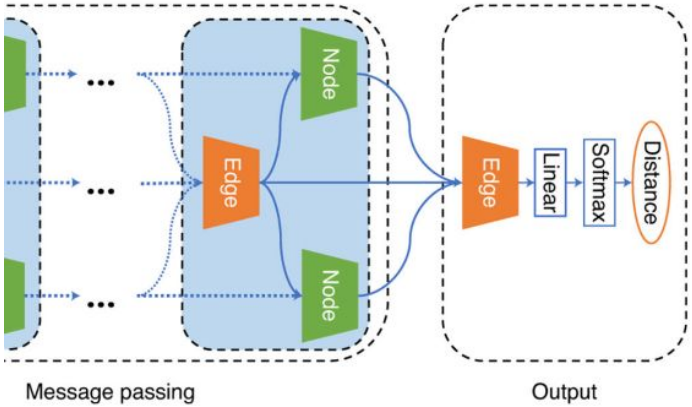
Type	Feature	Dimension
Residue	One-hot encoding of residue	21
	rPosition	1
	Dihedral, SS3 and RSA calculated by DSSP	6
	One-hot encoding and relative coordinate of heavy atoms	7
Residue pair	Distance ($C\alpha C\alpha$, $C_{\beta}C_{\beta}$ and NO)	3
	Orientation (ω, θ and φ)	3
	Sequential separation	9

Features (2)

- Graph construction
  - Edge definition (contact): C_β - C_β euclidean distance <10 Å
  - SE(3) - Equivariant
- 3 inter-residue orientations
  - As defined in trRosetta
- Sequential separation of the two residues (|i-j|)
  - Discretized into 9 bins
    - [1, 2, 3, 4, 5, 6–10, 11–15, 16–20, >20]

Output layer

- Softmax
  - Get distance distribution from edge feat following trRosetta
  - Output
    - Ex256⟶ Ex37
  - 37-dimensional vector
    - 36 bins representing the distances from 2 to 20 Å (0.5Å each)
    - 2 - 2.5 Å - bin 0
      2.5 - 3Å - bin 1..
    - 1 bin indicating the distance >20Å
- Loss function
  - Cross entropy loss (Ex37, actual bin/distance)



Building refined models

- Convert predicted probability distribution into distance potentials
  - SPLINE restraints
- Full atom relaxation, side-chain packing, energy optimization
  - PyRosetta, FastRelax (ref2015)
- Sequential inference; training likely stuck in local minima
  - 5 trained models on 5 different splits
  - GNNRefine - 5 models
  - GNNRefinePlus - 50 models
    - Rank by energy, reduce down to 5

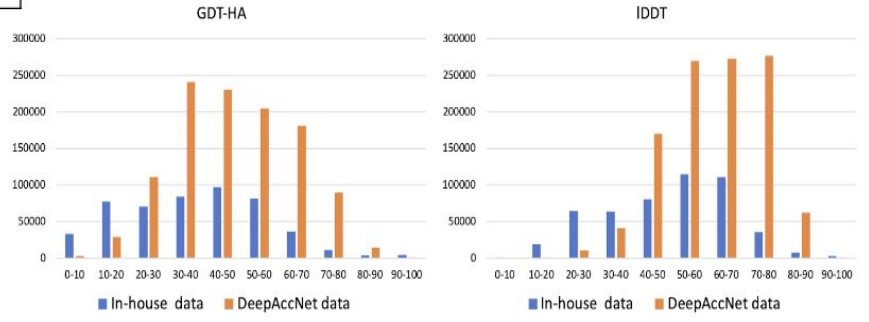
Distance (Å)	Probability	Energy (kcal/mol)
4.5	0.2	$E(4.5) = -\ln(0.2) \approx 1.61$
5.0	1.0	$E(5.0) = -\ln(1.0) = 0.0$
5.5	0.7	$E(5.5) = -\ln(0.7) \approx 0.36$
6.0	0.1	$E(6.0) = -\ln(0.1) \approx 2.30$

