



# Charge and Polarity Preferences for N-Glycosylation: A Genome-Wide *In Silico* Study and Its Implications Regarding Constitutive Proliferation and Adhesion of Carcinoma Cells

## OPEN ACCESS

Muhammad Ramzan Manwar Hussain<sup>1,2\*</sup>, Zeeshan Iqbal<sup>3,4\*</sup>, Wajahat M. Qazi<sup>5</sup> and Daniel C. Hoessli<sup>3,6\*</sup>

### Edited by:

Stephan Von Gunten,  
University of Bern, Switzerland

### Reviewed by:

Vered Padler-Karavani,  
Tel Aviv University, Israel  
Heinz Laubli,  
University of Basel, Switzerland

### \*Correspondence:

Muhammad Ramzan  
Manwar Hussain  
geniouschemist26@gmail;  
Zeeshan Iqbal  
zeeshaniqbal3506@yahoo.com,  
zeeshan@imsb.edu.pk;  
Daniel C. Hoessli  
danielhoessli@gmail.com

### Specialty section:

This article was submitted to  
Women's Cancer,  
a section of the journal  
Frontiers in Oncology

**Received:** 27 September 2017

**Accepted:** 29 January 2018

**Published:** 28 February 2018

### Citation:

Manwar Hussain MR, Iqbal Z,  
Qazi WM and Hoessli DC (2018)  
Charge and Polarity Preferences for  
N-Glycosylation: A Genome-Wide  
*In Silico* Study and Its Implications  
Regarding Constitutive Proliferation  
and Adhesion of Carcinoma Cells.  
*Front. Oncol.* 8:29.  
doi: 10.3389/fonc.2018.00029

<sup>1</sup>Key Laboratory of Genome Sciences & Information, Beijing Institute of Genomics (CAS), Chinese Academy of Sciences, Beijing, China, <sup>2</sup>University of Chinese Academy of Sciences, Beijing, China, <sup>3</sup>Institute of Molecular Sciences & Bioinformatics, Lahore, Pakistan, <sup>4</sup>Department of Physics, GC University Lahore, Lahore, Pakistan, <sup>5</sup>Center for Intelligent Machines and Robotics, Department of Computer Science, COMSATS Institute of Information Technology, Lahore, Pakistan, <sup>6</sup>Panjwani Center for Molecular Medicine and Drug Research, International Center for Chemical and Biological Sciences, University of Karachi, Karachi, Pakistan

The structural and functional diversity of the human proteome is mediated by N- and O-linked glycosylations that define the individual properties of extracellular and membrane-associated proteins. In this study, we utilized different computational tools to perform *in silico* based genome-wide mapping of 1,117 human proteins and unravel the contribution of both penultimate and vicinal amino acids for the asparagine-based, site-specific N-glycosylation. Our results correlate the non-canonical involvement of charge and polarity environment of classified amino acids (designated as L, O, A, P, and N groups) in the N-glycosylation process, as validated by NetNGlyc predictions, and 130 literature-reported human proteins. From our results, particular charge and polarity combinations of non-polar aliphatic, acidic, basic, and aromatic polar side chain environment of both penultimate and vicinal amino acids were found to promote the N-glycosylation process. However, the alteration in side-chain charge and polarity environment of genetic variants, particularly in the vicinity of Asn-containing epitope, may induce constitutive glycosylation (e.g., aberrant glycosylation at preferred and non-preferred sites) of membrane proteins causing constitutive proliferation and triggering epithelial-to-mesenchymal transition. The current genome-wide mapping of 1,117 proteins (2,909 asparagine residues) was used to explore charge- and polarity-based mechanistic constraints in N-glycosylation, and discuss alterations of the neoplastic phenotype that can be ascribed to N-glycosylation at preferred and non-preferred sites.

**Keywords:** N-glycosylation, cancer, human proteins, genome-wide mapping, charge and polarity, EGFR, cadherins, epithelial-to-mesenchymal transition

## INTRODUCTION

Glycosylation of proteins is a most complex form of co- and post-translational modifications introducing structural diversity to proteins in the form of *O*- and *N*-linked sugar moieties (1–8). The covalent addition of complex glycans to the amide side chain of asparagine (*N*-glycosylation) and hydroxyl groups of serine and threonine (*O*-glycosylation) generates a large number of glycoforms that are credited for the modulation of diverse cellular functions (4, 5, 9–11).

Proteins that undergo *N*-linked glycosylation are biosynthesized on membrane-associated ribosomes and their signal peptide is removed by a signal peptidase as they emerge into the lumen of the rough endoplasmic reticulum. In the endoplasmic reticulum (ER), the oligosaccharyl transferase (OT) mediates the co-translational transfer of a lipid-linked tetradecasaccharide (GlcNAc2-Man9-Glc3) from a dolichol phosphate to an asparagine included in a NXS/T sequon. The selective recognition by OT of the consensus sequence (NXS/T) has enabled investigation of the structural requirements for *N*-glycosylation. The rapid increase of substrate data for protein *N*-glycosylation has led to the development of different databases and prediction tools: dbPTMs, UniProt, NetNGlyc and MAPRes (Mining Association Patterns among preferred amino acid residues in the vicinity of amino acids targeted for post-translational modifications) (10, 12–15).

Human proteins including growth factors, growth factor receptors, cell-surface proteins and secretory proteins are among the substrates that are *N*-glycosylated to perform key biological functions (16–27). The statistical analysis of the sequence contexts for *N*-glycosylation (preferred and non-preferred motifs) is needed to explore the biological relationships between sequence, structure, and function of glycoproteins. MAPRes is a valuable tool to define the significantly preferred and non-preferred amino acids in the vicinity of a *N*-glycosylation site by resorting to the association rule mining technique (12, 28). The association pattern/rule is established between two or more frequently occurring entities that are in correlation. The new version of MAPRes has the capacity to analyze the sequence environment of the modified residues according to the biophysical and biochemical properties (polarity and charge) of the amino acids. NetNGlyc<sup>1</sup> is another important computational tool that predicts the *N*-glycosylation (*N*<sup>+</sup>) and non-*N*-glycosylation (*N*<sup>−</sup>) sites on the basis of potential score and consensus sequences within the target protein (29).

