Supporting Information

GLYCOENGINEERING OF ESTERASE ACTIVITY THROUGH METABOLIC FLUX-BASED MODULATION OF SIALIC ACID

Mohit P. Mathew,[1,2] Elaine Tan,[1,2] Jason W. Labonte,[1,3] Shivam Shah,[2] Christopher T. Saeui,[2] Lingshu Liu,[2] Rahul Bhattacharya,[2] Patawut Bovonratwet,[2] Jeffrey J. Gray,[3] and Kevin J. Yarema[2,3,4]

[2]Department of Biomedical Engineering and the Translational Tissue Engineering Center

[3]Department of Chemical and Biochemical Engineering

The Johns Hopkins University, Baltimore, Maryland, USA

[1]These authors contributed equally to this work

[4]Corresponding author:

Translational Tissue Engineering Center
5029 Robert H. & Clarice Smith Building
The Johns Hopkins University
400 North Broadway
Baltimore, Maryland, 21231
USA

Email: kyarema1@jhu.edu
Phone: 410.614.6835
Fax: 410.614.6840

Prediction of *N*-glycan site occupancy using NetNGlyc 1.0

FASTA formatted amino acid sequences were attained from the NCBI database (http://www.ncbi.nlm.nih.gov). Sequences were attained for carboxylesterase 1 (Accession no. AAI10339), carboxylesterase 2 (Accession no. AAH32095), carboxylesterase 3 (Accession no. ACD11491), carboxylesterase 7 (Accession no. AAH69548), carboxylesterase 8 (Accession no. AAH64573), acetylcholinesterase (Accession no. AAA68151), butyrylcholinesterase (Accession no. AAH18141), carboxyl ester lipase (Accession no. AAA51973) and esterase D (Accession no. AAC99788). These sequences were entered into the N-glycosylation prediction tool: NetNGlyc (http://www.cbs.dtu.dk/services/NetNGlyc/) and this online software tool then was used to predict the sites and probabilities of *N*-glycosylation.

The predicted sites were based on the following criterion (as explained at the above-listed NetNGlyc weblink):

The 'potential' score is the averaged output of nine neural networks. Any potential crossing of the default threshold of 0.5 represents a predicted glycosylated site (as long as it occurs in the required sequon Asn-Xaa-Ser/Thr without proline at Xaa). The 'potential' score is the averaged output of nine neural networks. For further information, the jury agreement metric indicates how many of the nine networks support the prediction. The "Probability of N-Glycosylation" column shows one of the following outputs for predictions indicating glycosylated sites:

\+ Glycosylation potential > 0.5

++ Glycosylation potential > 0.5 AND Jury agreement (9/9) OR Potential>0.75

+++ Glycosylation potential > 0.75 AND Jury agreement

++++ Glycosylation potential > 0.90 AND Jury agreement

and non-glycosylated sites:

\- Potential < 0.5

-- Potential < 0.5 AND Jury agreement (all nine < 0.5)

--- Potential < 0.32 AND Jury agreement

For prediction of N-glycosylation sites with high specificity (asparagines the most likely to be glycosylated), the algorithms use only (++) predictions (and better) for asparagines that occur within the Asn-Xaa-Ser/Thr triplet (no proline at the Xaa position). Of note, identifying sites using this method potentially compromises sensitivity (i.e., it is possible that positively glycosylated sites will not be identified but this adds confidence to the sites that are predicted to be glycosylated).

Table 1. *In silico* prediction of esterase *N*-glycosylation

Name	Accession Number	Predicted Sites	Sites	Jury Agreement	Probability of N-Glycosylation
ACHE	AAA68151	3	296	5/9	+
			381	7/9	+
			495	6/9	+
BCHE	AAH18141	9	45	9/9	++
			85	8/9	+
			134	8/9	+
			269	9/9	+++
			284	9/9	++
			369	5/9	+
			483	9/9	+++
			509	9/9	++
			514	4/9	+
CEL	AAA51973	1	210	9/9	++
ESD	AAC99788	NONE	N/A	N/A	N/A
CES1	AAI10339	1	79	7/9	+
CES2	AAH32095	1	175	9/9	++
CES3	ACD11491	1	105	6/9	+
CES7	AAH69548	3	363	6/9	+
			463	9/9	++
			474	6/9	+
CES8	AAH64573	3	95	9/9	+++
			157	9/9	++
			269	9/9	+++