GNNQA

- How to select the best refined model out of these 5 models?
- Developed their own QA methods
  - Based on updated node features
  - Targets labels are global lDDT metric

Training datasets

Dataset	Source	#Targets		#Decoy or starting models	
In-house training data	CASP 7-12	592	29455	115857	500255
	CATH	28863		384398	
DeepAccNet training data	PISCES	7992		1104080	
Test	CASP13	28		28	
	CASP14	37		37	
	CAMEO	208		208	
	CASP13 FM	28		4193	



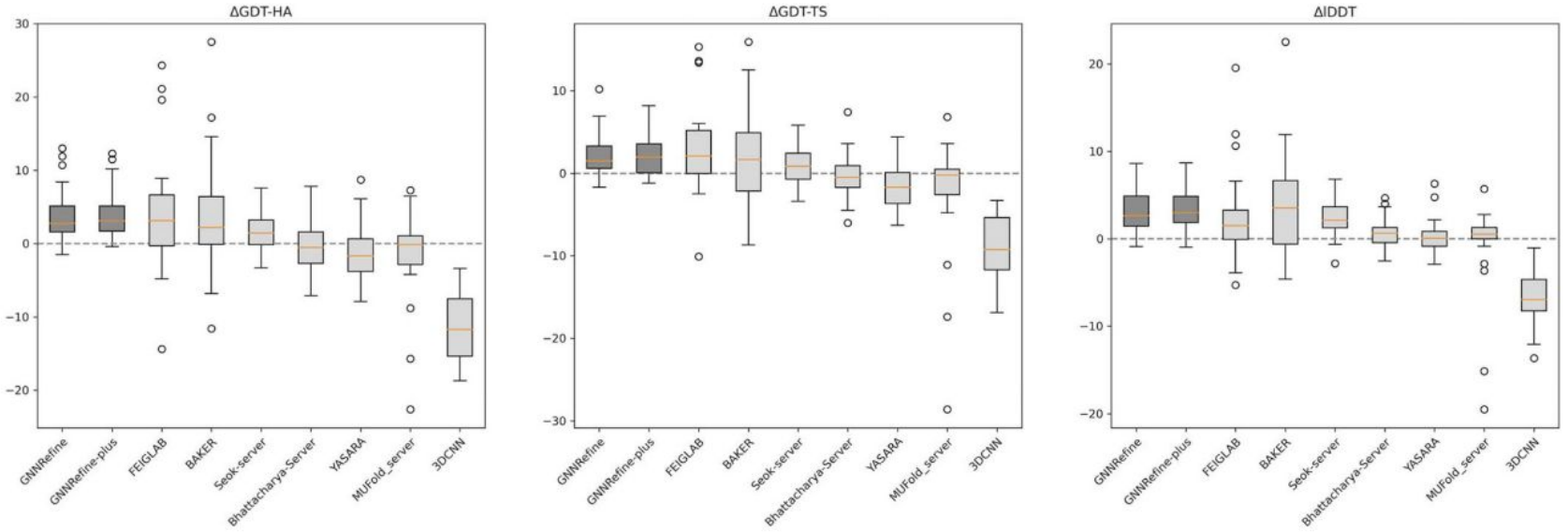
Evaluation metrics

- Global distance test
  - GDT-TS
  - GDT-HA
- Local distance difference test (lDDT)
  - 15 Å radius
- Range 0-100,higher the better
- “Degradation”
  - The number of refined models with worse quality than their initial models by a given threshold (0, −1 and −2).
  - 0 denotes that a refined model has worse GDT-HA than its starting model
  - −1 and −2 denote that a refined model is worse than its starting model by at least 1 and 2 GDT-HA units, respectively.

