

Analysis of N-Glycosylation Sites in HIV glycoprotein 160

Rajendra Mandage, Anand Varsale, Sumeet Deshmukh, Nitesh Gamare
Dept of Bioinformatics, Deogiri College, Aurangabad, India

ABSTRACT

HIV infection is a condition caused by the human immunodeficiency virus. The condition gradually destroys the immune system, which makes it harder for the body to fight infections. HIV presents a complex knot for scientists to unravel. An envelope protein of the human HIV that is encoded by the env gene contains numerous glycosylation sites. It serves as a precursor for both the GP120 and the GP41. Here statistical investigation was done to study the sequential aspects of amino acids around the N-glycosylated protein from HIV virus. Sequences containing N-glycosylated asparagine were selected from the uniprot database of N-glycosylated proteins. The frequency of occurrence of amino acid residues around the glycosylated asparagine showed that there are increased numbers of isoleucine and threonine residues around the N-glycosylation sites in comparison with the nonglycosylated asparagine residues. Preferential occurrence of amino acid residues around the glycosylation site shows that T has the maximum preference around the N-glycosylation site. T at 3 and/or -3 positions strongly favors glycosylation irrespective of other glycosylation sites. The data presented in the present work clearly indicate that there is a pronounced positional preference for the hydrophobic and neutral amino acids at various positions around the N-glycosylation site. In future it will be of much interest to investigate further the possible structural and conformational implications of some of these suggested positional preferences of the various amino acids around the site of glycosylation. This is a potentially important study, and such analyses will surely contribute an important part of our knowledge base in the future on HIV research. These results will be of interest to molecular biologists and protein engineers to identify N-glycosylation sites important in molecular recognition processes in HIV virus.

INTRODUCTION

Protein N-glycosylation is an important co-translational modification process wherein short sugar chains are covalently attached to the amide group of asparagine (N) residue in the amino acid chain. N-glycosylation affects a number of properties of proteins such as solubility, stability and turnover, secretion, protease resistance, protein-protein interaction/recognition and immunogenicity, and hence has an immense biological importance.

HIV GP 160.

- Encoded by 'env' gene.
- Protein synthesized is modified by addition of complex carbohydrate (Glycan).
- Host enzyme attack GP 160 to be splitting into GP 120 and GP 41.

MATERIALS & METHODS

The HIV-1 envelope gp160 sequences were retrieved from Uniprot protein resource using keyword based search. These 200 glycosylated sites were selected for the present investigation. A sequence sub database was in excel sheet prepared by considering the central ASn residue as the glycosylated. With a selection of 5 amino acid residues on either side of it (i-5 to i+5; i is the position of glycosylated Asn). Thus, the window size of the amino acid segments selected for our analysis is 11. Position specific matrix was calculated for 11 residues and conserved positions were identified colored according to clustal X color scheme.

RESULTS

Frequency of amino acids in glycosylated and non-glycosylated site differ:

Residues	Frequency of amino acids around	
	Glycosylated sites	Non-glycosylated sites
Tyrosine	High	Moderate
Serine	High	Moderate
Threonine	High	Moderate
Proline	Moderate	High
Alanine	High	Moderate
Tryptophan	Moderate	High
Glycine	High	Moderate
Methionine	Moderate	high
Isoleucine	High	Moderate
Valine	high	Moderate
Glycine	High	moderate

POSITION SPECIFIC COLOUR MATRICES

	-5	-4	-3	-2	-1		1	2	3	4	5	
E						A S P A R A G I N E						
V												
L												
T												
N												
F												
C												
D												
G												
I												
P												
Y												
R												
K												
S												
Q												
H												
A												
M												
W												

Amino acid	Single letter code	Color	Properties
Arginine, Lysine	R, K	Red	Hydrophilic, Positive charge, Basic
Theonine, Asparagine, Serine, Glutamine	T, N, S, Q	Green	Hydrophilic, Neutral charge, Max.H-bond.
Valine, leucine, Phenalanine, Isoleucine	V, L, F, I, A, M, W	Blue	Hydrophobic, Neutral charge, Aliphatic (V, L, I).
Glutamic acid, Aspartic acid	E, D	Magenta	Acidic
Histidine, Tyrosine	H, Y	Cyan	Hydrophilic, Aromatic
Glycine	G	Orange	Hydrophobic, Neutral charge.
Cysteine	C	Pink	Hydrophobic, Max.H-bond

CONCLUSION

The data presented in the present work clearly indicate that there is a pronounced positional reference for the amino acids at various positions around the N-glycosylation site. Ile and Thr occur preferentially at many positions close to the site of glycosylation and, in particular, strongly favor N-glycosylation when it is in the -2 and/or -3 and 2 and/or 3 positions. Serine more frequently occurs at the 1 and 5 positions when the site of glycosylation is a Asn. In addition, around glycosylation sites, the other amino acids preferred favorably are Ala, Gly, Thr and Val. The data presented in the present work clearly indicate that there is a pronounced positional preference for the hydrophobic and neutral amino acids at various positions around the N-glycosylation site Some potential sequence motif such as I-T-T-L-T-N-S-T-T-N-S, E-I-V-S-N-N-V-T-T-E-N occur frequently in the data set.

REFERENCE

- Spiro RG. Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiol.* 2002; 12: 43R-56R.
- Dennis JW, Granovsky M, Warren CE. Protein glycosylation in development and disease. *Bioessays.* 1999; 21: 412-421.