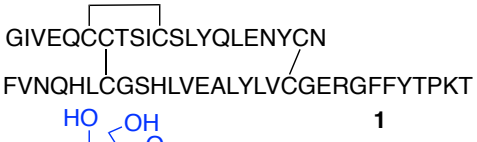
In Silico Evaluation of Insulin Analogs with Different Variations

The most important properties for developing oral insulin are the proteolytic stability, oligomerization propensity, and biological activity. An analog that is very resistant to protease digestion, does not form oligomers, and possesses the biological activity of wild-type insulin is ideal candidate we are aiming for. Previous studies have demonstrated that it is feasible to use computational approaches to study each of these properties. In the first year of funding period, we will use simulation approaches to study the proteolytic stability and oligomerization propensity of insulin **1**-**55**. The proteolytic stability will be studied using molecular dynamics (MD) simulations. The aggregation propensity will be studied using predictive algorithms based on empirical data and also MD simulations.

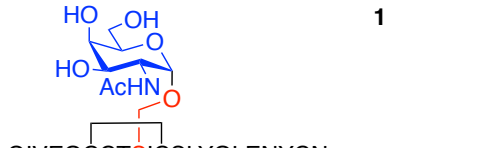
The effects of the engineering designs on the proteolytic stability will be investigated through MD simulations of the glycosylated and mutated insulin molecules in water. Similar methods have been used before with PEGylated insulin and other proteins, which will inform the design of our computational methods here. By using calculated measures of the molecule's dynamics, we will be able to predict how certain modifications will stabilize or destabilize the insulin molecule. Specifically, we will look at the root-mean squared deviation (RMSD), root-mean square fluctuation (RMSF), and radius of gyration values for each simulated molecule.

Protein aggregation will be examined through a combination of predictive algorithms and MD simulations. Much like proteolytic stability, the aggregation of proteins is intrinsically linked to flexibility, and several groups have taken advantage of this to predict aggregation propensity based on RMSF measures calculated during MD simulations. Solvent-accessible-surface (SAS) area has also been used to effectively predict aggregation propensity from simulated molecules. Predictive algorithms, based on experimentally-obtained, real-world aggregation propensities for many proteins, are also quite useful for our purpose here. In particular, the AGGRESCAN3D approach will be applied. This approach uses the 3D structure of the molecule as well as intrinsic aggregation values for each amino acid, which are based on empirical data, to predict aggregation-prone proteins.

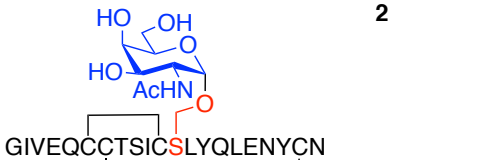
(A) Study insulin glyco-variants that have been characterized to optimize simulation approaches