In this study, we have identified 2,909 *N*-glycosylated sites (*N*<sup>+</sup> sites) in 1,117 human proteins in which the majority (96.5%) of *N*<sup>+</sup> sites is followed by the canonical motif of NXS/TY. According to our MAPRes analysis for general protein sequence analyses, Val at +1, Ser/Thr at 2, Leu/Val at 3 and Leu at −5 positions were found significantly preferred residues to mediate the glycosylation of Asn residues in the human proteome. After classifying amino acids' charge and polarity according to properties of their side-chain R-groups, significant preference for *N*-glycosylation was found for non-polar, uncharged R-groups

(Leu/Val/Gly/Ala/Ile: O) at position 1, polar R-group (Met/Thr/Ser/Cys/Asn/Gln: L) at position 2, polar, negatively charged acidic R-groups (Asp, Glu: N) at position 3/5/−4 and aromatic amino acids: Phe/Trp/Tyr: A, at position 3/−5/−1. Furthermore, we validated the MAPRes-predicted preferred association pattern for the wider *N*-glycosylation sequence contexts by using the NetNGlyc 1.0 server and 130 literature-reported UniProt proteins, and provided further evidence that charge and polarity of O amino acids (Gly/Ala/Val/Leu/Ile) at position 1, A amino acids (Phe/Trp/Tyr) at positions −6, −5, −2, −1, 1, 3, and 10, P-amino acids (Lys/Arg/His) at positions −9, −3, 9, 10, and N amino acids (Asp/Glu) at positions −4/3/5, in combination with L amino acids at position 2, is likely to generate significantly preferable environments for the *N*-glycosylation of human proteins. Any change in charge and polarity environments may, therefore, result in aberrant *N*-glycosylation on both preferred and non-preferred sites.

## MATERIALS AND METHODS

As a first step, preferred amino acids and association patterns were found around *N*-glycosylated (*N*<sup>+</sup>) and non-*N*-glycosylated (*N*<sup>−</sup>) residues by using MAPRes. Then, preferred amino acids and association patterns were determined on the basis of polarity and charge of the surrounding amino acids of *N*<sup>+</sup> and *N*<sup>−</sup> residues. Next, association patterns mined by MAPRes were validated with the NetNGlyc 1.0 server (see text footnote 1) for 15 biologically important proteins. The second level of validation was performed by using 130 literature-reported human proteins with 438 *N*-glycosylated motifs from the UniProt database.<sup>2</sup>

### Data Assembly

The primary sequences for *N*-glycosylated proteins were downloaded from dbPTM, version 3.0,<sup>3</sup> which is a database containing information about phosphorylation and glycosylation. The downloaded data for *N*-glycosylation included 16,915 entries in total and provided all required information for MAPRes analysis except for the protein primary sequences (from UniProt), which were added manually in the final dataset. All entries other than human proteins were removed. In the final dataset, only 1,117 proteins and 2,909 asparagine residues were found modified.

### Data Cleaning

The data cleaning of the final dataset was essential for estimating modified residues on vicinal amino acids. Several types of errors that can be generated during the compilation of the dataset such as incorrect position of the modified residues, presence of non-standard characters in the sequence, incorrect sequence length and repetitions of entries. These inconsistencies were identified and removed by utilizing the Data Inconsistency and Duplication Check module of MAPRes.

<sup>1</sup><http://www.cbs.dtu.dk/services/NetNGlyc/>.

<sup>2</sup><http://www.uniprot.org/>.

<sup>3</sup><http://dbptm.mbc.nctu.edu.tw/index.php>.

## Dataset

To the environments of the N+ (positive) and N- (negative), the negative sites were annotated in the final dataset. There were in total 31,052 negative sites (non-modified Asn) in the final dataset of human *N*-glycosylation proteins (Table 1). The ratio of the negative to positive sites is a very high number which could bias the results. To balance the number of negative and positive sites, a computational module was utilized, which selects the random entries in any dataset and can balance data one to one (1:1) or one to two (1:2) positive to negative sites. In this study, the positive to negative sites were analyzed on a 1 to 1 basis.

## Classification of the Amino Acids

The new version of MAPRes can mine association patterns for neighboring amino acids of modified residues on the basis of their specific biophysical and biochemical properties. In this study, the association patterns mined on the basis of polarity and charge of the amino acids were distributed into five different groups (Table 2). Specific letters were used to substitute the symbols of standard amino acids such that positively charged R-group amino acids were replaced with P, negatively charged R-group with N, polar, uncharged R-group with L, non-polar, aliphatic R-group with O, and aromatic R-group amino acids with A (Table 2).

## Preference Estimation and Association Rules Mining

After the classification of the amino acids according to the polarity and charge, the whole dataset was further divided into *N*-glycosylated and non-*N*-glycosylated datasets. MAPRes supplied the charge and polarity information to estimate the

significantly preferred amino acids and association rule mining for neighboring amino acids for all datasets (Table 3). In the first step, MAPRes generated 21 amino acid long peptides (Asn at 0 position and 10 amino acids on each side) from the protein dataset and estimated their preference (preferred amino acids corresponding to N+ and N- residues). Next, association rules/patterns were estimated by utilizing the preferred amino acids and their correlation with the modified residues. Association analyses were carried out at different possible support values for all datasets. There are four datasets in total, two for primary sequence of proteins (one is *N*-glycosylated “N+” and other is non-*N*-glycosylated “N-”) and two for classified/encoded sequence of proteins (*N*-glycosylated and non-*N*-glycosylated) for which MAPRes mined association rules (Table 4).

## Validity of Association Rules

The validity of the association patterns/rules were performed by the NetNGlyc 1.0 Server for the *N*-glycosylation and the 130 UniProt-reviewed human proteins comprising 438 *N*-glycosylated motifs (Table 5). At the first level of validation, the FASTA sequences of 15 biologically important (randomly selected) proteins were retrieved from UniProt. The modification potential for different residues was found by utilizing the NetNGlyc 1.0 Server (see text footnote 1). The 21 amino-acid long peptide assigned Asn at positions 0 and 10 amino acids on each side in the datasets of both predicted and non-predicted sites. The association patterns mined by MAPRes for N+ and N- were searched in the predicted and non-predicted peptide datasets. The percentage of peptides consistent with the association patterns was calculated. The same procedure was applied to the association rules found on

TABLE 1 | Statistics of assembled data and MAPRes output.

Observations	Glycosylated		Non-Glycosylated	
	General protein dataset	Encoded protein dataset	General protein dataset	Encoded protein dataset
Total number of proteins			1,117	
Number of targeted sites		2,909	31,052	
Number of significantly preferred sites	43	16	32	8
Number of association patterns	23	39	18	17
Number of unique association patterns	8	24	17	9
Support level (%)	5–55	5–95	5–10	5–30

TABLE 2 | Classification of for charge and polarity specific analysis of amino acids.

