Modeling All-Atom Glycan Structures via Hierarchical Message Passing and Multi-Scale Pre-training

Minghao Xu

Glycans are ubiquitous

Starch

Bread

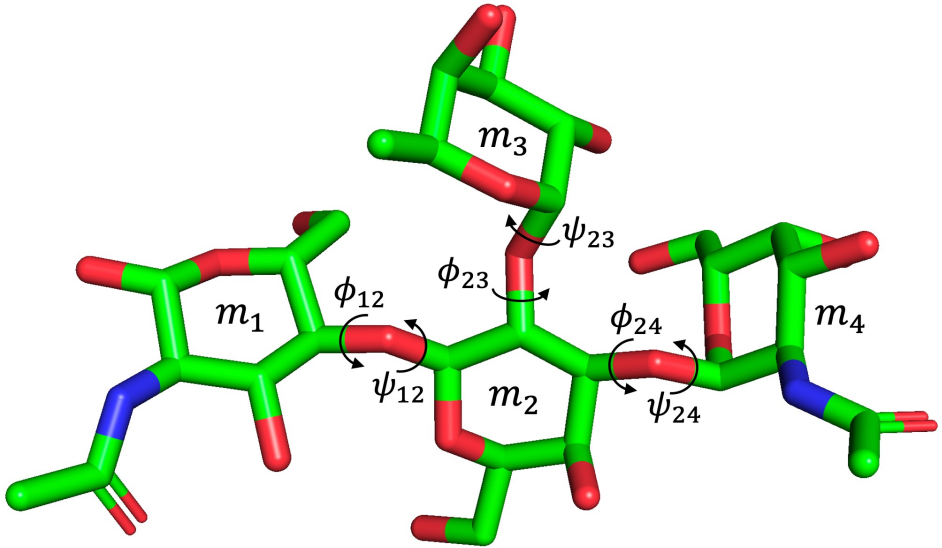
Rice

Sugars

Orange
(fructose)

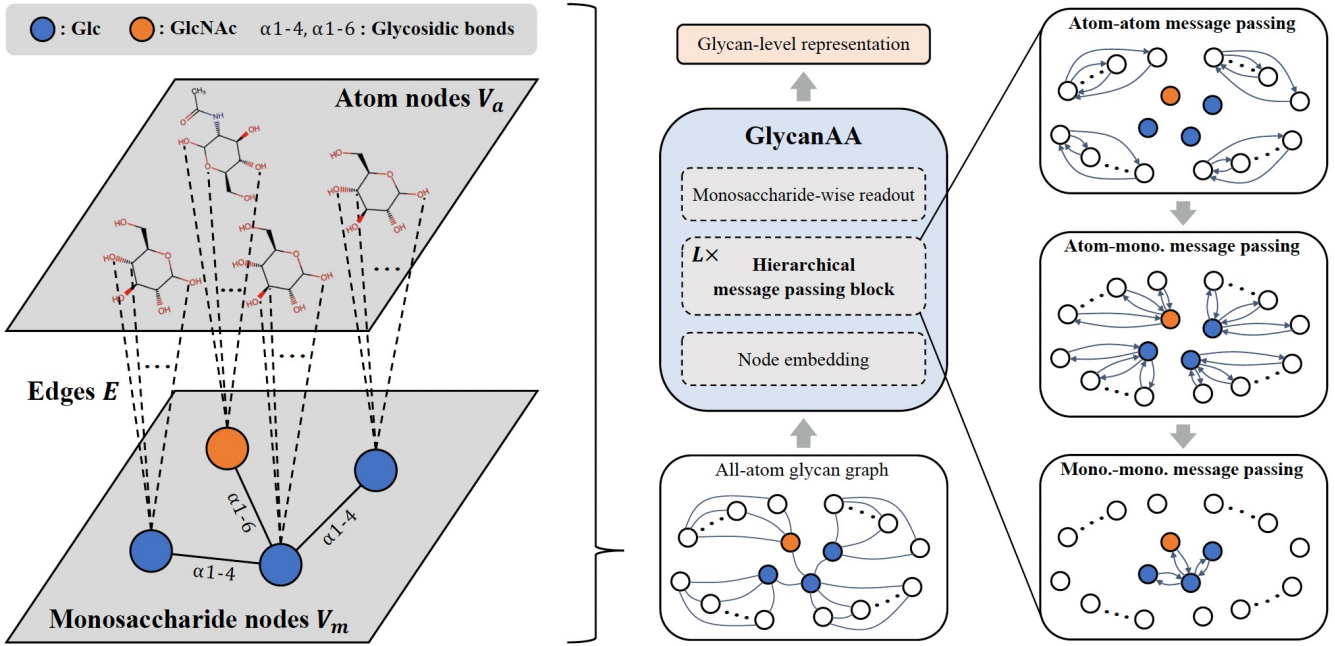
Milk
(lactose)

