

Glycopeptide analysis of individual glycosylation sites of HIV envelope oligomers that induce breadth in neutralizing antibody responses

H Desaire, S Alam, EP Go, H Liao, J Irungu, Y Zhang, D Dalpathado, LL Sutherland, F Gao, BF Haynes.



DUKE UNIVERSITY SCHOOL OF MEDICINE

Duke Human Vaccine Institute

dhvi.duke.edu

Glycobiology Meets Immunology



Glycobiology of Env effects its immunogenicity.

- 1) Glycans shield immunogenic epitopes of Env.
Glycosylation in V2 can prevent binding of the MAb, 2909.**
- 2) Glycans are required for binding the neutralizing antibody, 2G12**
- 3) Glycans can dictate the conformation of the protein.**

Since glycosylation effects protein immunogenicity, glycobiology could provide novel approaches to improve immunogen design.

Our immediate Goal: Use glycobiology to understand key characteristic differences between good immunogens and poor immunogens

This goal is possible because:

- 1. The glycosylation profiles of different Env. Immunogens are different.**
- 2. Glycosylation profiles can correlate to *in vivo* structure of the protein.**
- 3. High mannose glycans indicate protected (structurally stable) sites on the protein.**

Hypotheses: A better immunogen may have a more stable, conserved 3D structure, evidenced by a more homogeneous glycan population and/or high-mannose glycans.

A better immunogen may have more unoccupied glycosylation sites.

Characteristics of the proteins we analyzed

JR-FL gp140 Δ CF

CON-S gp140 Δ CFI

A **wild-type** strain
(subtype B)

A **synthetic** strain
Centralized Consensus Gene Sequence
2001 LANL Group M Sequence Database

Proteins have **81% sequence identity**
Both proteins expressed in **293T cells**
Both proteins isolated as **oligomers.**

Poor Immunogen
Induced limited tier 1
Env. pseudovirus neutralization

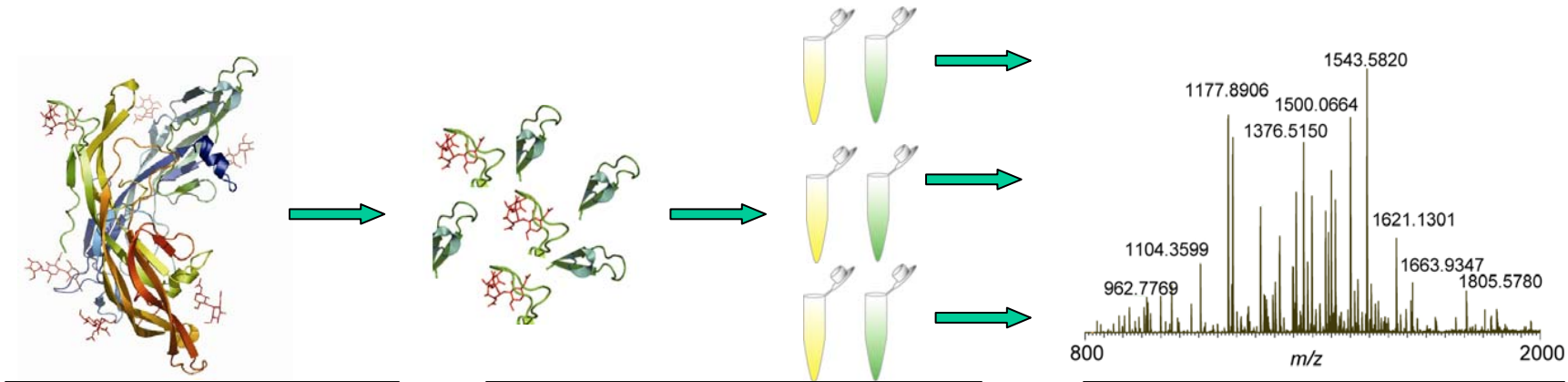
Better Immunogen
Elicited tier 1 and select tier 2
neutralizing antibodies for
multiple clades*

