

Response to reviewers' comments on "Identifying signatures of proteolytic stability and monomeric propensity in O-glycosylated insulin using molecular simulation"

Wei-Tse Hsu¹, Dominique A. Ramirez², Tarek Sammakia³, Zhongping Tan⁴, Michael R. Shirts¹

¹Department of Chemical & Biological Engineering, University of Colorado Boulder, Boulder, CO, USA 80309; ²Department of Biochemistry, University of Colorado Boulder, Boulder, CO, USA 80309; ³Department of Chemistry, University of Colorado Boulder, Boulder, CO, USA 80309; ⁴Institute 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)

Comments from Reviewer 1

To understand the impact of glycosylation on insulin proteolytic stability, Hsu et al. correlate structural features computed from molecular dynamics simulations of glycosylated insulins with experimentally measured alpha-chymotrypsin half lives. They find that the solvent accessible surface area of proteolytic sites, glycan-associated hydrogen bonds, and glycan-dimer occlusion are all partially predictive. Though quite focused in scope, I think the paper is a nice illustration of how computation can aid the evaluation of experimental data for molecular design. While the paper is clearly written overall, I have a number of suggestions/questions that I think can strengthen the study if addressed.

Comment: In the introduction, the authors discuss previous simulations of insulin. One highly relevant study that is missing is Busto-Moner et al. *Biochemistry* 2021, 60, 3125–3136. This study analyzes 165 microseconds of molecular dynamics data to characterize the structural ensemble of the human insulin monomer at low pH.

Our response: We thank the reviewers for drawing attention to this paper, published just before we submitted our paper, and we have added this study to the introduction.

Comment: Not only is this a useful comparison point for the present simulations, but I think it makes the authors remark at the bottom of p.3 that their aggregate simulation time is the longest for human insulin incorrect.

Our response: As the paper indeed exploited a longer simulation of insulin compared to our simulations, we have restated the aggregate length of our simulation to be "one of the longest" among all the computational studies of human insulin. We have also compared our simulations to this paper to better validate the computational ensemble, as detailed in Section 1 of the current supporting information. Specifically, the referred paper presents a pipeline for sampling wild-type insulin ensemble and concludes that 60% of insulin structures exhibits at least one of the following elements of disorder: melting of A-chain N-terminus helix (A1-A9), detachment of B-chain N-terminus (B1-B7) and detachment of B-chain C-terminus (B20-B30).

Despite the fact that the studied ensemble in the referred paper is not directly comparable to our simulations, given the significantly lower pH value it adopted, we still repeated the analysis presented in the referred paper, as there is likely to be a partial agreement in some of the metrics. The details of the analysis are presented in Section 1 of the supporting information. As a result, qualitative agreement can be observed from the results summarized in Supplemental Table S1 that our simulations did capture the 3 important elements of disorder described in the referred paper. Notably, we do not expect quantitative

agreement in fraction of the ensemble with these disordered elements because simulations done by Busto-Moner were carried out at pH 2.5, unlike the neutral pH of our study. In the main text, the analysis in the SI is referred to with the text "Our simulations captured the key important disordered elements investigated by Busto-Moner et al., though at different percentages. This discrepancy is not surprising given the pH difference between the two studies, with the differences elaborated in Section 1 in the supporting information."

Feb 18, 2022 12:23 PM

Comment: *Ref. 26 from the same group is a study of dimer dissociation, not monomer unfolding, as stated on p. 2.*

Our response: We modified the description about the referred paper titled "Insulin Dissociates by Diverse Mechanisms of Coupled Unfolding and Unbinding" mentioned on page 2 (Ref. 26 in both the previous and current bibliography). The original sentence is as follows:

"More recently, with steered molecular dynamics and replica-exchange umbrella sampling, Antoszewski et al. determined the relative energies of unfolding and identified different unfolding pathways of insulin."

The sentence has been modified to the following:

"More recently, with steered molecular dynamics and replica-exchange umbrella sampling, Antoszewski et al. revealed different pathways of insulin dimer dissociation and characterized relevant energetic and structural features of the process."

Comment: *The authors need to state the force field they are using.*

Our response: We have modified our main text to describe the adopted force fields for building proteins more clearly, specifically stating in the text: "After extracting the monomer structure from each of the initial models, we used the H++ server (version 3.2), which also parameterizes the protein with the AMBER ff14SB force field, to assign reasonable protonation states corresponding to the pH values adopted in experiments."