Performance on the 28 CASP13 refinement targets

Type	Methods	GDT-HA	GDT-TS	lDDT	Degradation		
					0	−1	−2
	Starting	52.27	71.51	61.74			
Human	FEIGLAB	+4.04	+2.97	+2.48	8	6	4
	BAKER	+3.35	+1.86	+3.73	7	6	6
Server	GNNRefine	+3.83	+2.31	+3.19	3	1	0
	GNNRefine-plus	+3.90	+2.31	+3.33	4	0	0
	Seok-server	+1.73	+0.89	+2.23	7	3	1
	Bhattacharya-Server	−0.44	−0.37	+0.64	17	12	8
	YASARA	−1.23	−1.57	+0.26	18	16	13
	MUFold_server	−1.61	−2.33	−0.70	13	11	7
	3DCNN	−11.47	−8.78	−6.92	22	22	22

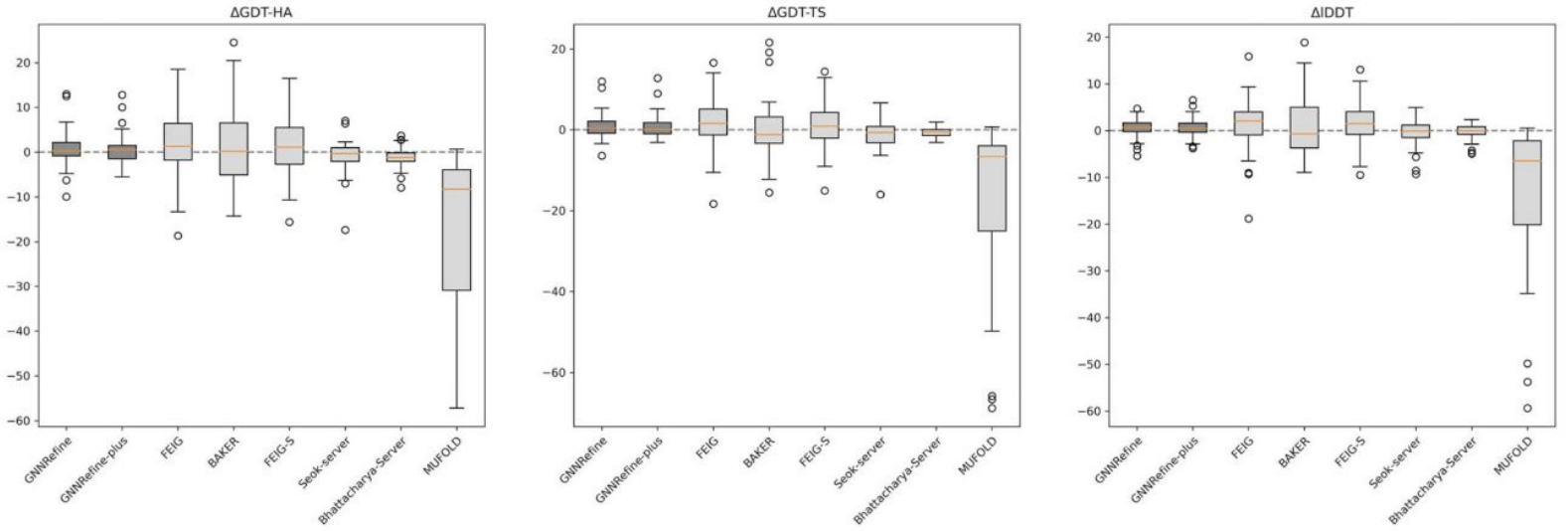
Performance on the 28 CASP13 refinement targets (2)



Performance on the 37 CASP14 refinement targets

Type	Methods	GDT-HA	GDT-TS	lDDT	Degradation		
					0	−1	−2
	Starting	54.12	72.65	65.98			
Human	FEIG	+2.01	+1.49	+1.13	14	12	9
	BAKER	+1.13	−0.03	+0.90	17	15	13
Server	GNNRefine	+0.84	+0.82	+0.50	17	9	7
	GNNRefine-plus	+0.80	+0.77	+0.67	14	10	6
	FEIG-S	+1.59	+1.05	+1.16	15	14	11
	Seok-server	−1.14	−1.32	−0.52	21	15	11
	Bhattacharya-Server	−1.24	−0.68	−0.45	29	22	10
	MUFOLD	−15.37	−17.91	−13.28	36	35	32