27 N-X-T/S sites
(potential glycosylation sites)

31 N-X-T/S sites

**Virol. 2006, 353, 268.*

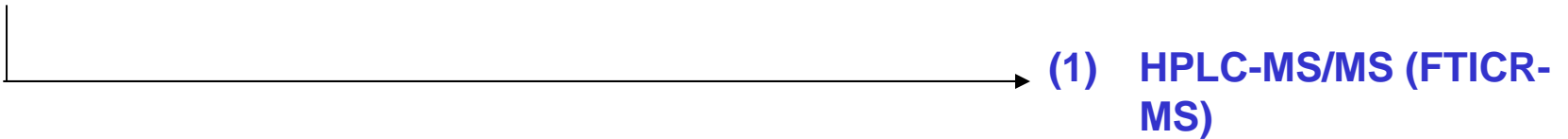
Sample preparation for glycopeptide analysis



- (1) Denature protein
- (2) Reduce/Alkylate
- (3) Trypsin digest

- (1) Fractionate on HPLC
(60 fractions)
- (2) Concentrate (x100)

- Acquire MS data for each fraction on:
- (1) MALDI TOF-TOF



Custom Data Analysis Tools

GlycoPep DB : A Glycopeptide Database

ABOUT GLYCOPEP DB Carbohydrate Database All Carbohydrate Entries

GLYCOPROTEINS Peptide Sequence QAHQNSR or Peptide Mass

TOOLS Cysteine Modification carbamidomethyl Variable Modification Pysin_SUMO3 or Pysin-Cox2C

HELP Charge State +1 Charge Carrier R

CONTACT INFO Mass Tolerance ± 100 ppm

PeakList

- 1028.817383
- 1032.811646
- 1035.599854
- 1037.689453
- 1039.749756
- 1043.772583
- 1045.752197
- 1046.720947
- 1053.774902
- 1055.801270
- 1057.809326

Submit Data Clear

GlycoPep ID

ABOUT GlycoPep ID Protein Sequence

HELP

CONTACT INFO

HOME

Enzyme Trypsin Charge +1 Mass Tolerance ± 100 ppm

Cysteine Modification carbamidomethyl Variable Modification Pysin_SUMO3 or Pysin-Cox2C