Comment: *In the text, on p. 5 it is stated that the pH is tuned around pH 8 such that all the systems have charge of -2, but in Table S1 all the charges are -1. Besides harmonizing these, it would be nice if the authors could discuss the impact of pH on their conclusions. The experiments in ref. 19 are at pH 8, while the physiological pH in the small intestine will be lower. Can the simulations be used to assess the impact of pH on the conclusions?*

Our response: In our study, we found pH ranges that given total charges of -1, so the original information in the main text (charge=-2) was incorrect. We apologize for this important error. To correct this, we have harmonized the values described in the main text and the supporting information so now they are all indicating total charges of -1. Also, we found that the pH values of the models 4EY1, 3I3Z, and 4EY9 were mistakenly described in the original Supplemental Table S1 (now Supplemental Table S3), especially that the pH values of 3I3Z and 4EY9 were accidentally switched. In addition to correcting the values, instead of presenting the maximum pH value that led to total charges of -1 as we originally did, now we present the whole range of allowed pH values that correspond to total charges of -1 to reduce confusion. For each model, any externally set pH value with H++ within the range leads to exactly the same protonation states of residues. Table 1 shows the original and updated pH values for Table S1 for each wild-type model.

	4EYD	4EY1	3I3Z	4EY9	2MVC
Original pH	8.0	8.0	7.9	6.9	7.3
Updated pH	6.9-7.8	6.9-7.8	6.6-6.9	7.7-7.9	6.5-7.3

Table 1. The original and update pH values of each wild-type model.

We note that the pH ranges corresponding to the fixed total charges of -1 will vary between different models, since the value from H++ is an approximation based on a single input structure instead of the entire ensemble. The fact that the ranges are not completely overlapping is not a serious problem, as all structures have the exact same sequence, and so the ranges are a reflection of the shortcomings of the

H++ protonation algorithm and the use of single crystal structures. As we chose to fix the total charges of insulin at -1, the allowed pH values summarized in Table 1 above collectively range from approximately 6.6 to 7.8. This range overlaps both the pH ranges in the small intestine and the pancreatic tract, which are pH ranges adopted in the experiments done by Guan et al. (Guan, Xiaoyang, et al. "Chemically precise glyco-engineering improves human insulin." ACS chemical biology 13.1 (2018): 73-81.) to mimic the environments where dimerization and proteolysis typically occur. We therefore expect that a pH change within this range would not have any noticeable impact on the usage of our metrics and our conclusions. We have updated the main text to incorporate this discussion of choosing total charges and pH values in our study on page 5.

Comment: *More information about the statistics of the metrics themselves is needed. The authors should give their distributions in the Supporting Information.*

Our response: We thank the reviewers for bringing this up. We have added a figure (Supplemental Figure S2) and its discussion in Section 3 of the supporting information, and described the presence of this SI data in the main text. The figure shows the distributions of the SASA at the B25-B26 scissile bond, B26-B27 scissile bond, residue B24, and residue B25. For all SASA measures, GF 13 peaks at apparently lower SASA values compared to other variants, which is consistent with our finding mentioned in the main text that Metric 1 and Metric 2 had better predictiveness for the more proteolytically stable variants.

Comment: *Figure S2 suggests that there are infrequent transitions within the simulations—are the metrics switching their ranges with these?*

Our response: We have added Section 2 in the supporting information to address this comment. Specifically, to see if the ranges of the metrics change significantly upon transitions between states, for each wild-type model, we repeated data analyses not involving the glycan moiety, for the states before and after the major transition in pairwise RMSD. These analyses include 8 measures involved in Metric 1 to Metric 3 (scissile bond SASA, P1-site SASA, and β -sheet propensity) for the proteolytic stability. As a result, SASA measures generally did not vary significantly in their values upon transition between states, while at least one of the β -sheet propensity measures (e.g. the β -sheet propensity at residue B25) showed a large change after the major transition occurred. This implies that our simulation might not comprehensively sample the configurational space of wild-type insulin, which is also reflected by the fact observed from the pairwise RMSD that the system did not have frequent major transitions back and forth between different states. However, frequent minor transitions between states shown in pairwise RMSD suggest that the simulation still captures a large amount of configurational diversity and long-timescale events. We emphasize that the goal of our study is not to comprehensively sample the whole configurational space of insulin and its glyco-variants but to develop reasonable metrics that work with MD simulations requiring manageable computational cost and distinguish variants with certain properties from their counterparts. A key point is that the uncertainties are computed over the five independently generated structures, meaning the variations within each individual run are less important.