Chemical nature of side chain	Amino acids	Abbreviation
Positively charged R-groups	K, R, H	P
Negatively charged R-groups	D, E	N
Polar, uncharged R-groups	S, T, C, M, N, Q	L
Non-polar, aliphatic R-groups	G, A, V, L, I	O
Aromatic R-groups	F, Y, W	A

TABLE 3 | Significantly preferred sites around *N*-Glycosylated and non-*N*-Glycosylated residues for general protein sequence analyses.

Amino Acids	<i>N</i> -Glycosylated residues	Non- <i>N</i> -Glycosylated residues
<b>Significantly preferred positions for general protein dataset</b>		
C	-10, 1, 7	-2, 7, 9
D	-6, 5, 9	
E	-4, 3, 5	-7, -1
F	-5, -2, -1, 10	10
G	1	
H	-3, -2	
I	1, 4, 6	-4, -3, -2, 1, 2, 4, 5, 7
K	-3, 9, 10	4, 8
L	-5, 3	-3
N	-10, 8, 9, 10	-10, -7, -6, -1, 1, 7, 8
Q	-9, -8, -1, 10	-1
R	-7	-1
S	1, 2	
T	2	-5, 2
V	1, 3, 5	
W	3,	
Y	-2, -1, 1	-6, -5, -1, 3
<b>Significantly preferred positions for encoded protein dataset</b>		
A	-6, -5, -2, -1, 1, 3, 10	-6, -3, -2, 3, 10
L	2	-7, 2
N	-4, 3, 5	-
O	1	-
P	-9, -3, 9, 10	8

**TABLE 4** | Association rules/patterns mined by MAPRes for N-Glycosylated residues for general and classified protein sequences.

Association patterns	Confidence level (%)	Support level (%)
<b>Rules for general protein dataset</b>		
<L,-5><T,2>	100	5
<T,2><L,3>	100	5
<V,1><T,2>	100	5
<T,2><V,3>	100	5
<L,-5>	100	10
<L,3>	100	10
<S,2>	100	10, 15, 20, 25, 30, 35, 40
<T,2>	85.87	10, 15, 20, 25, 30, 35, 40, 45, 50, 55
<b>Rules for encoded protein dataset</b>		
<A,-5><O,1><L,2>	100	5
<A,-2><O,1><L,2>	100	5
<N,-4><O,1><L,2>	100	5
<O,1><L,2> <N,3>	100	5
<O,1><L,2><N,5>	100	5
<P,-9><O,1><L,2>	100	5
<P,-3><O,1><L,2>	100	5
<O,1><L,2><P,9>	100	5
<O,1><L,2><P,10>	100	5
<A,-5><L,2>	100	10
<A,-2><L,2>	77.25	10
<A,-1><L,2>	100	10
<L,2><A,3>	74.88	10
<L,2><A,10>	77.31	10
<N,-4><L,2>	100	10
<L,2><N,3>	100	10
<L,2> <N,5>	100	10
<P,-9><L,2>	100	10
<P,-3><L,2>	100	10
<L,2><P,9>	100	10
<L,2> <P,10>	100	10
<O,1><L,2>	100	15, 20, 25, 30, 35, 40
<L,2>	76.12	45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95

**TABLE 5** | Validation of association patterns mined by MAPRes with the NetNGlyc prediction model.

	Predicted	Non-Predicted
Total no of residues (Asn)	229	642
Number of peptides in which rules were found	216	378
Number of encoded peptides in which rules were found	195	397
Percentage of conformity with patterns mined by MAPRes (%)	94	59
Percentage of conformity with patterns mined by MAPRes for encoded dataset (%)	85	62

the basis of the classification of the amino acids. At this step, the predicted and non-predicted peptide datasets were encoded according to the previously defined classification of the amino acids. In the second validation step, 438 N-glycosylated motifs from 130 literature-reported proteins were analyzed and the percentage of N-glycosylated motifs validating the MAPRes association patterns was calculated. The protein modeling was performed by using Iterative Threading ASSEMBLY Refinement (I-TASSER) (30).

## RESULTS

### Rules Mined by MAPRes for N+ Sites

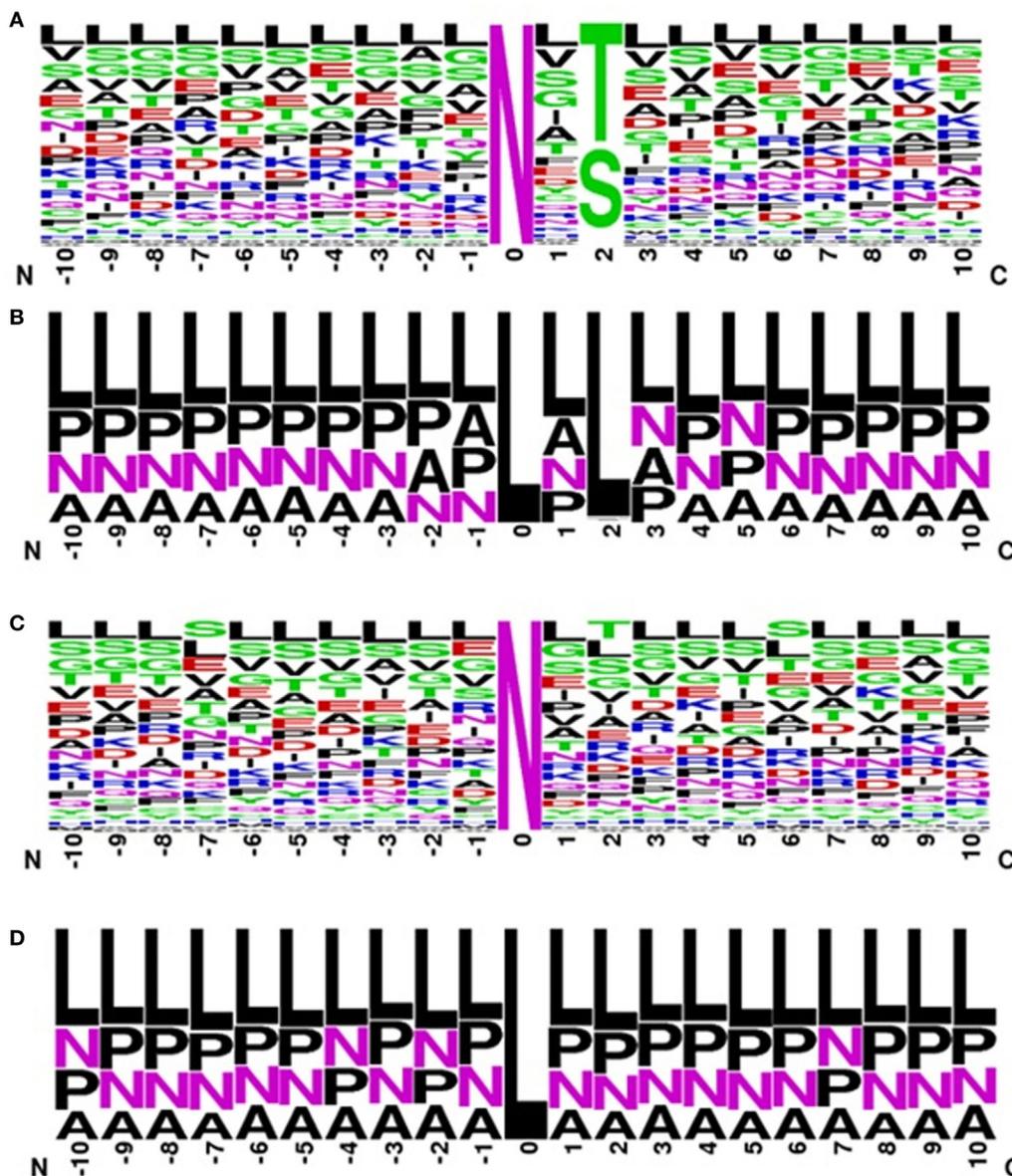
MAPRes identified 43 significantly preferred amino acid positions around N-glycosylated residues (Table 1). The Phe, Asn, and Gln were found preferred at four different positions and Asp, Gly, Ile, Lys, Val, and Tyr at three positions (Table 3). MAPRes mined association patterns on the basis of these significantly preferred amino acids allowed to develop a correlation between significantly preferred amino acids and modified residues. MAPRes suggested 23 association patterns that were the sum of all patterns mined at different support levels. After removal of repeated patterns/rules found at different support levels, only 8 distinctive rules for N-glycosylated residues were retained. From those patterns, Thr and Ser at position 2 (Figure 1A) were found at the highest support levels. MAPRes also found some other residues about which a correlation could be developed with N-glycosylated residues with Leu at positions -5/3 and Val at positions 1 and 3. Seven out of eight association patterns mined at 100% confidence level (Table 4).

