Supporting Information: Identifying signatures of proteolytic stability and monomeric propensity in *O*-glycosylated insulin using molecular simulation

Wei-Tse Hsu[1], **Dominique A. Ramirez**[2], **Tarek Sammakia**[3], **Zhongping Tan**[4], **Michael R. Shirts**[1]

[1]Department of Chemical & Biological Engineering, University of Colorado Boulder, Boulder, CO, USA 80309; [2]Department of Biochemistry, University of Colorado Boulder, Boulder, CO, USA 80309; [3]Department of Chemistry, University of Colorado Boulder, Boulder, CO, USA 80309; [4]Institute of Materia Medica, Chinese Academy of Medical Sciences, Peking Union Medical College, Beijing, 100050, China

***For correspondence:**
michael.shirts@colorado.edu (MRS); zhongping.tan@imm.pumc.edu.cn (ZT)

Submitted in the Journal of Computer-Aided Molecular Design.

1 Supplemental Tables

	4EYD	4EY1	3I3Z	4EY9	2MVC
Total charges	-1	-1	-1	-1	-1
pH value	8.0	8.0	7.9	6.9	7.3
HisB5	HIP (+1)	HIP (+1)	HIP (+1)	HIP (+1)	HIE (+0)
HisB10	HIE (+0)	HIE (+0)	HIE (+0)	HID (+0)	HIP (+1)

Table S1. The comparison of histidine protonation states of wild-type structures.

	GF 2	GF 3	GF 4	GF 5	GF 6	GF 7	GF 8	GF 9	GF 10	GF 11	GF 12	GF 13
4EYD	NA	NA	3.45	63.56	69.11	NA	NA	45.79	121.70	21.65	140.40	24.77
4EY1	NA	NA	6.89	371.83	71.04	NA	19.37	205.24	159.14	58.17	360.45	637.20
4EY9	NA	NA	18.53	274.66	97.34	NA	NA	12.55	496.98	0.39	89.13	57.77
3I3Z	NA	NA	93.49	62.86	199.65	NA	NA	45.27	62.80	42.78	97.81	281.52
2MVC	3.98	NA	21.31	54.78	105.28	NA	NA	114.61	59.62	58.75	66.00	81.57

Table S2. Dimer occlusion autocorrelation lag times for each of the listed models. All numbers are listed in units of nanoseconds.

#	4EYD-based GFs	4EY1-based GFs	4EY9-based GFs	3I3Z-based GFs	2MVC-based GFs
1	N/A	N/A	N/A	N/A	N/A
2	None	ThrA8(OG1)-GalNAc[1](O2N):10%	None	None	GlnA5(NE2)-GalNAc[1](O6): 14% GlnA15(NE2)-GalNAc[1](O6): 13%
3	None	None	None	None	GlnA15(NE2)-52(O2N):11%
4	None	None	None	None	None
5	None	PheB24(N)-GalNAc[1](O3): 48% ThrB30(N)-GalNAc[1](O2N): 16%	PheB24(N)-GalNAc[1](O3): 11%	ThrB27(N)-GalNAc[1](O3):15% Glu4(N)-GalNAc[1](O2N): 14%	PheB24(N)-GalNAc[1](O3): 10%
6	ThrB27(N)-GalNAc[1](O3):26%	ThrB27(N)-GalNAc(1)(O3):26%	None	None	CysA7(N)-GalNAc[1](O2N): 18% ThrB27(N)-GalNAc(1)(O3): 12%
7	None	None	None	None	GlnA15(NE2)-Man[1](O4): 18%
8	GlnA5(NE2)-Man[2](O2): 12%	GlnA5(NE2)-Man[2](O2): 11%	None	GlnA5(NE2)-Man[1](O5): 12%	None
9	PheB24(N)-Man[1](O3): 22% TyrB16(OH)-Man[1](O4): 13%	None	PheB24(N)-Man[1](O3): 27% TyrB16(OH)-Man[1](O4): 19%	None	None
10	PheB24(N)-Man[1](O3): 52% ThrB30(N)–Man[2](O2): 35% ThrB27(N)–Man[2](O6): 23%	ThrB27(N)-Man[2](O6): 46% PheB24(N)-Man[1](O3): 30% TyrB16(OH)-Man[1](O4): 15%	ThrB27(N)-Man[2](O6): 45% PheB24(N)-Man[1](O3): 28% TyrB16(OH)-Man[1](O4): 11%	ThrB27(N)-Man[2](O6): 47% PheB24(N)-Man[1](O3): 15%	ThrB27(N)-Man[2](O6): 47% PheB24(N)-Man[1](O3): 37% TyrB16(OH)-Man[1](O4): 14%
11	None	None	None	None	None
12	ThrB30(N)-Man[2](O6): 31% GlyB23(N)-Man[1](O3): 10%	TyrA19(OH)-Man[1](O4): 51% ThrB27(N)-Man[1](O3): 43% ThrB30(N)-Man[2](O6): 19%	ThrB30(N)-Man[2](O6): 24% ValA3(N)-Man[2](O2): 15%	ThrB30(N)–Man[2](O6): 38%	ThrB30(N)-Man[2](O6): 49% GlyB8(N)-Man[2](O4): 10%
13	ThrB27(N)-Man[2](O6): 61% PheB24(N)-Man[1](O3): 49% TyrB16(OH)-Man[1](O4): 16%	ThrB27(N)-Man[2](O6): 31% ThrB27(N)-Man[3](O6): 14% PheB24(N)-Man[1](O3): 12% TyrB16(OH)-Man[1](O4): 11%	ThrB27(N)-Man[2](O6): 27% PheB24(N)-Man[1](O3): 17% ThrB27(N)-Man[3](O6): 15% TyrB16(OH)-Man[1](O4): 10%	ThrB27(N)-Man[2](O6): 43% PheB24(N)-Man[1](O3): 25% TyrB16(OH)-Man[1](O4): 12%	ThrB27(N)-Man[2](O6): 44% PheB24(N)-Man[1](O3): 22%