Fragment Ions

Specify the singly charged parent ion

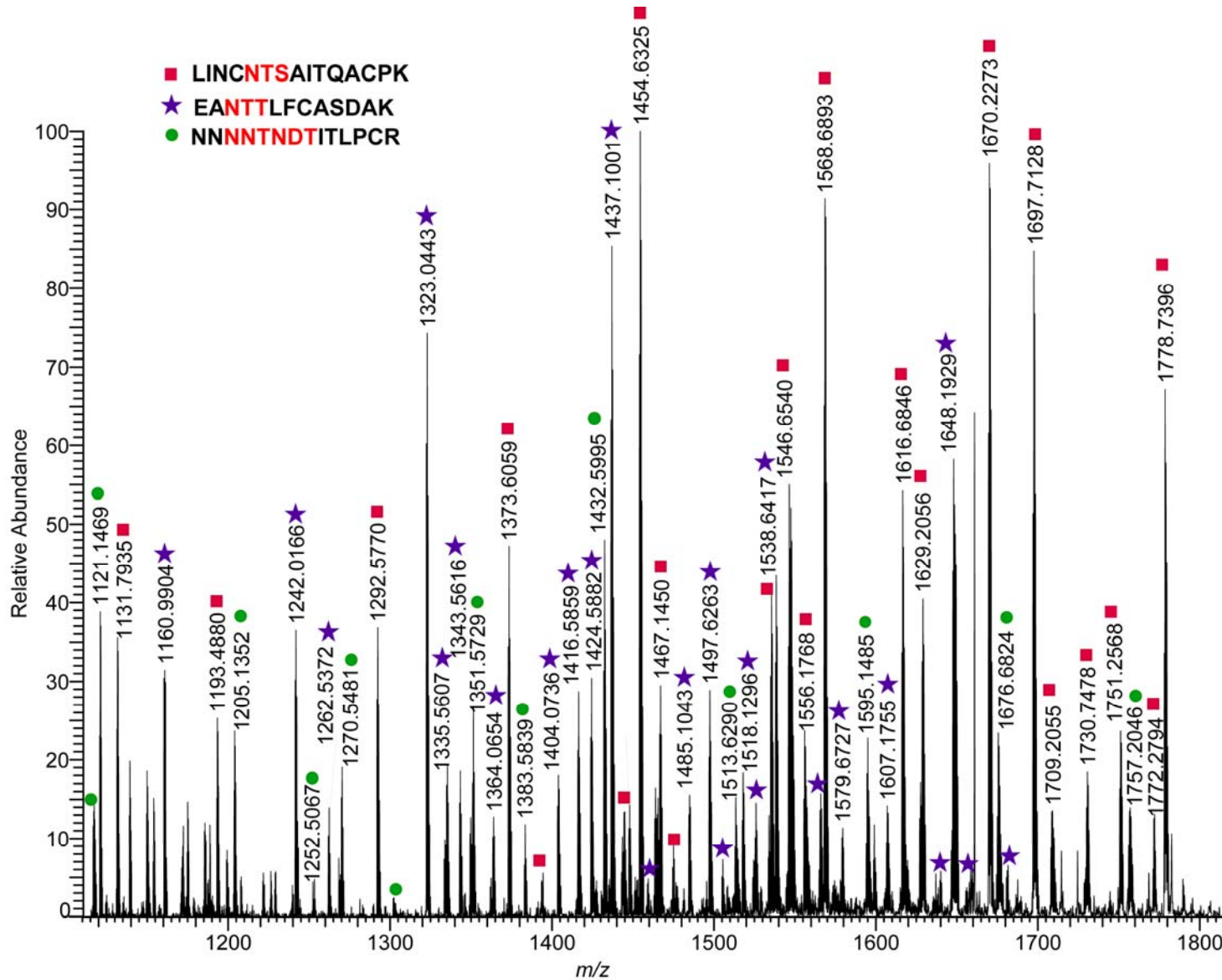
PeakList

- 1030.038940
- 1032.042333
- 1033.049194
- 1035.052879
- 1036.055187
- 1037.059907
- 1038.064023
- 1040.108040
- 1047.059847
- 1049.062457
- 1051.065071

Submit Data

Mass spectrometry produces extensive data for analysis

EX: 60 Different glycopeptides ID'd in one spectrum.



Example: Glycoforms Identified near the C-terminus of the gp120 V3 loop

Envelope Protein	Glycopeptide	Glycoforms	
JR-FL gp140ΔCFI	<p>ESVEINCTRPNNTR</p>	[Hex]11 [HexNAc]4 [Hex]12 [HexNAc]4 [Hex]13 [HexNAc]4 [Hex]14 [HexNAc]4 [Hex]16 [HexNAc]4 [Hex]17 [HexNAc]4 [Hex]18 [HexNAc]4 [Hex]6 [HexNAc]7[Fuc]2 [Hex]6 [HexNAc]8[Fuc]2 [Hex]6 [HexNAc]9[Fuc]2 [Hex]6 [HexNAc]10[Fuc]2 [Hex]7 [HexNAc]7[Fuc]2 [Hex]7 [HexNAc]8[Fuc]2 [Hex]8 [HexNAc]7[Fuc]2 [Hex]10 [HexNAc]6[Fuc]2 [Hex]10 [HexNAc]8[Fuc]2 [Hex]10 [HexNAc]9[Fuc]2 [Hex]11 [HexNAc]8[Fuc]2 [Hex]11 [HexNAc]10[Fuc]2 [Hex]12 [HexNAc]10[Fuc]2	[Hex]8 [HexNAc]6[Fuc]1 [Hex]8 [HexNAc]7[Fuc]1 [Hex]9 [HexNAc]7[Fuc]1 [Hex]9 [HexNAc]6[Fuc]1 [Hex]10 [HexNAc]5[Fuc]1 [Hex]10 [HexNAc]6[Fuc]1 [Hex]11 [HexNAc]5[Fuc]1 [Hex]11 [HexNAc]6[Fuc]1 [Hex]11 [HexNAc]7[Fuc]1 [Hex]11 [HexNAc]8[Fuc]1 [Hex]11 [HexNAc]9[Fuc]1 [Hex]11 [HexNAc]10[Fuc]1 [Hex]12 [HexNAc]5[Fuc]1 [Hex]12 [HexNAc]6[Fuc]1 [Hex]12 [HexNAc]7[Fuc]1 [Hex]13 [HexNAc]6[Fuc]1 [Hex]14 [HexNAc]6[Fuc]1 [Hex]14 [HexNAc]7[Fuc]1
CON-S gp140ΔCFI	<p>TIIVQNESVEINCTRPNNTR</p>	[Hex]5 [HexNAc]2 [Hex]6 [HexNAc]2 [Hex]7 [HexNAc]2 [Hex]9 [HexNAc]2 [Hex]8 [HexNAc]4 [Hex]9 [HexNAc]4 [Hex]10 [HexNAc]4 [Hex]11 [HexNAc]4 [Hex]12 [HexNAc]4 [Hex]13 [HexNAc]4 [Hex]14 [HexNAc]4 [Hex]16 [HexNAc]4 [Hex]17 [HexNAc]4 [Hex]18 [HexNAc]4 [Hex]8 [HexNAc]6 [Fuc]1 [Hex]8 [HexNAc]7 [Fuc]1 [Hex]9 [HexNAc]5 [Fuc]1 [Hex]9 [HexNAc]6 [Fuc]1 [Hex]12 [HexNAc]6 [Fuc]1 [Hex]12 [HexNAc]7 [Fuc]1	[Hex]4 [HexNAc]3 [Hex]5 [HexNAc]3 [Hex]3 [HexNAc]4[Fuc]1 [Hex]3 [HexNAc]5[Fuc]1 [Hex]5 [HexNAc]7 [Hex]6 [HexNAc]7 [Hex]7 [HexNAc]7 [Hex]8 [HexNAc]7 [Hex]8 [HexNAc]8 [Hex]9 [HexNAc]7 [Hex]9 [HexNAc]8 [Hex]10 [HexNAc]8 [Hex]11 [HexNAc]8 [Hex]12 [HexNAc]8 [Hex]13 [HexNAc]8