In case of classified amino acids, MAPRes found 16 significantly preferred sites for N-glycosylated residues and 24 unique association patterns. The A amino acids were found significantly preferred at 7 different positions. Similarly, L, O, A, P, and N amino acids were also found preferred at several positions (Table 3). The range of confidence levels was 74.88–100%. The range of support levels was from 5 to 95%. The L amino acids at position 2 (Figure 1B) were found at the highest support level (95%). Another rule found at multiple support level was <O,1><L,2> (Figure 1B) and remaining association patterns were mined at only 5 and 10% support levels (Table 4).

### Rules Mined by MAPRes for N-Sites

MAPRes also mined association rules for N-residues at different support levels. There were 32 significantly preferred sites identified by MAPRes for N-sites (Table 2). Moreover, 18 association patterns in total were mined by MAPRes for N-sites and only one pattern was found at multiple support levels. Lys at position -3 (Figure 1C) was the only residue which was mined at 10% support level. The Ile and Asn were found significantly preferred at 8 and 7 different positions, respectively, and all of these positions were also correlated with N-sites. Tyr was also found preferred at four different positions but none of these developed a correlation with N-residues. The range of confidence level was 14.13–100% for N-residues. The <T,2> (Thr at position +2) had the lowest confidence level (14.13%).

The results concerning the significantly preferred positions and association patterns on the basis of polarity and charge of the surrounding amino acids indicated that the A amino acids were highly preferred at five positions (Table 3). L amino acids were also preferred at two positions and all of these preferred residues were found in association patterns for N-sites. The two association patterns <L,-7> and <L,2> (Figure 1D) were found at multiple support levels. The range of the confidence levels was 23.88–100% for the association patterns mined by MAPRes for classified dataset of non-N-glycosylation.



**FIGURE 1** | Frequency logo for sequence environment of N + residues for (A) general dataset, (B) classified dataset and of N- residues for (C) general dataset (D) classified dataset.

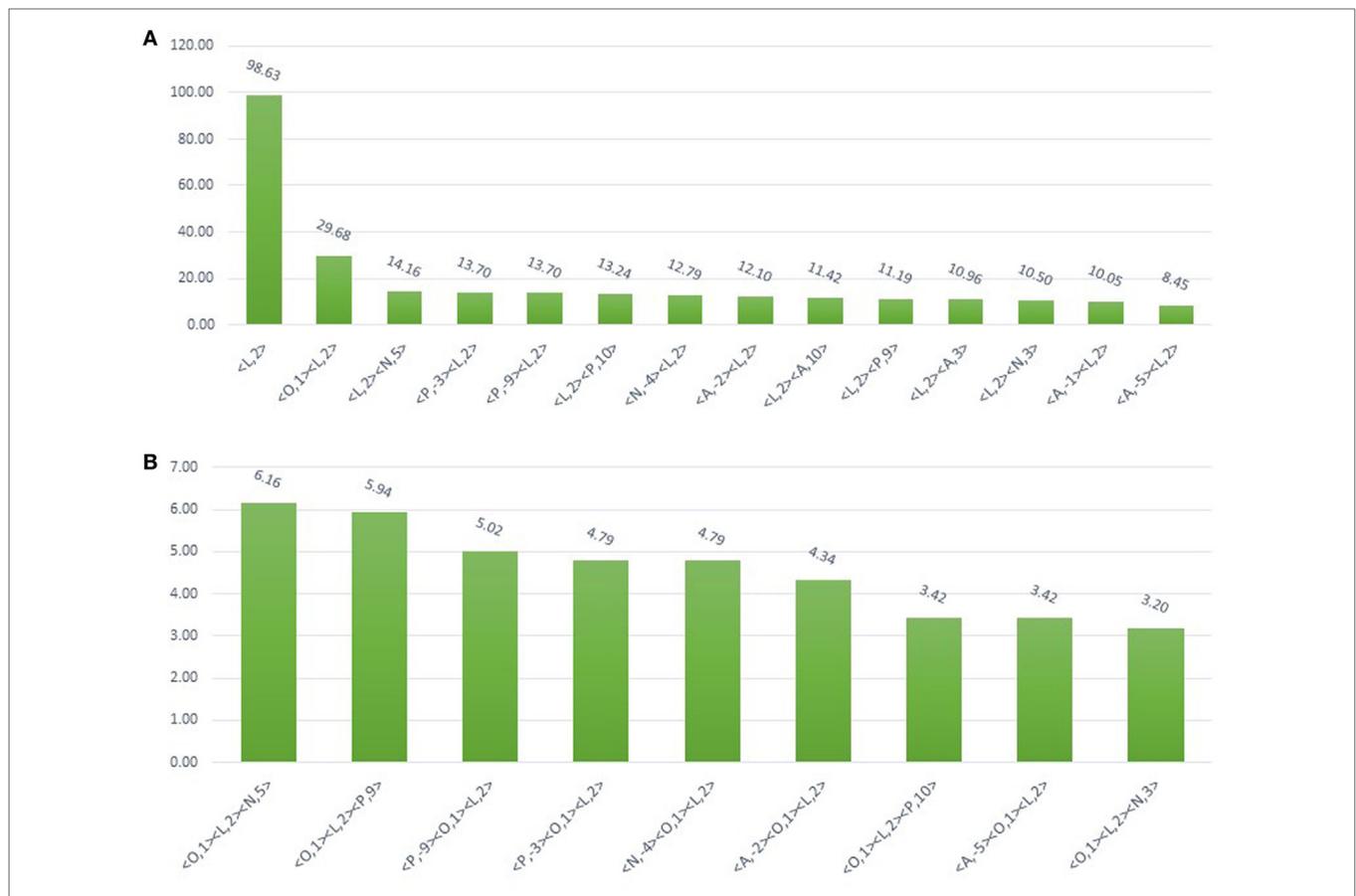
## Validation of the Association Patterns Mined by MAPRes