Comment: *How relevant are the averages used in the main text?*

Our response: For each variant, the averages presented in the correlation plots in the main text are averages over the simulations started from 5 different wild-type models, and uncertainties are calculated over the 5 models. Since the transitions in the time series are less significant than the variations between the models, we conclude that the averages are still relevant even if there are transitions between states.

Comment: *Also, how are different metrics correlated with each other?*

Our response: In the supporting information, we have added correlation plots (Supplemental Figures S3 to S5) between the 8 measures involved in Metric 1 to Metric 3 for the proteolytic stability. As a result, correlations between any two measures involved in the more predictive metrics (the scissile bond SASA and P1-site SASA) (Figure S3) tend to be stronger compared to the ones that involve at least one measure of the less predictive metric (β -sheet propensity) (Figure S4 and Figure S5). This is expected because correlating

two measures that are highly associated with proteolytic stability naturally leads to a stronger correlation. The only exception is the correlation between β -sheet propensity of B22 and B23, which both are not predictive for the proteolytic stability but highly correlated with each other. This is probably because these two residues are adjacent to each other and contribute to similar secondary structures.

Comment: *I am confused by how the authors are presenting Kendall's tau statistics. It says in the text that p-values are calculated, but I don't see these. Instead, there is an uncertainty (+/-) reported. Are they related?*

Our response: We apologize for mistakenly using the p-value as the uncertainty of Kendall's tau correlation coefficients in the correlation plots. Now in the updated version, we present the uncertainty of the correlation coefficients determined from bootstrapping.

Specifically, in the main text, all the correlation plots are characterizing the relationship between one metric variable and the experimental data. Therefore, in each of the 500 bootstrap iterations we performed, for each variant, we drew 5 bootstrap samples for the metric variable from values based on different wild-type models, and 5 bootstrap samples for the experimental reference from normal distributions centered at the experimental values. From each of these 500 bootstrap iterations, we calculated one Kendall's tau correlation coefficient by correlating the sample mean of the two variables. Lastly, we calculated the standard deviation of these 500 values and reported it as the uncertainty of the correlation coefficient.

The same bootstrap approach was used to estimate the uncertainty of the Pearson correlation coefficients for the correlations between metrics reported in Section 4 in the supporting information. In this case, the bootstrap samples for both metrics of interest were drawn from values from each of the five simulations starting from different models. The details of the bootstrap methods have been included in the newly added section 2.4 in the main text.

Comment: *Would the authors gain predictive power by combining metrics (e.g., by thresholding each metric or a linear combination of them, a linear model, or a decision tree)?*

Our response: The difficulty of applying a linear combination or a decision tree of the metrics as a predictor is that it is hard to justify what weight or threshold should be used for each metric. Even if we figure out a set of weights for a linear combination that leads to correct classification, the lack of experimental data would cause high uncertainty of these weights or decision cuts. In our study, instead of finding the best way to predict the proteolytic stability and dimerization propensity of insulin, it is our focus to identify physically motivated, obvious structural features that could distinguish insulin glyco-variants with high proteolytic stability/low dimerization propensity from their counterparts.

We recognize the desirability of having more automated ways of identifying structural characteristics leading to predictability. Therefore, we are working on applying DiffNets (Ward, Michael D., et al. "Deep learning the structural determinants of protein biochemical properties by comparing structural ensembles with DiffNets." *Nature communications* 12.1 (2021): 1-12.), self-supervised autoencoders where the classification labels are updated with the expectation-maximization algorithm, to the MD trajectory of each variant. Mathematically, the hidden layers in the neural network combine the 3D coordinates of the systems of interest in a non-linear fashion. It has been shown that with the aid of an additional classification task, DiffNets are able to automatically find the structural determinants useful for distinguishing different structural protein ensembles. We hope that in a further study, DiffNets or equivalent tools could identify relevant features that capture what we found in our study. This discussion has been rephrased in the conclusions section in the main text to capture these ideas.

Comment: *In Figure 1, there is an alpha missing from GF 12.*

Our response: We have corrected Figure 1. Thanks for catching that!

Comment: *In Figure 2, the meaning of the green coloring is not given in the caption.*

Our response: We did not mean to color the region that was originally in green. We have removed the green color from the figure in the updated version of our paper.

Comment: On p. 4, the authors write "our study...is also unique as the first study that assesses proteolytic stability of a protein by molecular dynamics." The proteolytic stability is experimentally assessed. Molecular dynamics simulations are used to identify features that correlate with these measurements.