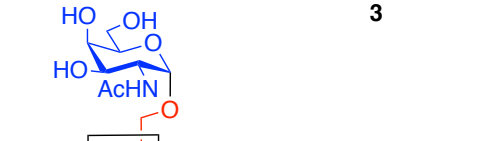
1



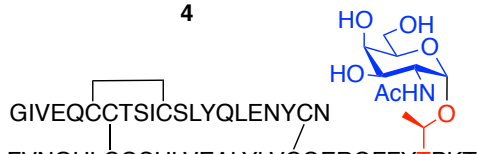
2



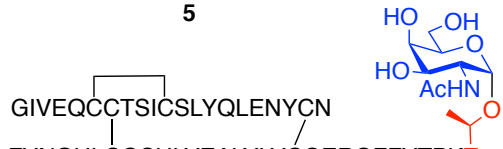
3



4



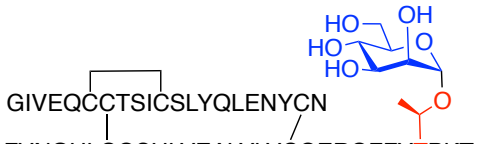
5



6

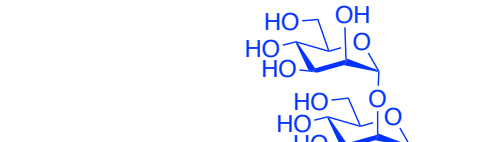
7

8

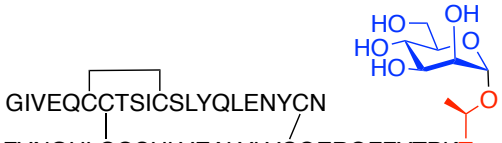


9

10



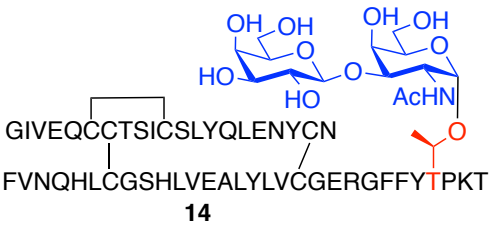
11



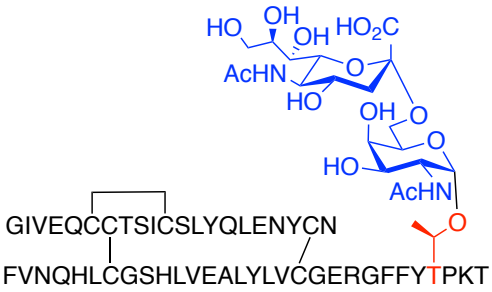
12

13

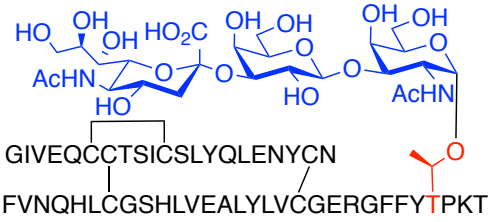
(B) Use optimized approaches to predict the effects of core-1 and -2 types of O-linked glycans



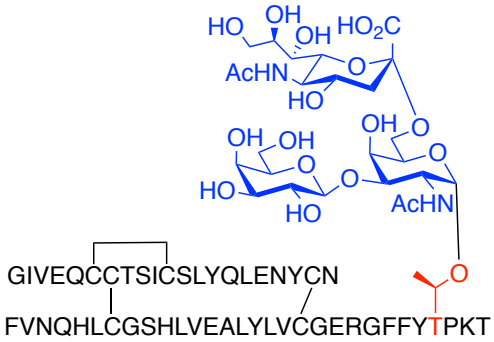
14



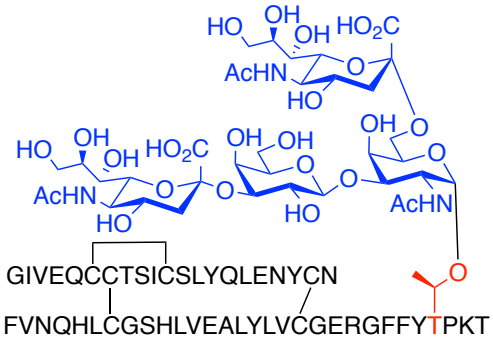
15



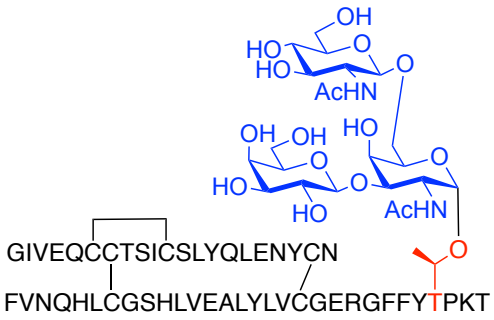
16



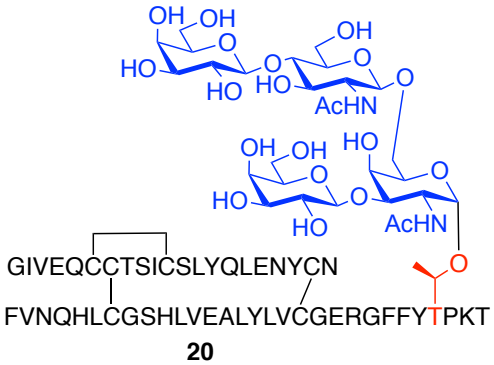
17



18

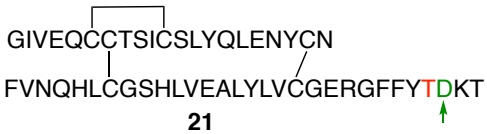


19

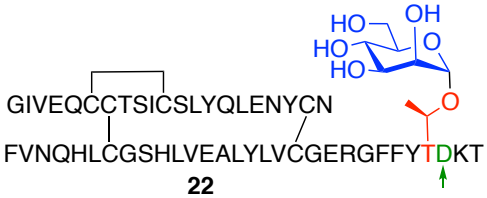


20

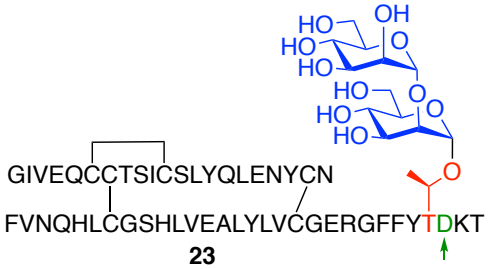
(C) Use optimized approaches to predict the effects of varied amino acid sequences