In the first level of validation, the association patterns mined by MAPRes for *N*-glycosylated and non-*N*-glycosylated residues were determined by utilizing the NetNGlyc server. There were 226 potential sites for *N*-glycosylation (N+) found by NetNGlyc from the selected 15 proteins. The rest of the 642 Asn were considered to be N-residues. Peptides of 21 amino acids were generated for both N+ and N- residues and patterns mined by MAPRes for *N*-glycosylated and non-*N*-glycosylated, and found in both datasets. The confirmatory percentage for association patterns in the peptide dataset was found to be 94% for the

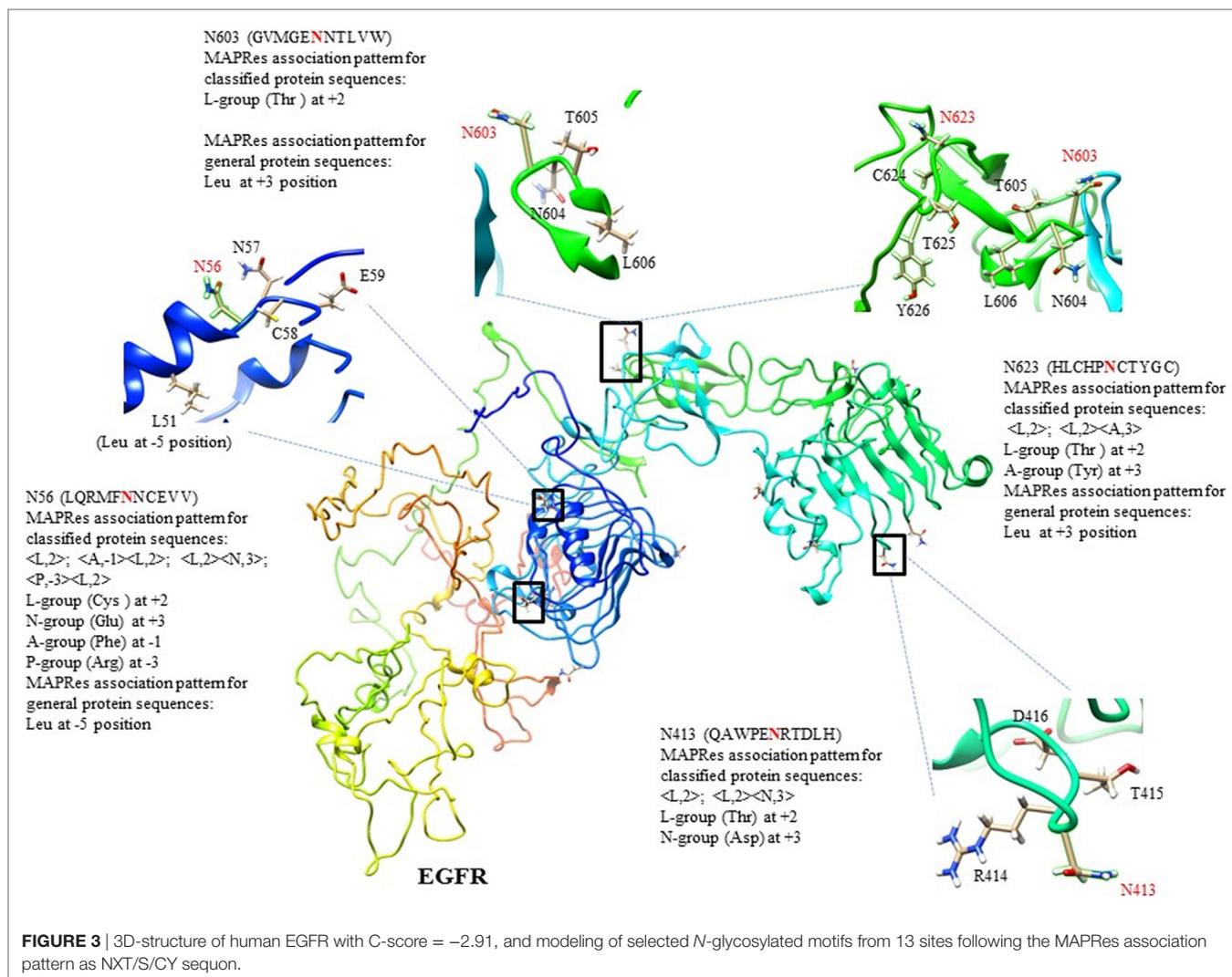
*N*-glycosylated dataset and 59% for the non-glycosylated dataset (Table 5). As for the validation of the association patterns mined by the MAPRes on the basis of polarity and charge, an explicit technique was used. Namely, the amino acids of the predicted and non-predicted dataset were decoded according to the defined classification of the amino acids, and association patterns searched in this dataset. This validation procedure for encoded datasets provided a good percentage of conformity, with 85% for the *N*-glycosylated and 62% for the non-*N*-glycosylated sites (Table 5). In the second level of validation, the 438 *N*-glycosylated motifs from 130 experimentally known proteins were retrieved from UniProt (reviewed human proteins) and sequence-based

statistical analysis of *N*-glycosylated motifs validated the association rules of MAPRes. From the general dataset, around 99% of *N*-glycosylated motifs were found in the NXS/TY sequon (40% of *N*-glycosylated motifs were found in the NXS/TY sequon, 59% in the NXTY sequon). From classified datasets, among 438 residues, 99% were L amino acids at position 2 (<L,2>), and 30% of total were O-amino acids at position 1, in combination with L amino acids at position 2, to validate the <O,1><L,2> rule. Therefore, a strong correlation was observed between polar and non-polar R-groups in the NXS/TY sequon. The presence of O amino acids (G, A, V, I, L) at position 1 and of L amino acids at position 2 were characteristic for *N*-glycosylation (Figure 2A). From 99% *N*-glycosylated motifs with L amino acids at position 2, 14%, 12% and 10% residues were occupied with amino acids containing negatively charged polar R-group (N amino acids: Asp/Glu) at 5, -4, and 3 positions. In addition to O and N amino acids, A amino acids also showed preference with at -2, 10, 3, -1, and -5 positions, in combination with L amino acids at 2 (Figure 2A).