Table S3. The glycan-involved hydrogen bonds and their existence percentages of each glycoform.

Atom type	Role	Description
N	Donor	An sp^2 nitrogen in amide group
NE2	Donor	An epsilon nitrogen.
OH	Donor	An alcohol oxygen in Tyr
OG1	Donor	An alcohol oxygen in Thr
O2	Acceptor	The oxygen atom connected to the second carbon atom of the sugar
O3	Acceptor	The oxygen atom connected to the third carbon atom of the sugar
O4	Acceptor	The oxygen atom connected to the fourth carbon atom of the sugar
O5	Acceptor	The oxygen atom connected to the fifth carbon atom of the sugar
O6	Acceptor	The oxygen atom connected to the sixth carbon atom of the sugar
O2N	Acceptor	The oxygen atom of the N-acetyl group

Table S4. The atom types involved in the glycan-involved hydrogen bonds.

	least occlusion								*most occlusion*			
4EYD	2	3	7	8	4	11	12	6	**9**	**10**	5	**13**
4EY1	2	3	7	8	4	11	6	12	**9**	**10**	5	**13**
4EY9	2	3	7	8	11	4	6	12	5	**9**	**10**	**13**
3I3Z	2	3	7	8	4	6	11	12	**9**	5	**13**	**10**
2MVC	3	7	8	2	4	11	6	12	**9**	5	**10**	**13**
	low batch				*medium batch*				*high batch*			

Table S5. Glycoforms ordered from most to least proportion occlusion, based on proportion of simulation with measured occlusion.

2 Supplemental Figures

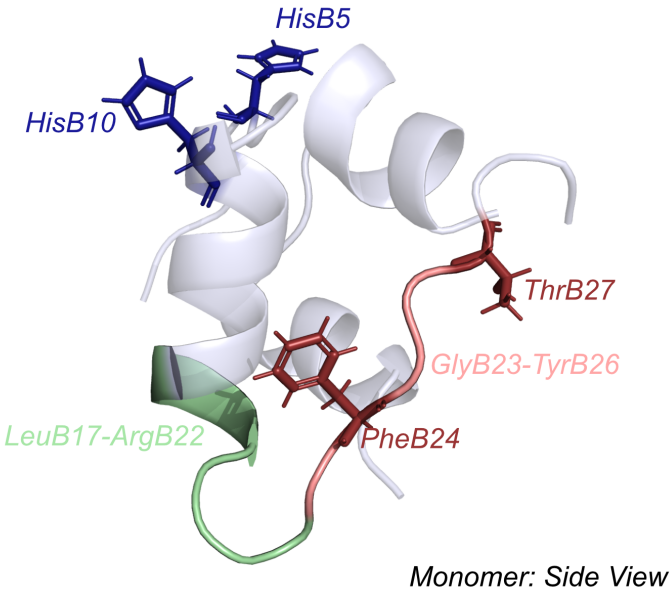

Figure S1. The histidine residues (HisB5, and HisB10, colored in blue) are shown with residues PheB24 and ThrB27 (colored red) which are important in enhancing proteolytic stability and residues LeuB17-ArgB22 (colored green) and GlyB23-TyrB26 (colored salmon), important in determining dimerization potential.