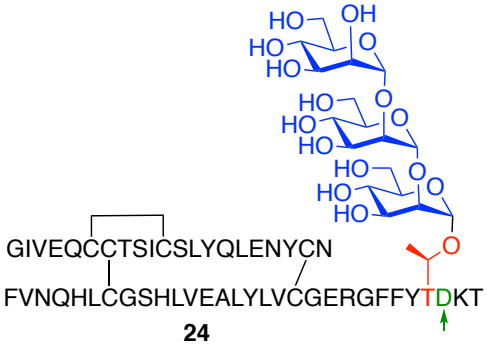
21



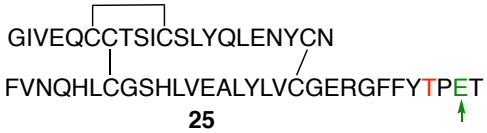
22



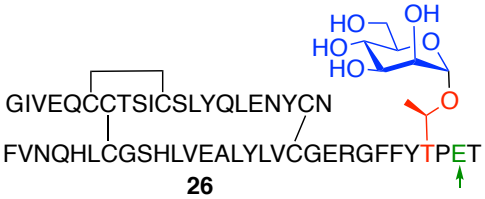
23



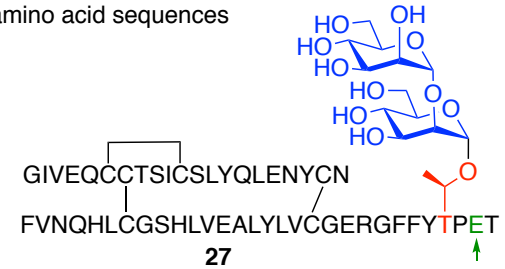
24



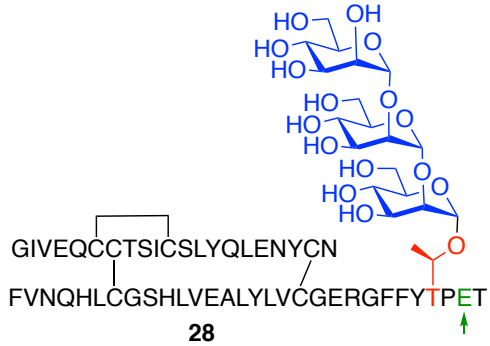
25



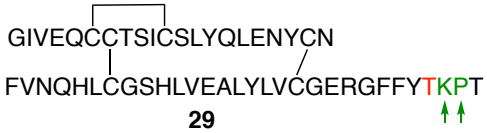
26



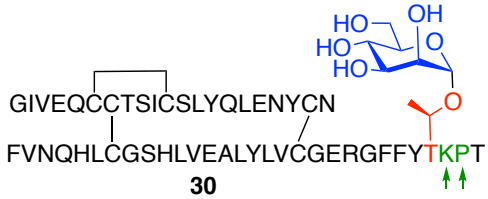
27



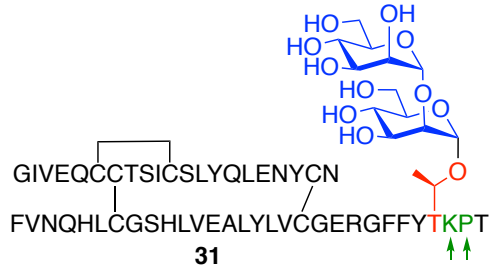
28



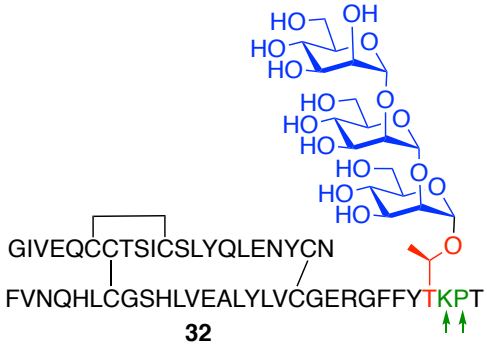
29



30

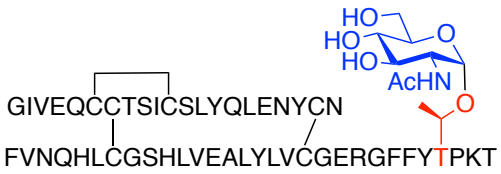


31

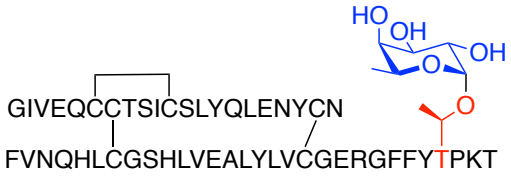


32

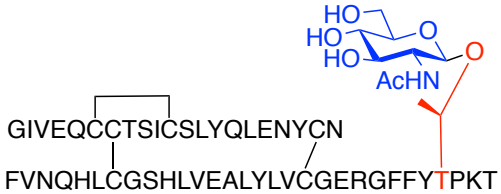
(D) Use optimized approaches to predict the effects of O-linked glycan cores