Our response: We thank the reviewer for this catch. We have changed the original sentence from: "In addition, our study is not only one of the very few studies that characterize conformations of insulin with a covalently attached moiety, but is also unique as the first study that assesses proteolytic stability of a protein." to now read: "In addition, our study is not only one of the very few studies that characterize conformations of insulin with a covalently attached moiety, but is also unique as the first study that correlates molecular dynamics simulations to the proteolytic stability of these proteins."

Comment: On p. 6, "swiching" should be "switching."

Our response: We have fixed that in the main text. Thanks for catching this!

Comment: On p. 8, the authors refer to "converted trajectories." I'm not sure to which trajectories these refer and what is meant by "converted."

Our response: In the original version, there are the following two sentences that contained "converted trajectories":

- (1) "We used the same converted trajectories described above to calculate this metric."
- (2) "We used Python package ProDy to calculate the total number of atom neighbor pairs between the glycan and dimer interface for each frame in the converted trajectories."

In the original text, we described the trajectories for data analysis as converted trajectories because we converted the timesteps of the output time series between adjacent time frames from 2 ps, as originally specified in the simulation, to 250 ps. The purpose was mainly for reducing the computational cost and memory required for performing data analysis. Note that this conversion had a negligible impact on the statistics of our data as our simulations are long enough for us to collect sufficient data.

To avoid confusion, in the updated text, we have removed the first sentence from the paragraph since we have mentioned the same conversion elsewhere in the paper. We have also removed the word "converted" in the second sentence.

Comment: On p. 13, there is a closing parenthesis missing after "Supplemental Table S3."

Our response: We have addressed this in the supporting information. Thank you!

Comments from Reviewer 2

Comment: The authors use MD simulations to identify the molecular basis of glycosylated insulin that lead to enhanced proteolytic stability or reduced propensity to dimerize. The former is related to stability in the gastrointestinal tract, and the latter is related to permeability through the intestinal epithelium. 12 insulin glycoforms are studied in this work. Metrics based on solvent accessible surface area (SASA) of scissile bonds, SASA of the P1 sites, and the β -sheet propensity of the P1-P3 region were found to be weakly correlated with alpha-chymotrypsin half-life. The existence of glycan-involved hydrogen bonds, however, is generally predictive of proteolytic stability. Furthermore, glycoforms whose glycans come in proximity to the dimer interface were found to be predictive of low dimerization propensity.

We are glad that the paper is clear enough that reviewer was able to succinctly summarize the findings.

Comment: Could sites P1—P3 be labeled in the figures?

Our response: To address this, we have added Figure 3 in the main text, where the P1-P2' sites are clearly labeled assuming different cleavage sites of interest.

Comment: Also, the Schechter-Berger nomenclature should be briefly explained in the text in order for this work to be more accessible to a broad audience.

Our response: To address this, we have added the following paragraph in the description for Metric 2 in

Section 2.2.1 in the main text: "In Schechter-Berger nomenclature, the residues N-terminal to the cleavage site are denoted as P1, P2, P3, ... etc, while the residues in the opposite direction are denoted as P1', P2', P3' ... etc."

Additional changes

In addition to the changes mentioned in the responses above, we made the following changes for the plots originally presented in the main text:

- (1) We replotted Figure 5A as we found that the time series of scissile bond SASA of B25-B26 and B26-B27 of the wild-type models 4EYD and 4EY1 were spaced with 500 ps instead of 250 ps as others. We have also recalculated the error bars as the standard deviation across the 5 different wild-type models. Note that the difference upon the change is negligible and does not influence any of our conclusions described in the paper.
- (2) We replotted the correlation plots of β -sheet propensity in Figure 6 because we found that the β -sheet propensity was actually calculated for residues B21 to B24, not B22 to B25, the residues we intended to analyze. In addition, the correlation coefficients originally annotated in the figure were Pearson correlation coefficients, not Kendall's tau correlation coefficients as intended. In the updated version, we have recalculated the metric for correct residues with correct correlation coefficients. This change does not influence our conclusion that the SASA metrics were only partially predictive since the Kendall's tau correlation coefficients were all below 0.5 as the Pearson correlation coefficients.
- (3) We recalculated the error bars of the SASA metrics as the standard deviation across the simulations from the five different wild-type models, as the uncertainties in the time series were underestimated. We have also modified relevant paragraphs in the result discussion accordingly. The change in the calculation of uncertainty increases the uncertainty of the computed metrics, but not enough to influence the conclusions of the paper in terms of which metrics are predictive. It additionally clarifies some of the unexpected changes between glycoforms by showing the differences are less than the uncertainty.