Coverage Statistics

CON-S: 100% coverage of 31 potential sites
 ~400 glycopeptide ions analyzed.

JR-FL: 100% coverage of 27 potential sites
 ~400 glycopeptide ions analyzed.

Hex = Mannose or Galactose

HexNAc = N-Acetylglucosamine

Fuc = Fucose

Variably Occupied and Unoccupied Glycosylation sites.

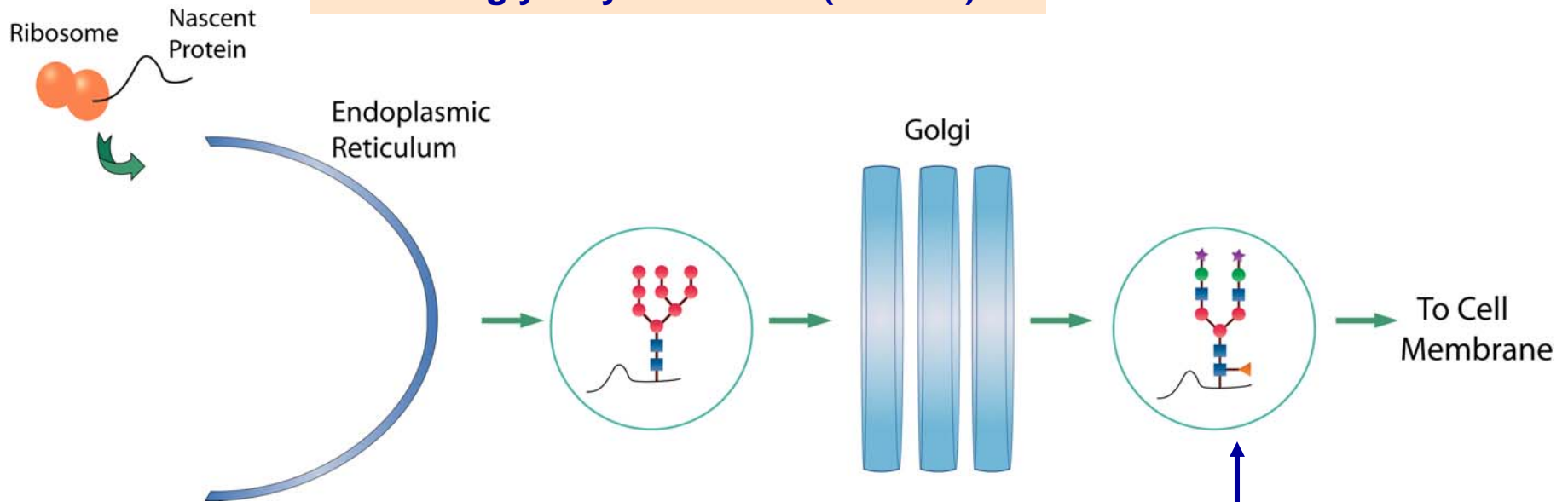
		signal peptide	70
CON-S gp140 Δ CFI		MRVIRGIQRNCQHLWRWGTLILGMLMICSAENLWVTVVYGVVWKEA NTT LFCASDAKAYDTEVHNWVAT	
JR-FL gp140 Δ CF		M-PMGSLQPLATLYLLGMLVASVL-----AVEKLWVTVVYGVVWKEATTTLFCASDAKAYDTEVHNWVAT	
CON-S gp140 Δ CFI		HACVPTDPNPQEIIVLE NVTE NFNMWKNNMVEQMHEDEIISLWDQSLKPCVKLTPLCVTL NCTNV - NVTNTT	140
JR-FL gp140 Δ CF		HACVPTDPNPQEVVLE NVTE HFNMWKNNMVEQMHEDEIISLWDQSLKPCVKLTPLCVTLNCKDV- NATNTT	
CON-S gp140 Δ CFI		NNTE -----EKGEIK NCSFNIT TEIRDKKQKVYALFYRLDVVPIDD--NN NSS NYRLIN NTS AITQACP	210
JR-FL gp140 Δ CF		NDSEG -TMERGEIK NCSFNIT TSIRDEVQKEYALFYKLDVVP-----DN NNTS YRLISCDTSVITQACP	
CON-S gp140 Δ CFI		KVSFEPIPIHYCAPAGFAILKCNDK KFNGT GPK NVST TVQCTHGIRPVVSTQLLL NGS LAEDEEIIIRSEN	280
JR-FL gp140 Δ CF		KISFEPIPIHYCAPAGFAILKCNDKTFNGKGPK NVST TVQCTHGIRPVVSTQLLL NGS LAEDEEVVIRSDN	
CON-S gp140 Δ CFI		IT NNAKTIIVQL NES VEIN CTRPN NNT RKSIIRI--GPGQAFYATGDIIGDIRQAHC NIS GTKW NKT LQOV	350
JR-FL gp140 Δ CF		FT NNAKTIIVQLKESVEIN CTRPN NNT RKSIHI--GPGRAFYTTEIGDIRQAHC NIS RAKW NDT LKQI	
CON-S gp140 Δ CFI		AKKLRHFN- NKT IIFKPSSGGDLEITTHSFNCRGEFFY NTS GLFN STWIGN --- GTKN - NNNT NDT LT	420
JR-FL gp140 Δ CF		VIKLREQFE- NKT IV FNHSS GGDPEIVMHSFNCGGEFFY NTS QLFN STWNNN --- TEGS NNT EGNTIT	
CON-S gp140 Δ CFI		LPCRIKQIINMWQGVGQAMYAPPIEGKITCKSNITGLLLTRDGG NNNT NETE IFRPGGGMRDNWRSELY	490
JR-FL gp140 Δ CF		LPCRIKQIINMWQEVGKAMYAPPIRGQIRCSSNITGLLLTRDGGINE- NGTE IFRPGGGMRDNWRSELY	
CON-S gp140 Δ CFI		KYKVVKIEPLGVAPTKAK-----L--TVQARQLLSGIVQQQSN	560
JR-FL gp140 Δ CF		KYKVVKIEPLGVAPTKAK-----T-----LTVQARLLLSGIVQQQNN	
CON-S gp140 Δ CFI		LLRAIEAQQHLLQLTVWGIKQLQARVLAVERYLKDQQL-----EIID	630
JR-FL gp140 Δ CF		LLRAIEAQQRMLQLTVWGIKQLQARVLAVERYLGDQQLLG IWGCSGKLIC TTAVPW NASWSNKS LDRIWN	
CON-S gp140 Δ CFI		NMT WMEWERE IN NYT DIIYS LIEESQNQQEKNEQELLALDKWASLWNWFDITNWLW-----	700
JR-FL gp140 Δ CF		NMT WMEWERE ID NYT SEIYTLIEESQNQQEKNEQELLELDKWASLWNWFDITKWLW-----	