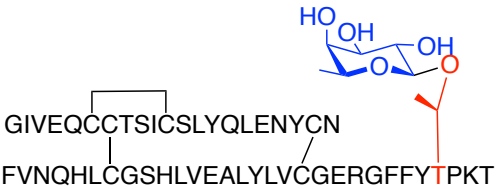
33



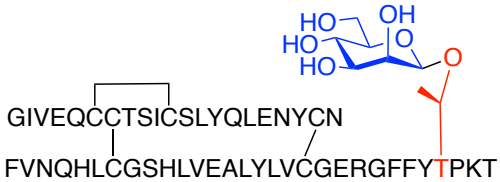
39



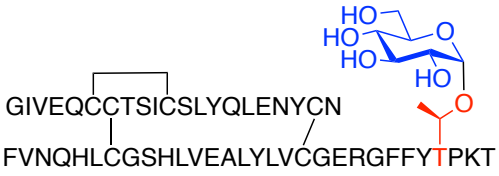
34



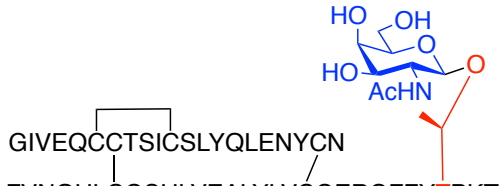
40



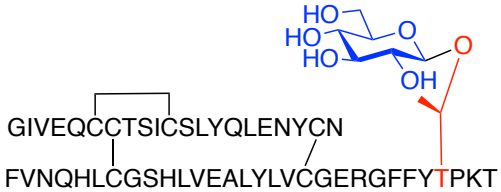
35



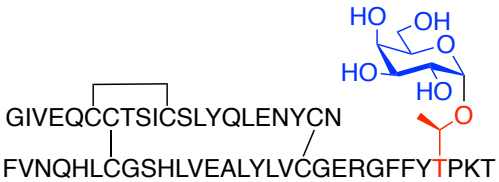
41



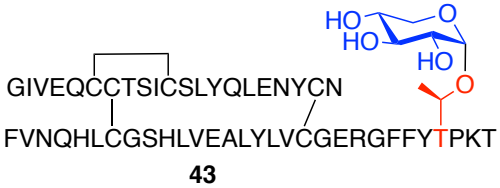
36



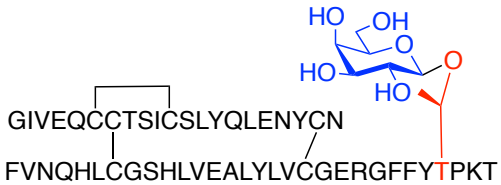
42



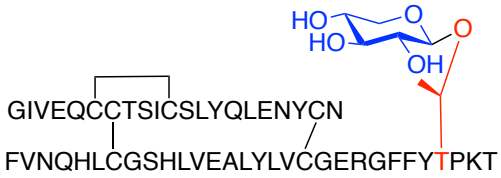
37



43

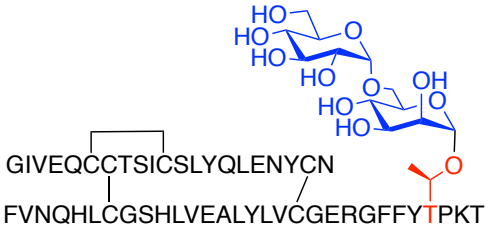


38

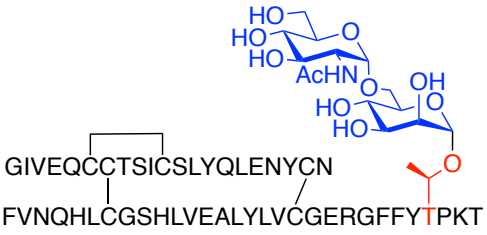


44

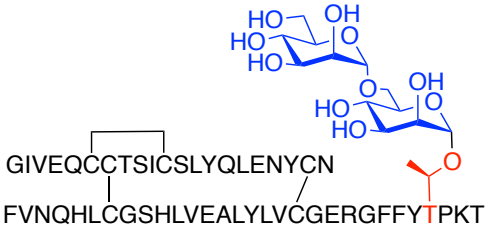
(E) Use optimized approaches to predict the effects of large O-linked glycans