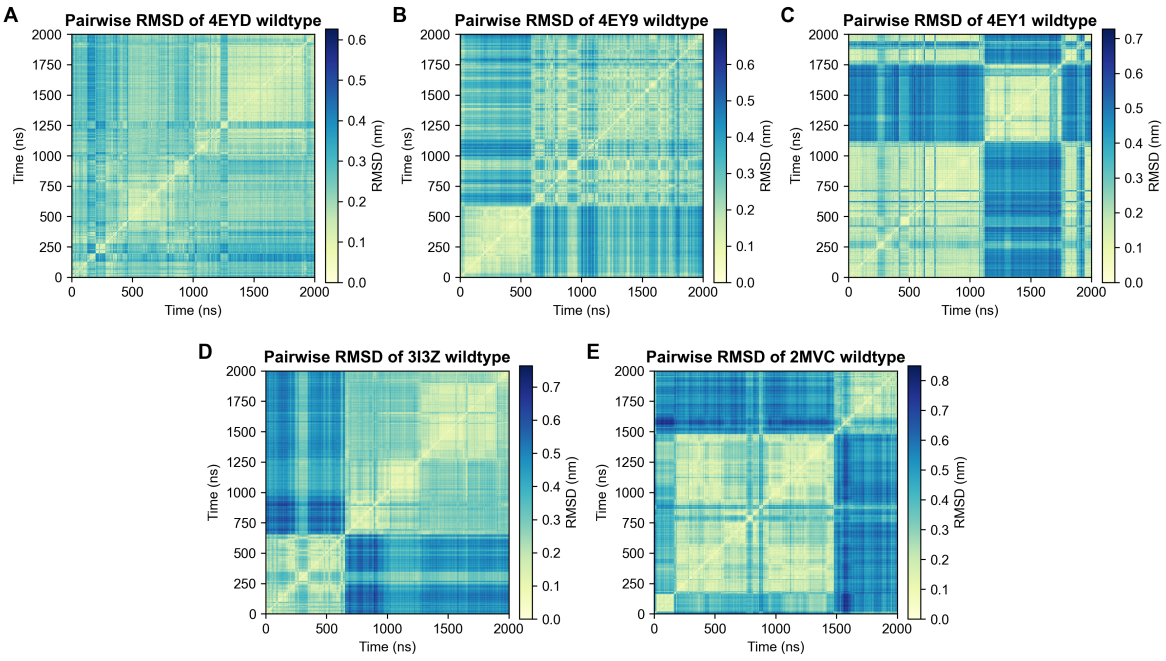

Figure S2. The pairwise RMSD of calculated from the 250ps-spaced MD trajectory of each wild-type model, including 4EYD (A), 4EY9 (B), 4EY1 (C), 3I3Z (D), and 2MVC (E). As shown in the figure, the major transitions occur around 500–1500 ns, we therefore concluded that at least 2000 ns was required to sample the configurational ensemble of insulin.

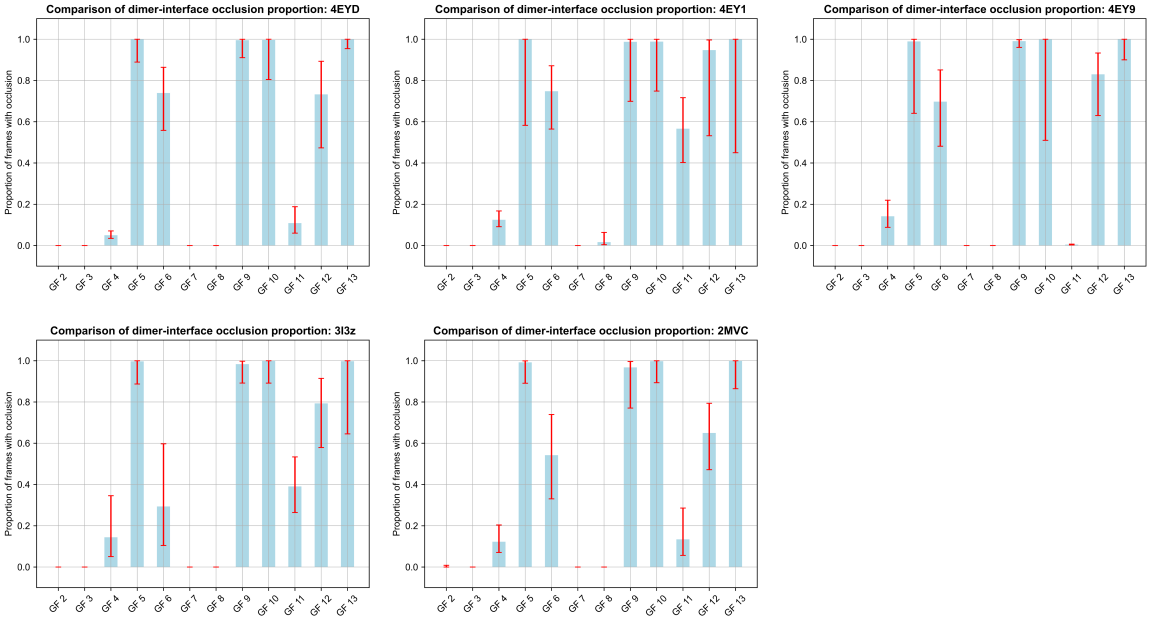

Figure S3. Proportion of frames with glycan-dimer occlusion for each glycoform. Red bars represent the asymmetric 95% Wilson score confidence interval.