Fully occupied sites: CON-S: 19/31

JR-FL: 20/27

Glycan Processing Depends on Protein Structure

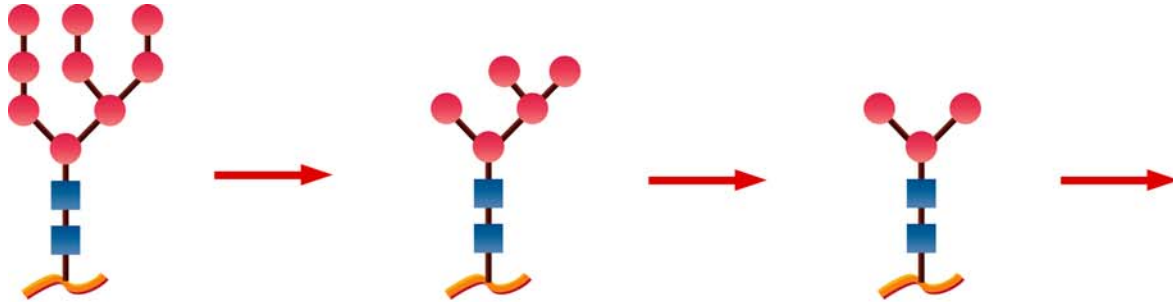
First, a carbohydrate precursor may be added at glycosylation sites (N-X-T/S)



Addition of carbohydrate requires: The N-(X)-T/S sequence, localized enzymes and cofactors, and substrate recognition.

The carbohydrate is modified *if* the correct enzymes and cofactors are localized and they have access to the carbohydrate.

Glycosylation processing



"Man 9"
High mannose

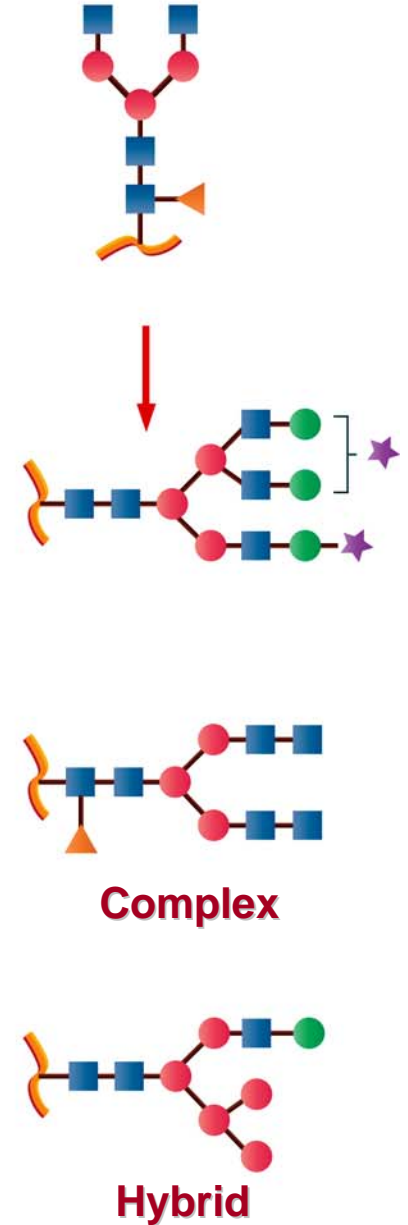
"Man 5"
High mannose

"Trimannose core"

High mannose glycans are present at occluded sites on the protein.

High mannose glycans stabilize protein structure.