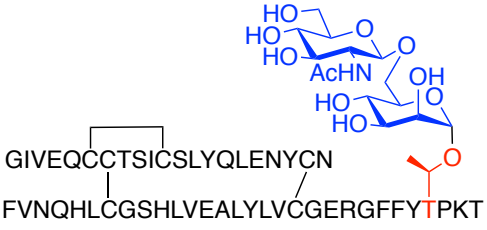
45



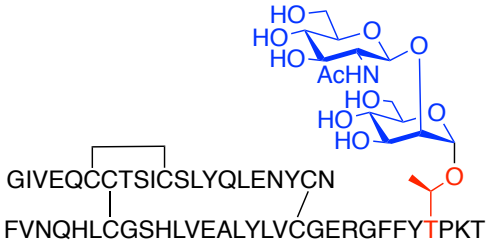
46



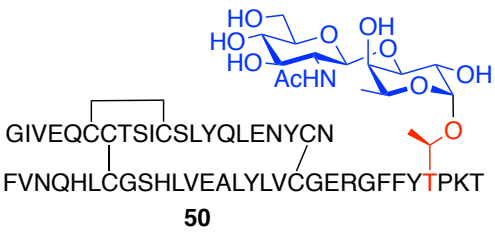
47



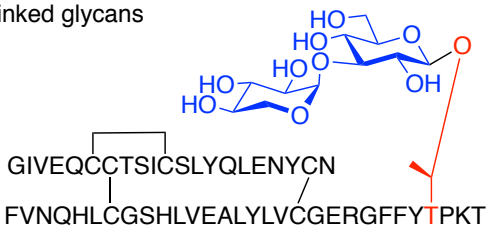
48



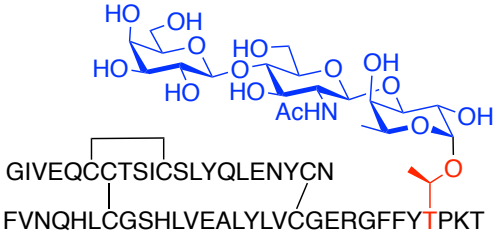
49



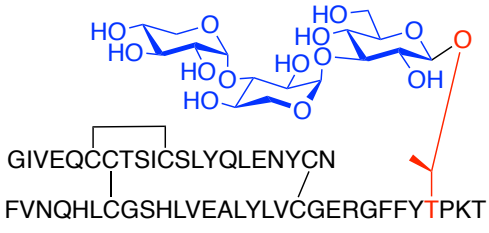
50



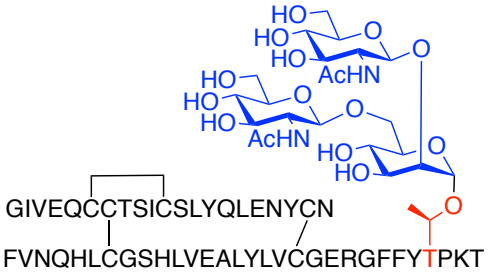
51



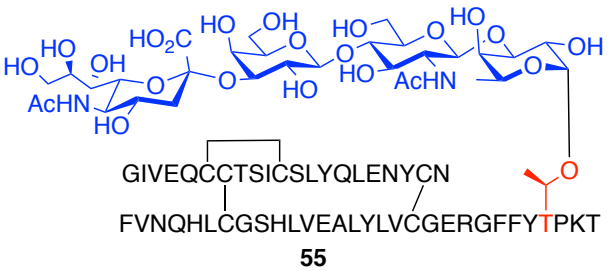
52



53



54



55

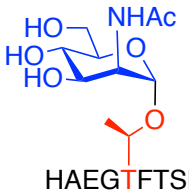
In Silico Evaluation of GLP-1 Analogs with Different Variations

In the first year of funding period, we will use simulation approaches to study the proteolytic stability of GLP-1 **1**-**42**.