Atomic interactions are Determinants of glycan functions



Glycan functions: inducing immune response, glycosylation, acting as substrates in enzymatic reactions

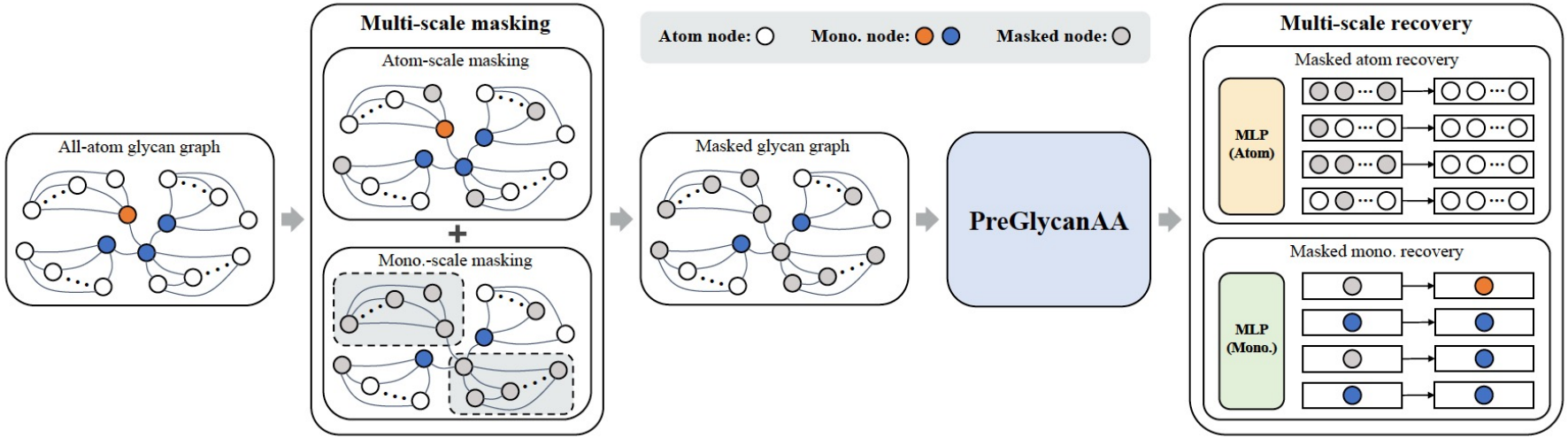
GlycanAA: Modeling all-atom glycan structures via Hierarchical message passing