The presence of aromatic amino acids penultimate or vicinal to the *N*-glycosylated site supports the *N*-glycosylation mechanism in the presence of Cyst/Ser/Thr at 2. The P-amino acids, in combination with L at 2, occupied 11 to 13% of positions -3, -9, and 9. Hence, the particular combination of non-polar, polar, aromatic, positive, and negatively charged residues in the vicinity of Asn supports the *N*-glycosylation process (Figure 2). For instance, in EGFR, an important oncogenic driver in various carcinoma, the MAPRes association patterns for classified protein sequences validated L (Thr/Ser/Cys) at position 2 (Figure 3). Similarly, the occupancy by O amino acids at position 1, N at 3/-4, at -1/-2/3/10, and P at -3/9/10 was validated by the MAPRes association pattern in 13 *N*-glycosylated motifs [N56 (NNCE), N73 (NYDL), N128 (NKTG), N175 (NMSM), N196 (NGSC), N352 (NATN), N361 (NCTS), N413 (NRTD), N444 (NITS), N528 (NVSr), N568 (NITC), and N603 (NNTL)], localized in the extracellular EGFR (Figure 3), suggesting the preference of non-polar aliphatic, negatively charged polar, positively charged



**FIGURE 2** | Bar plot showing the 438 *N*-glycosylated residues from 130 UniProt-reviewed proteins validating the MAPRes association rule for preferred *N*-glycosylation sites. **(A)** From 438 *N*-glycosylated motifs of 130 proteins, based on charge and polarity preference, 99% of *N*-glycosylated residues were found with L-group at 2 (<L,2>) indicating the significant preference for *N*-glycosylation. The other classified groups (O, N, P, and A) were found in charge and polarity combination with L at 2. For instance, 30% *N*-glycosylated motifs were observed with O at 1 and L at 2 (<O,1><L,2>), 14% N at position 5 (<L,2> <N,5>), 13% P (<L,2> <P,10> ; <P,-9><L,2> ; <P,-3> <L,2>), and 12% A at -2 (<A,-2><L,2>). **(B)** Three groups combination (<O,1><L,2><N,5> ; <O,1><L,2><P,9> ; <P,-9><O,1> <L,2> ; <P,-3><O,1> <L,2> ; <N,-4><O,1> <L,2> ; <A,-2><O,1> <L,2> ; <O,1><L,2> <P,10> ; <A,-5><O,1> <L,2> ; <O,1><L,2><N,3>) were ranging from 6 to 3% motifs.



polar, and aromatic amino acids in combination of L-group amino acids for *N*-glycosylation in EGFR. In addition to EGFR, in E-cadherin (E-cad), another key regulator of neoplastic cells, the extracellular *N*-glycosylation sites “N558 (NSTY), N570 (NGSP), N622 (NTSP), N637 (NWTI)” were modeled to validate the MAPRes association pattern (Figure 4).

## DISCUSSION

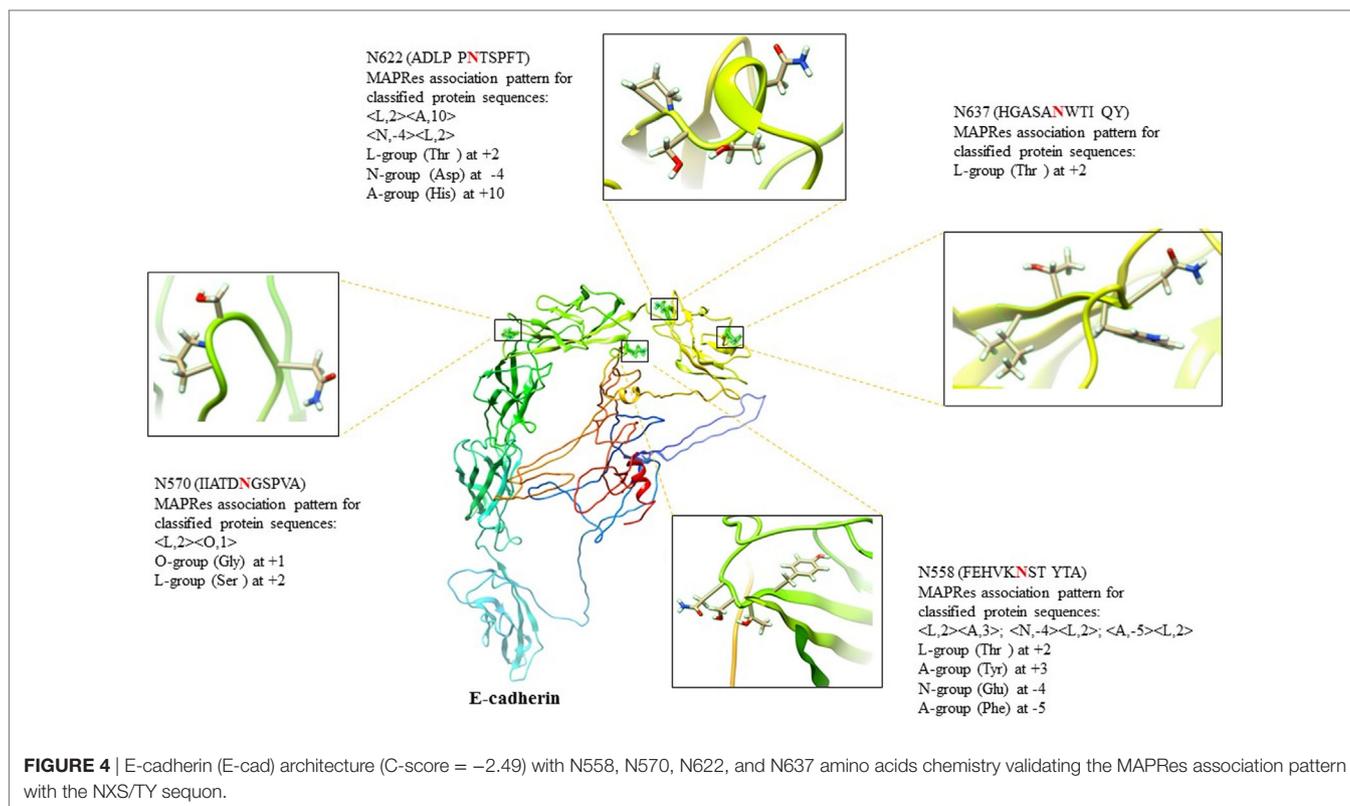
Different tools and techniques have been developed to understand the decisive roles of glycoconjugates in biological systems (3, 4, 6, 7, 10, 18, 27, 31–34). However, many important facets of glycosylation in protein function remain to be explained.