(A) Study GLP-1 glyco-variants that have been characterized to optimize simulation approaches

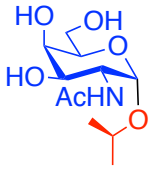
HAEGTFTSDVSSYLEGQAAKEFIAWLVKGRG

1



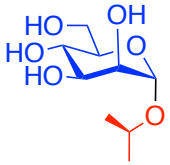
HAEGTFTSDVSSYLEGQAAKEFIAWLVKGRG

2



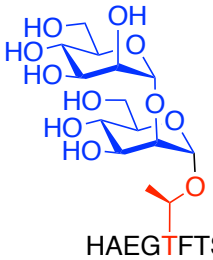
HAEGTFTSDVSSYLEGQAAKEFIAWLVKGRG

3



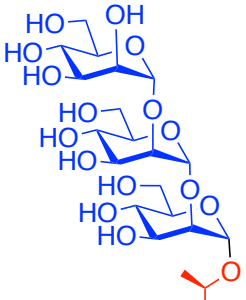
HAEGTFTSDVSSYLEGQAAKEFIAWLVKGRG

4



HAEGTFTSDVSSYLEGQAAKEFIAWLVKGRG

5



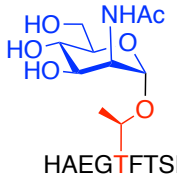
HAEGTFTSDVSSYLEGQAAKEFIAWLVKGRG

6

(B) Use optimized approaches to predict the effects of varied amino acid sequences

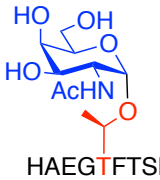
HAEGTFTSDVSSYLEGQAAKEFIAWLVKGR$_{amide}$

7



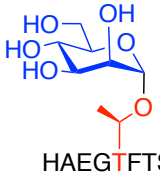
HAEGTFTSDVSSYLEGQAAKEFIAWLVKGR$_{amide}$

8



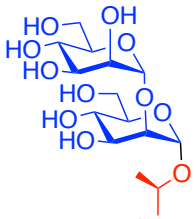
HAEGTFTSDVSSYLEGQAAKEFIAWLVKGR$_{amide}$

9



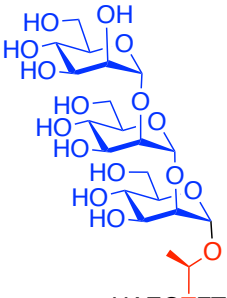
HAEGTFTSDVSSYLEGQAAKEFIAWLVKGR$_{amide}$