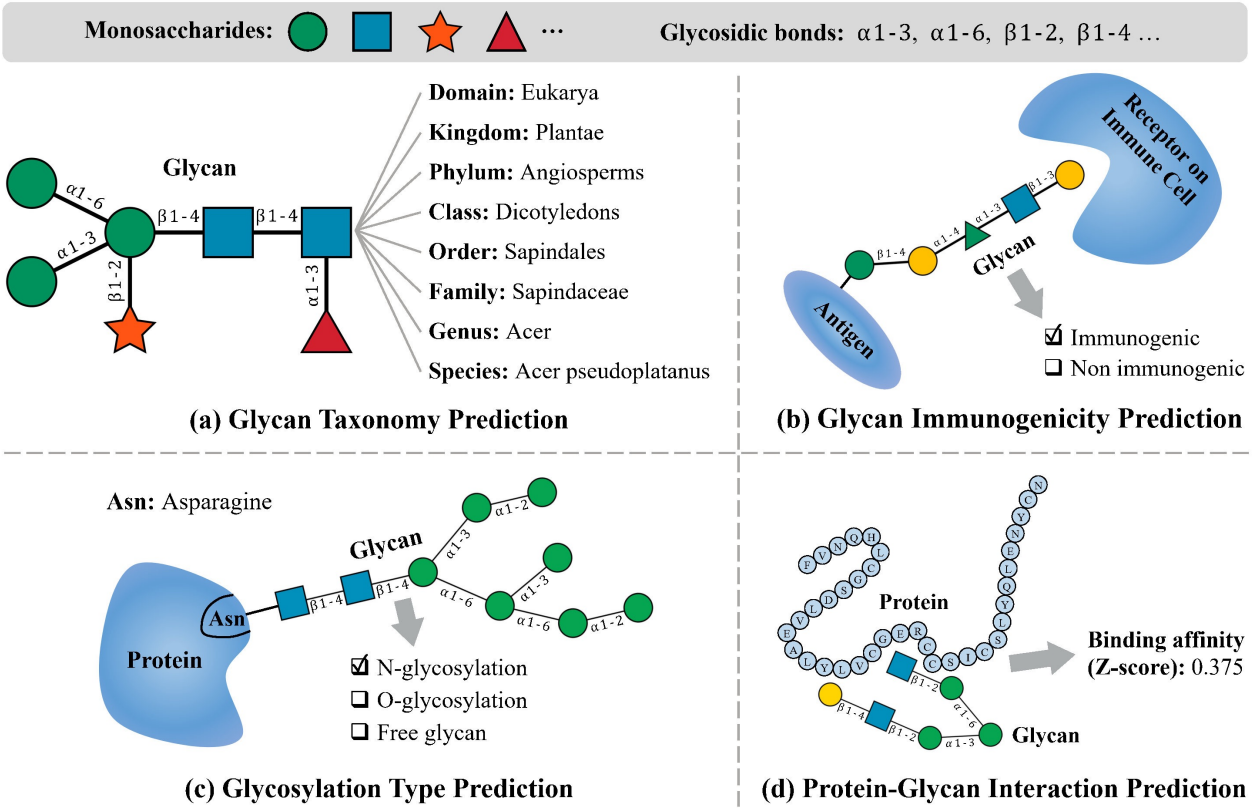
(a) All-Atom Glycan Graph

(b) GlycanAA with Hierarchical Message Passing

PreGlycanAA: Boosting all-atom glycan understanding by Multi-scale pre-training



Evaluating on diverse glycan understanding tasks



GlycanAA outperforms previous models PreGlycanAA achieves SOTA performance

Model	Taxonomy prediction								Immuno. (*AUPRC*)	Glycos. (*Macro-F1*)	Interaction (*Spearman's ρ*)	Weighted Mean Rank
	Domain (*Macro-F1*)	Kingdom (*Macro-F1*)	Phylum (*Macro-F1*)	Class (*Macro-F1*)	Order (*Macro-F1*)	Family (*Macro-F1*)	Genus (*Macro-F1*)	Species (*Macro-F1*)				
Monosaccharide-level Glycan Sequence Encoders												
Transformer	0.612$_{(0.009)}$	0.546$_{(0.079)}$	0.316$_{(0.014)}$	0.235$_{(0.022)}$	0.147$_{(0.007)}$	0.114$_{(0.039)}$	0.065$_{(0.001)}$	0.047$_{(0.008)}$	0.856$_{(0.012)}$	0.729$_{(0.069)}$	0.244$_{(0.009)}$	16.09
Shallow CNN	0.629$_{(0.005)}$	0.559$_{(0.024)}$	0.388$_{(0.024)}$	0.342$_{(0.020)}$	0.238$_{(0.016)}$	0.200$_{(0.014)}$	0.149$_{(0.009)}$	0.115$_{(0.008)}$	0.776$_{(0.027)}$	0.898$_{(0.009)}$	0.261$_{(0.008)}$	12.53
LSTM	0.621$_{(0.012)}$	0.566$_{(0.076)}$	0.413$_{(0.036)}$	0.272$_{(0.029)}$	0.174$_{(0.023)}$	0.145$_{(0.012)}$	0.098$_{(0.016)}$	0.078$_{(0.008)}$	**0.912$_{(0.068)}$**	0.862$_{(0.016)}$	*0.280$_{(0.001)}$*	11.00
ResNet	0.635$_{(0.009)}$	0.505$_{(0.025)}$	0.331$_{(0.061)}$	0.301$_{(0.010)}$	0.183$_{(0.082)}$	0.165$_{(0.019)}$	0.112$_{(0.018)}$	0.073$_{(0.007)}$	0.754$_{(0.124)}$	0.919$_{(0.004)}$	0.273$_{(0.004)}$	12.09
Monosaccharide-level Glycan Graph Encoders												
MPNN	0.632$_{(0.007)}$	0.638$_{(0.050)}$	0.372$_{(0.019)}$	0.326$_{(0.015)}$	0.235$_{(0.046)}$	0.161$_{(0.004)}$	0.136$_{(0.008)}$	0.104$_{(0.009)}$	0.674$_{(0.119)}$	0.910$_{(0.006)}$	0.217$_{(0.002)}$	18.34
GCN	0.635$_{(0.001)}$	0.527$_{(0.006)}$	0.325$_{(0.024)}$	0.237$_{(0.009)}$	0.147$_{(0.005)}$	0.112$_{(0.010)}$	0.095$_{(0.009)}$	0.080$_{(0.006)}$	0.688$_{(0.023)}$	0.914$_{(0.011)}$	0.233$_{(0.009)}$	18.38