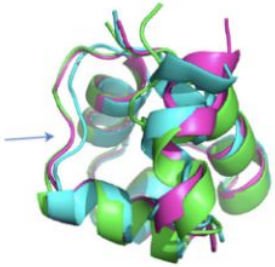
Performance on the 37 CASP14 refinement targets (2)



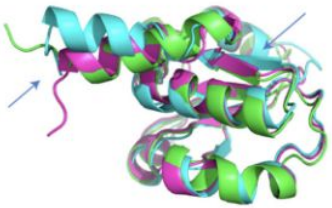
Deep learning vs Human input

- GNNRefine doesn't use MSA and only use start model.
- FEIG-S uses template, BAKER uses MSA.
- For 14 targets FEIG used the other models of the same server generating the starting model assigned by CASP14.
- BAKER used inter-residue distance predicted by trRosetta4 from MSAs as an input feature of DeepAccNet.
- Model selection is manual.

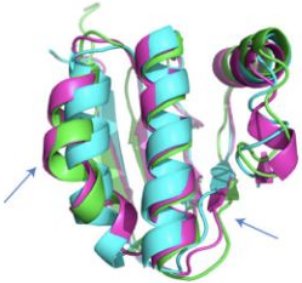
Case study (ΔGDT-HA >=10)



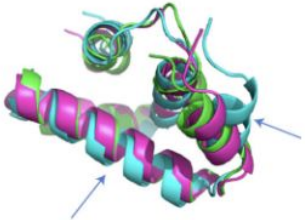
R0974s1 (GDT-HA: 65.58→79.35)



R0976-D2 (GDT-HA: 64.31→76.61)



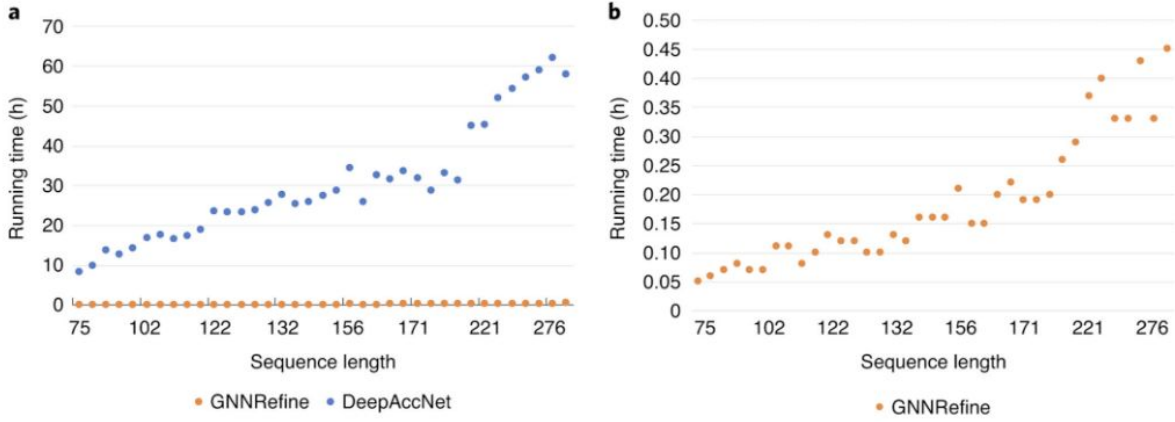
R0993s2 (GDT-HA: 50.00→63.01)



R1082 (GDT-HA: 53.33→66.33)

Run time

- Run time to refine one structure
  - GNNRefine
    - 10 minutes 1 cpu
  - BAKER
    - 30 hours 60 cpus
  - FEIG
    - 16 GPU hours



CASP14 excluding AlphaFold2 targets

All start models

Type	Methods	GDT-HA	GDT-TS	lDDT	Degradation		
					0	−1	−2
	Starting	54.12	72.65	65.98			
Human	FEIG	+2.01	+1.49	+1.13	14	12	9
	BAKER	+1.13	−0.03	+0.90	17	15	13
Server	GNNRefine	+0.84	+0.82	+0.50	17	9	7
	GNNRefine-plus	+0.80	+0.77	+0.67	14	10	6
	FEIG-S	+1.59	+1.05	+1.16	15	14	11
	Seok-server	−1.14	−1.32	−0.52	21	15	11
	Bhattacharya-Server	−1.24	−0.68	−0.45	29	22	10
	MUFOLD	−15.37	−17.91	−13.28	36	35	32