10



HAEGTFTSDVSSYLEGQAAKEFIAWLVKGR$_{amide}$

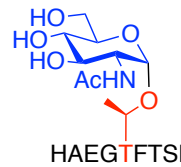
11



HAEGTFTSDVSSYLEGQAAKEFIAWLVKGR$_{amide}$

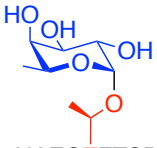
12

(C) Use optimized approaches to predict the effects of O-linked glycan cores



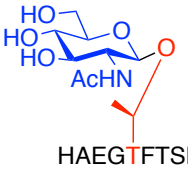
HAEGTFTSDVSSYLEGQAAKEFIAWLVKGRG

13



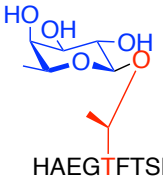
HAEGTFTSDVSSYLEGQAAKEFIAWLVKGRG

19



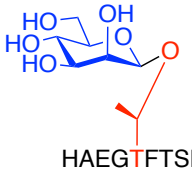
HAEGTFTSDVSSYLEGQAAKEFIAWLVKGRG

14



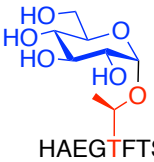
HAEGTFTSDVSSYLEGQAAKEFIAWLVKGRG

20



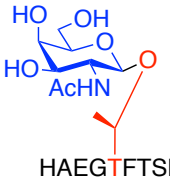
HAEGTFTSDVSSYLEGQAAKEFIAWLVKGRG

15



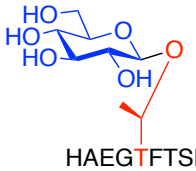
HAEGTFTSDVSSYLEGQAAKEFIAWLVKGRG

21



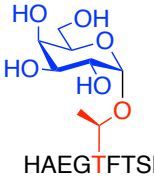
HAEGTFTSDVSSYLEGQAAKEFIAWLVKGRG

16



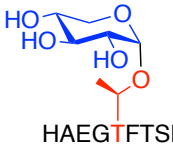
HAEGTFTSDVSSYLEGQAAKEFIAWLVKGRG

22



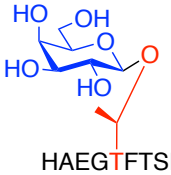
HAEGTFTSDVSSYLEGQAAKEFIAWLVKGRG

17



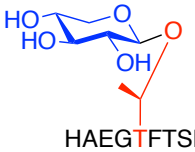
HAEGTFTSDVSSYLEGQAAKEFIAWLVKGRG

23



HAEGTFTSDVSSYLEGQAAKEFIAWLVKGRG

18

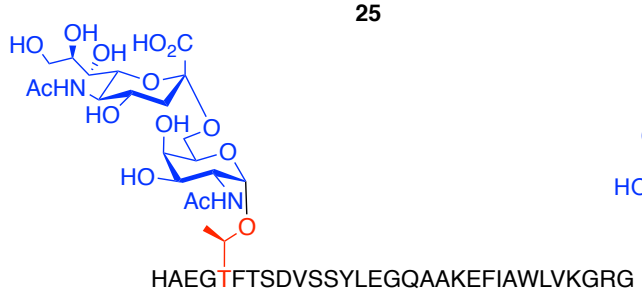


HAEGTFTSDVSSYLEGQAAKEFIAWLVKGRG

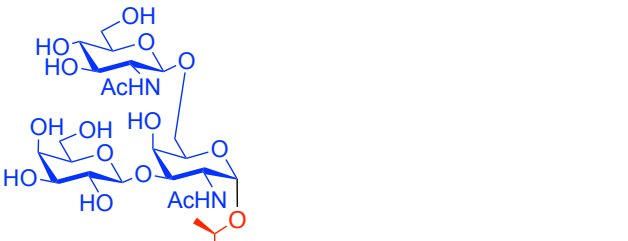
24

(D) Use optimized approaches to predict the effects of core-1 and -2 types of O-linked glycans