GAT	0.636$_{(0.003)}$	0.523$_{(0.007)}$	0.301$_{(0.014)}$	0.265$_{(0.012)}$	0.190$_{(0.009)}$	0.130$_{(0.005)}$	0.125$_{(0.010)}$	0.103$_{(0.009)}$	0.685$_{(0.053)}$	0.934$_{(0.038)}$	0.229$_{(0.002)}$	16.94
GIN	0.632$_{(0.004)}$	0.525$_{(0.007)}$	0.322$_{(0.046)}$	0.300$_{(0.027)}$	0.179$_{(0.002)}$	0.152$_{(0.005)}$	0.116$_{(0.022)}$	0.105$_{(0.011)}$	0.716$_{(0.051)}$	0.924$_{(0.013)}$	0.249$_{(0.004)}$	15.06
CompGCN	0.629$_{(0.004)}$	0.568$_{(0.047)}$	0.410$_{(0.013)}$	0.381$_{(0.024)}$	0.226$_{(0.011)}$	0.193$_{(0.012)}$	0.166$_{(0.009)}$	0.138$_{(0.014)}$	0.692$_{(0.006)}$	0.945$_{(0.002)}$	0.257$_{(0.004)}$	12.19
RGCN	0.633$_{(0.001)}$	0.647$_{(0.054)}$	*0.462$_{(0.033)}$*	0.373$_{(0.036)}$	0.251$_{(0.012)}$	0.203$_{(0.008)}$	0.164$_{(0.003)}$	0.146$_{(0.004)}$	0.780$_{(0.006)}$	0.948$_{(0.004)}$	0.262$_{(0.005)}$	6.78
PreRGCN	0.636$_{(0.005)}$	0.664$_{(0.032)}$	0.451$_{(0.023)}$	0.389$_{(0.016)}$	0.265$_{(0.015)}$	0.205$_{(0.006)}$	0.172$_{(0.010)}$	0.139$_{(0.008)}$	0.781$_{(0.019)}$	*0.949$_{(0.015)}$*	0.263$_{(0.018)}$	*5.34*
GearNet	0.471$_{(0.005)}$	0.577$_{(0.036)}$	0.395$_{(0.025)}$	0.389$_{(0.010)}$	0.256$_{(0.007)}$	0.189$_{(0.004)}$	0.165$_{(0.003)}$	0.136$_{(0.003)}$	0.740$_{(0.015)}$	0.892$_{(0.027)}$	0.248$_{(0.004)}$	15.66
GearNet-Edge	0.628$_{(0.009)}$	0.573$_{(0.030)}$	0.396$_{(0.010)}$	0.384$_{(0.010)}$	0.262$_{(0.006)}$	0.200$_{(0.010)}$	0.177$_{(0.008)}$	0.140$_{(0.005)}$	0.768$_{(0.023)}$	0.909$_{(0.010)}$	0.250$_{(0.003)}$	12.25
ProNet	0.627$_{(0.007)}$	0.590$_{(0.015)}$	0.438$_{(0.012)}$	0.380$_{(0.008)}$	0.242$_{(0.005)}$	0.192$_{(0.018)}$	0.146$_{(0.010)}$	0.128$_{(0.004)}$	0.778$_{(0.019)}$	0.930$_{(0.015)}$	0.252$_{(0.002)}$	10.31
All-Atom Glycan Encoders												
All-Atom RGCN	0.637$_{(0.001)}$	0.624$_{(0.007)}$	0.293$_{(0.014)}$	0.156$_{(0.028)}$	0.112$_{(0.023)}$	0.096$_{(0.006)}$	0.063$_{(0.007)}$	0.035$_{(0.005)}$	0.520$_{(0.017)}$	0.928$_{(0.017)}$	0.215$_{(0.003)}$	19.88
Graphormer	*0.640$_{(0.006)}$*	0.468$_{(0.054)}$	0.249$_{(0.041)}$	0.201$_{(0.013)}$	0.142$_{(0.019)}$	0.112$_{(0.009)}$	0.077$_{(0.006)}$	0.054$_{(0.044)}$	0.637$_{(0.062)}$	0.856$_{(0.009)}$	0.211$_{(0.027)}$	22.91