*N*-glycosylation is a biologically relevant protein modification for the regulation of signaling, protein–protein interactions and protein folding and stability (6, 31, 35, 36). *N*-glycosylation is initiated in the ER and modified within the Golgi stacks, with the removal of mannose residues and sequential addition of GlcNAc and fucose. However, during passage through the Golgi,

*N*-glycosylation depends both upon the structure of the protein as well as the amount and quality of processing enzymes available (31, 36).

Based on the hypothesis that the nature of both penultimate and vicinal amino acids contributes to the regulation of site-specific glycosylation, different computational tools have been developed to unravel the upstream and downstream characteristics of glycosylation sites (12, 37). Such computational tools should be useful in screening glycosylation sites in oncologically relevant proteins from cancer tissues (12, 38, 39). MAPRes is one of such computational tools that estimates significantly preferred sites around modified residues and mines association rules/patterns (10). The preference estimation of significantly preferred sites and association pattern-mining at various support levels has already been studied extensively for large amounts of data (38).

In this study, MAPRes supports the Ser and Thr at position 2 at the highest level compared to the other mined patterns (Table 4). At the level of charge and polarity, amino acids with polar-uncharged R-groups encoded in as L residues at 2 position



showed the highest support level (95%) for *N*-glycosylation. In other studies, the same motif was identified by the presence of Ser/Thr at position 2 around glycosylated Asn that can affect the polarity of the protein (39–41). Our results are consistent with a recent study stating that the Asn is preferably glycosylated when position 2 is occupied by Ser/Thr, and provided position 1 is not Pro (42). In addition, our results also highlight the marked preference for Thr at position 2 instead of Ser in NXS/T motif which is also consistent with the literature (40, 41). The combined effects of both O amino acids, carrying non-polar aliphatic R-groups, at position 1, and L amino acids with polar R-groups at position 2, were found to enhance *N*-glycosylation, as exemplified by the rule <A,-1> <L,2> (Figure 2A). Hence, to support *N*-glycosylation, non-polar R-group containing O amino acids were found sandwiched between polar R-groups containing Asn and Cys/Ser/Thr. Moreover, the A or aromatic acid amino acids, when juxtaposed to glycosylation sites in the tertiary fold, promoted protein folding along with the *N*-glycans during *N*-glycosylation (31, 33). In our study, presence of A amino acids at positions 3/-1/-2/-5 in combination with L at 2, indicates the preference of aromatic amino acids in the vicinity of the *N*-glycosylation site to promote protein folding via glycan-protein hydrophobic and nucleophilic interactions (Table 5; Figure 3). Interestingly, the same behavior of aromatic amino acids was suggested in the *N*-glycosylation process (37, 41, 43). In addition, our results also support the presence of polar acidic (N: Asp/Glu) residues preferably at positions 3/5/-4 (Table 5; Figure 2) which may help maintaining the solubility and ionic interactions of proteins.

In the first level of validation, the association patterns mined by MAPRes for *N*-glycosylated and non-*N*-glycosylated residues were determined by utilizing the NetNGlyc server. There were 226 potential sites for *N*-glycosylation (N+) found by NetNGlyc from the selected 15 proteins. The rest of the 642 Asn were considered to be N-residues. From the second level of validation, we selected the 438 experimentally defined glycopeptides from 130 reviewed proteins and found validation of MAPRes rules in all 438 glycopeptides. We further modeled two cell membrane proteins, epidermal growth factor receptor (EGFR), and E-cad, because these proteins are heavily decorated with *N*-glycans and involved in cancers as a result of aberrant or excessive glycosylation (18, 31, 34, 44, 45). Inappropriate *N*-, and *O*-glycosylation often results in malfunction of EGFR—a validated oncogenic target in cancers, such as lung and breast (31, 44, 46–48). Actually, complete abrogation of EGFR *N*-glycosylation by tunicamycin led to increased susceptibility of EGFR to the tyrosine kinase (TK) inhibitor erlotinib, a frequently utilized drug to downregulate EGFR activation supporting constitutive proliferation of non-small cell lung cancer (49). This effect of *N*-glycosylation (a modification of the extracellular portion of EGFR) on the intracellular TK enzyme strongly suggests that the allosteric organization of EGFR TK is dependent on extracellular *N*-glycosylation events and that EGFR functions are indeed linked to the *N*-glycosylation status of EGFR.

In EGFR, 13 canonical *N*-glycosylation sites (N56, N73, N128, N175, N196, N352, N361, N413, N444, N528, N568, and N603) in its extracellular domain were reported in the UniProt database (44, 50–53). Amino acids 1–165 (domain I) and 310–480

(domain III) are involved in ligand binding, and domain II (165–309) constitutes the dimerization arm (46). These three domains, with domain IV (481–620) are essential for ligand binding and receptor dimerization/multimerization of the EGFR and result in the formation of an active EGFR. In this study, all 13 canonical motifs [N56 (NNCE), N73 (NYDL), N128 (NKTG), N175 (NMSM), N196 (NGSC), N352 (NATN), N361 (NCTS), N413 (NRTD), N444 (NITS), N528 (NVS), N568 (NITC), and N603 (NNTL)] were found, validating the rules mined with MAPRes for *N*-glycosylation (Figure 3). In the first glycosylated motif NNCE, the presence of Cys at 2 indicates the involvement of NXCYS sequon in *N*-glycosylation (44), and validates the respective MAPRes rules ( $\langle A, -1 \rangle \langle L, 2 \rangle$ ;  $\langle P, -3 \rangle \langle L, 2 \rangle$ ;  $\langle L, 2 \rangle \langle N, 3 \rangle$ ;  $\langle L, 2 \rangle \langle A, 10 \rangle$ ). In addition to the L amino acids at 2, N acidic amino acid at 3 and A-group at -1 positions cumulatively promote *N*-glycosylation in NNCE. The structural modeling of NNCE suggests that the aromatic R-group of phenylalanine at position -1 (in close proximity of N56) achieves the conformation favoring *N*-glycosylation of N56 (Figure 3). The same behavior of aromatic amino acids in the vicinity of Asn residues were proposed in the *N*-glycosylation process (41, 43). From our collective results for 13 EGFR *N*-glycosylated motifs, A residues were localized at positions +3/-1/-2, L at +2, N at 3/-4/5, and P at -3/9/10 position to support the *N*-glycosylation process. Most probably, the charge and polarity environments of these 13 motifs represent optimal environments for *N*-glycosylation, as the EGFR studied were isolated from proliferating cultured cancer cells.