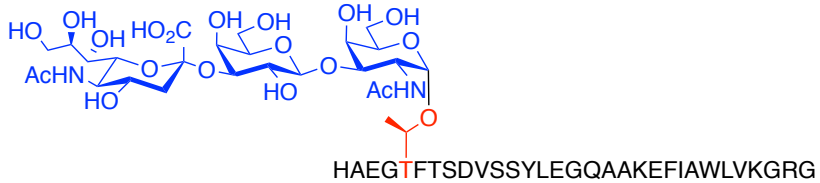
25



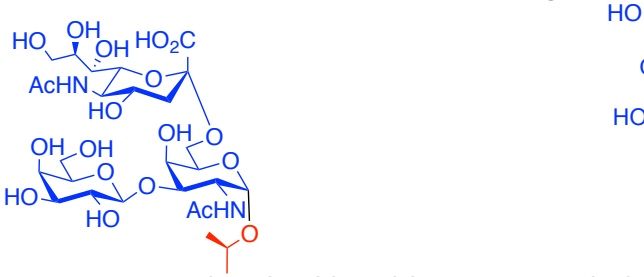
26



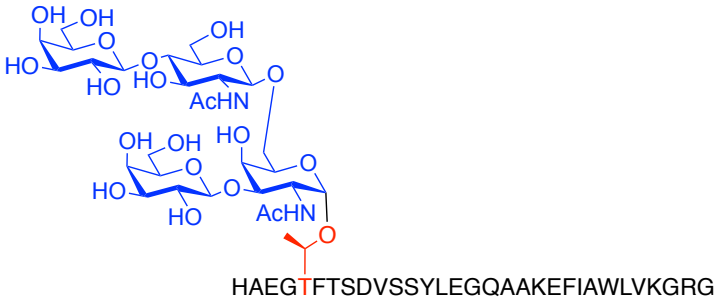
30



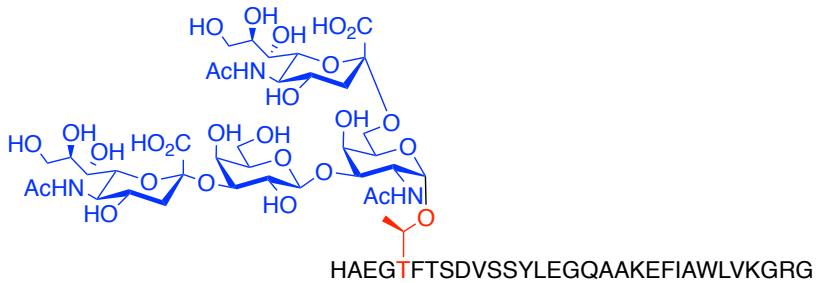
27



28

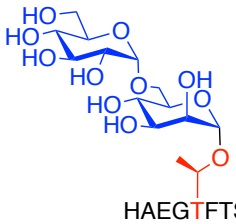


31



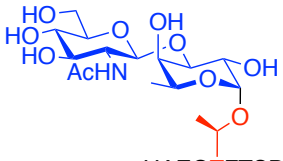
29

(E) Use optimized approaches to predict the effects of other large O-linked glycans



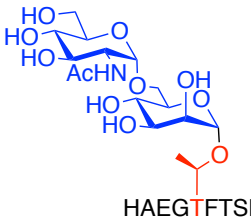
HAEGTFTSDVSSYLEGQAAKEFIAWLVKGRG

32



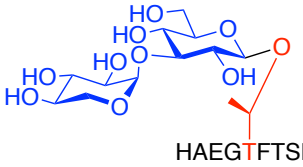
HAEGTFTSDVSSYLEGQAAKEFIAWLVKGRG

37



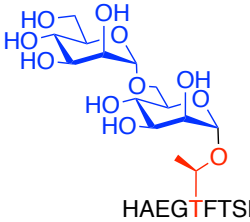
HAEGTFTSDVSSYLEGQAAKEFIAWLVKGRG

33



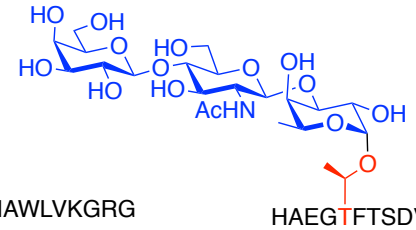
HAEGTFTSDVSSYLEGQAAKEFIAWLVKGRG

38



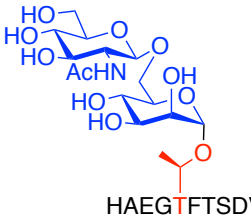
HAEGTFTSDVSSYLEGQAAKEFIAWLVKGRG

34



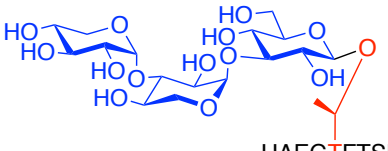
HAEGTFTSDVSSYLEGQAAKEFIAWLVKGRG

39



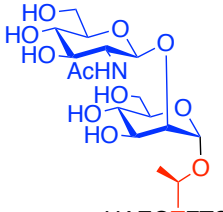
HAEGTFTSDVSSYLEGQAAKEFIAWLVKGRG

35



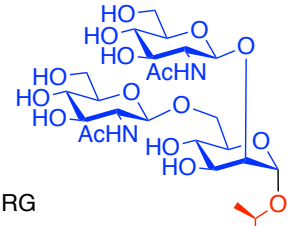
HAEGTFTSDVSSYLEGQAAKEFIAWLVKGRG

40



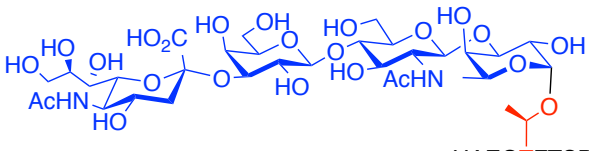
HAEGTFTSDVSSYLEGQAAKEFIAWLVKGRG

36



HAEGTFTSDVSSYLEGQAAKEFIAWLVKGRG

41



HAEGTFTSDVSSYLEGQAAKEFIAWLVKGRG

42