GraphGPS	0.477$_{(0.002)}$	0.511$_{(0.040)}$	0.314$_{(0.022)}$	0.261$_{(0.051)}$	0.153$_{(0.018)}$	0.134$_{(0.008)}$	0.105$_{(0.006)}$	0.065$_{(0.017)}$	0.637$_{(0.075)}$	0.883$_{(0.032)}$	0.247$_{(0.016)}$	20.38
Uni-Mol+	0.639$_{(0.004)}$	0.446$_{(0.034)}$	0.227$_{(0.023)}$	0.174$_{(0.019)}$	0.128$_{(0.020)}$	0.109$_{(0.017)}$	0.077$_{(0.012)}$	0.056$_{(0.003)}$	0.789$_{(0.099)}$	0.885$_{(0.045)}$	0.241$_{(0.007)}$	16.56
GlycanAA-SP	0.589$_{(0.073)}$	0.635$_{(0.078)}$	0.444$_{(0.019)}$	0.395$_{(0.009)}$	*0.270$_{(0.006)}$*	0.205$_{(0.005)}$	0.176$_{(0.015)}$	0.154$_{(0.009)}$	0.755$_{(0.010)}$	0.946$_{(0.017)}$	0.241$_{(0.003)}$	11.22
GlycanAA-AN	0.609$_{(0.028)}$	*0.685$_{(0.001)}$*	0.453$_{(0.037)}$	*0.427$_{(0.027)}$*	*0.270$_{(0.009)}$*	0.199$_{(0.012)}$	0.179$_{(0.007)}$	*0.155$_{(0.003)}$*	0.765$_{(0.024)}$	0.947$_{(0.025)}$	0.241$_{(0.004)}$	10.44
GlycanAA	0.642$_{(0.002)}$	0.683$_{(0.002)}$	0.484$_{(0.009)}$	0.429$_{(0.022)}$	0.291$_{(0.003)}$	0.221$_{(0.002)}$	0.198$_{(0.011)}$	0.157$_{(0.011)}$	0.792$_{(0.021)}$	0.950$_{(0.020)}$	0.288$_{(0.003)}$	2.56
Pre-trained All-Atom Glycan Encoders												
VabsNet	0.607$_{(0.004)}$	0.622$_{(0.022)}$	0.363$_{(0.006)}$	0.261$_{(0.023)}$	0.175$_{(0.015)}$	0.125$_{(0.003)}$	0.104$_{(0.005)}$	0.068$_{(0.006)}$	0.742$_{(0.040)}$	0.903$_{(0.015)}$	0.160$_{(0.008)}$	19.03
GlycanAA-Attribute	0.628$_{(0.007)}$	0.687$_{(0.001)}$	0.457$_{(0.028)}$	0.392$_{(0.033)}$	0.263$_{(0.011)}$	*0.208$_{(0.004)}$*	*0.188$_{(0.001)}$*	0.143$_{(0.003)}$	0.722$_{(0.009)}$	0.925$_{(0.011)}$	0.263$_{(0.009)}$	10.47
GlycanAA-Context	0.637$_{(0.002)}$	0.643$_{(0.048)}$	0.453$_{(0.026)}$	0.386$_{(0.038)}$	0.259$_{(0.033)}$	0.205$_{(0.005)}$	0.177$_{(0.004)}$	0.144$_{(0.007)}$	0.768$_{(0.013)}$	0.946$_{(0.018)}$	0.270$_{(0.010)}$	7.06
PreGlycanAA	**0.661$_{(0.025)}$**	**0.688$_{(0.001)}$**	**0.502$_{(0.018)}$**	**0.447$_{(0.014)}$**	**0.297$_{(0.005)}$**	**0.233$_{(0.010)}$**	**0.203$_{(0.003)}$**	**0.174$_{(0.004)}$**	*0.850$_{(0.044)}$*	**0.961$_{(0.011)}$**	**0.297$_{(0.002)}$**	**1.5**

Thank you!

Paper

Code