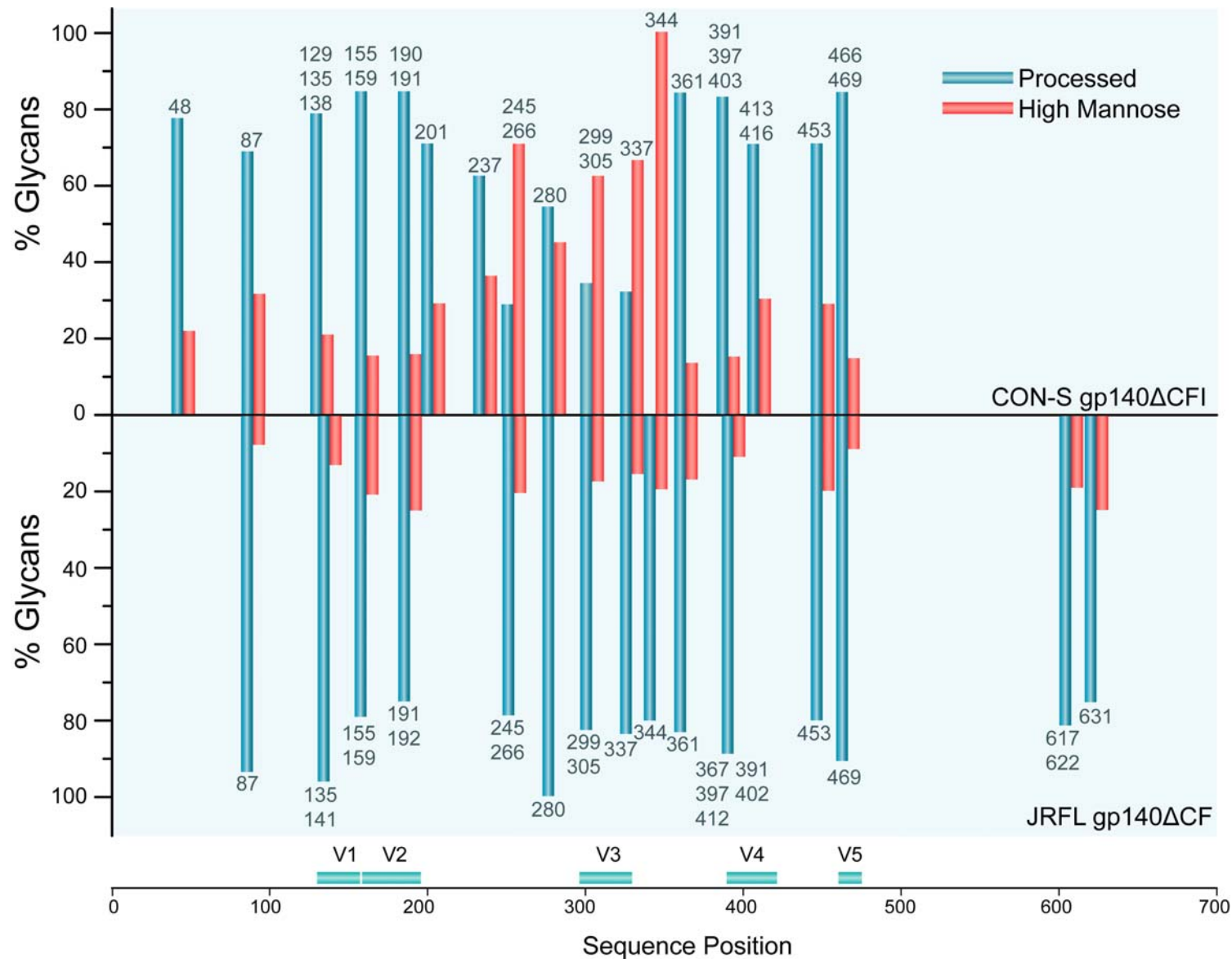
All "additions" take place after protein folding



Complex

Hybrid

Differences in glycosylation processing for CON-S and JR-FL



The glycan population at *the same site* in different sequence variants varies. CON-S has a higher proportion of unprocessed (high mannose) glycans.

Conclusions

Glycobiology:

- Changing the sequence of Env effects the glycosylation profile.
- Not all potential glycosylation sites on Env are utilized: CON-S and JR-FL had 12 and 7 sites respectively that were unoccupied.

Potential Correlations to Immunogenicity:

- CON-S's open glycosylation site in V3 may explain its enhanced ability to elicit neutralizing antibodies of greater breadth to V3.
- CON-S contained a higher proportion of high-mannose glycans which suggests a more stable 3D structure.

Future work:

- **Ultimate Goal: Characterize glycosylation on Env virions produced in relevant cell types (T cell or macrophage cell line) to identify the most biologically relevant glycan profile – then design strategies to mimic that profile in immunogenic experimental vaccines.**

Acknowledgements

NIH: RO1GM077226 (HD)

PO1AI61734 (BFH)

**Bill and Melinda Gates Foundation: Collaboration
For AIDS Vaccine Development (BFH)**