Excluding 7 AF2 start models

Type	Methods	GDT-HA	GDT-TS	lDDT	Degradation		
					0	-1	-2
	Starting	50.40	69.6	62.51			
Human	FEIG	+4.59	+3.48	+3.36	7	5	2
	BAKER	+2.71	+1.06	+2.34	11	9	8
Server	**GNNRefine**	**+1.94**	**+1.51**	**+1.26**	**10**	**3**	**3**
	GNNRefine-plus	**+1.63**	**+1.29**	**+1.35**	**8**	**4**	**2**
	FEIG-S	+3.75	+2.64	+2.86	8	7	5
	Seok-server	+0.27	-0.09	+0.58	14	8	4
	Bhattacharya-Server	-0.54	-0.39	+0.27	22	15	4
	MUFOLD	-12.32	-15.45	-10.75	29	29	26

Performance on 7 CASP14 AF2 refinement targets

Type	Methods	GDT-HA	GDT-TS	lDDT	Degradation		
					ΔGDT-HA< -5	ΔGDT-TS < -5	ΔlDDT < -5
	Starting	70.13	85.56	80.84			
Human	FEIG	-9.04	-7.06	-8.39	4	4	6
	BAKER	-5.61	-4.69	-5.26	5	3	4
Server	GNNRefine	**-3.84**	**-2.14**	**-2.76**	**2**	**1**	**1**
	GNNRefine-plus	**-2.69**	**-1.49**	**-2.24**	**1**	**0**	**0**
	FEIG-S	-7.67	-5.73	-6.14	5	3	4
	Seok-server	-7.16	-6.59	-5.23	5	5	3
	Bhattacharya-Server	-4.23	-1.96	-3.51	2	0	0
	MUFOLD	-28.44	-28.47	-24.12	6	5	7

Performance of standalone tools on CASP13

Methods	GDT-HA	GDT-TS	lDDT	Degradation			Running Time[1]
				0	-1	-2	
Starting	52.27	71.51	61.74				
GNNRefine	**+3.83**	**+2.31**	**+3.19**	**3**	1	**0**	**~0.16**
GalaxyRefine	+0.21	+0.05	+1.23	15	8	4	~2.53
ModRefiner-100	+0.16	+0.05	+0.73	13	3	0	~0.57
ModRefiner-50	-0.05	+0.04	+0.78	11	6	0	~0.59
ModRefiner-0	-0.70	-0.58	+0.99	17	12	2	~0.54
FastRelax	-2.00	-1.96	+0.17	18	14	13	~0.03
3DRefine	+0.24	+0.12	+0.30	9	**0**	0	--[2]
ReFold	-0.45	-0.24	-0.04	17	14	6	--[2]

Feature ablation

Features	Training data	GDT-HA	GDT-TS	lDDT	Degradation		
					0	−1	−2
All features	In-house	+3.15	+1.96	+2.88	1	0	0
All features	DeepAccNet data	+3.19	+1.75	+2.74	3	1	1
All features	CASP models only	+1.42	+0.92	+1.35	8	6	3
no Orientation	In-house	+2.21	+1.28	+2.26	4	2	0
no Dihedral&SS&RSA	In-house	+2.53	+1.67	+2.31	2	0	0
no AtomEmb	In-house	+3.25	+2.03	+2.57	2	0	0
AtomEmb (with local frame)	In-house	+3.05	+1.82	+2.50	3	1	1

Limitations or Future works

- Not an end-to-end refinement model
  - Energy minimization
- If model is poor→then graph is poor
  - Can’t detect or recover edges
- Poor performance on AF2 models
  - Better training set with higher model quality
- No MSA
- Chemical context in atom embedding