Activation of EGFR depends upon its multimerization, and a model has been recently proposed whereby residues in the extracellular domain IV have been identified that promote multimerization and enhance intracellular TK activity (54). Although abolition of *N*-glycosylation at Asn 544 did not alter the TK activity of the mutant, whether inappropriately *N*-glycosylated sites in domain IV would do so as well has not been investigated. Multimerization of EGFR occurs in plasma membrane domains enriched in various glycoconjugates (55) and *N*-glycosylated membrane proteins have been identified that may control EGFR activity. For instance, *N*-glycosylated  $\alpha 5$  integrin interacts with EGFR, promotes complex formation with  $\alpha 5 \beta 1$   $\alpha 6 \beta 4$  heteropolymers and integrins, therefore, constitute proteins that may express mutated *N*-glycosylation sites in neoplastic diseases (56, 57).

E-cadherin, a calcium-dependent cell–cell adhesion molecule, is another key regulator of normal and neoplastic cells (18, 32, 34). The genome-wide association studies highlighted a strong correlation between *N*-acetylglucosaminyltransferase-III (GnT-III)-based structural modification of E-cad *N*-glycosylation in epithelial-to-mesenchymal transitions and susceptibility to colorectal cancer (18). In addition, the aberrant glycosylation on E-cad by GnT-V resulted in a poorer survival rate for gastric cancer patients (18, 34). E-cad contains four canonical significantly preferred extracellular *N*-glycosylation sites “N558 (NSTY), N570 (NGSP), N622 (NTSP), N637 (NWTI)” (45), supporting the mining rules of O amino acids at 1, L at 2, at -4, and A at 3, 5, and -10 position in *N*-glycosylation. In recent studies on site-directed mutagenesis in E-cad, the N558 (as Asn554), and N637 (as Asn633) were found aberrantly decorated with complex-type

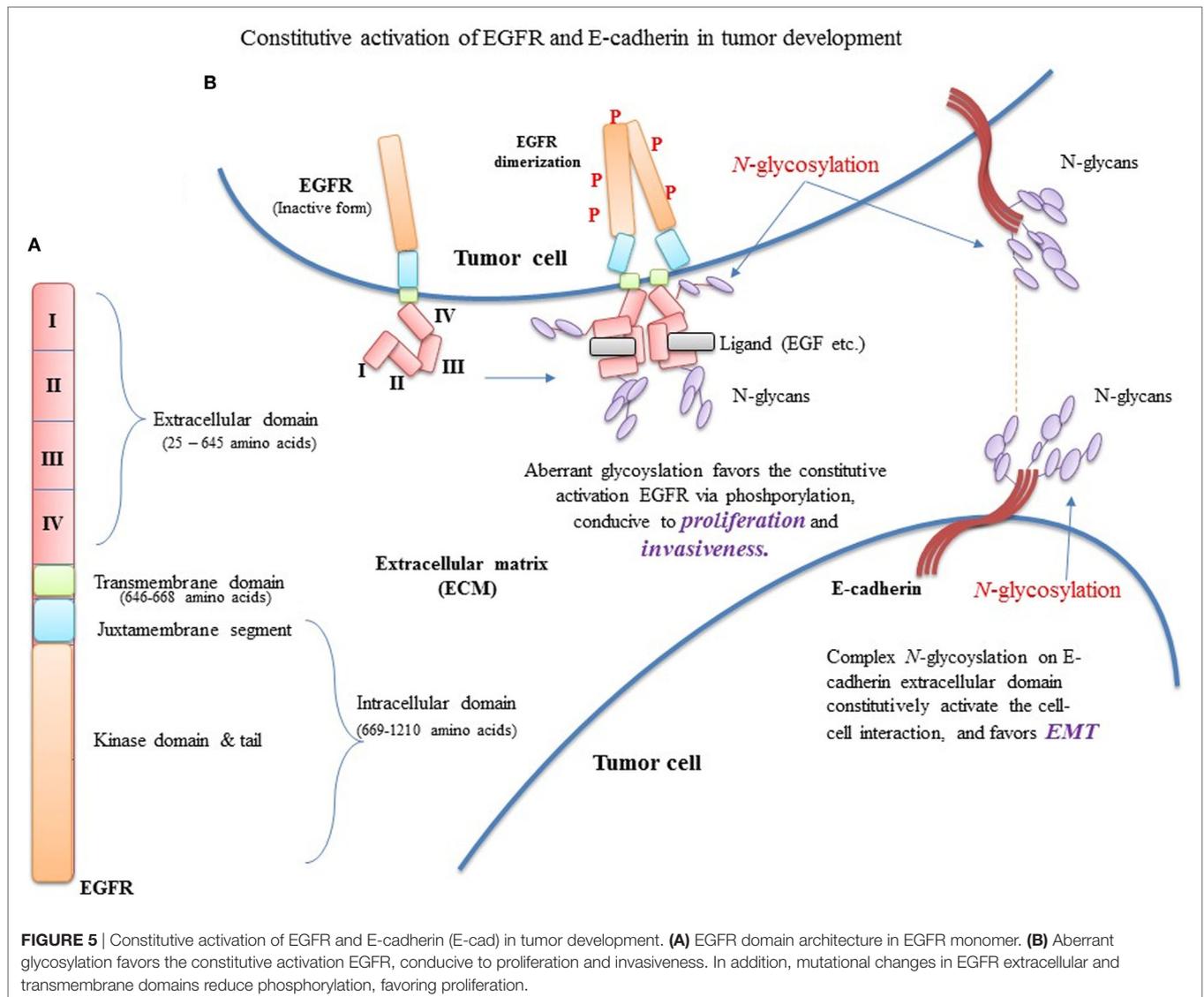
high-mannose *N*-glycans ( $\beta 1,6$  GlcNAc-branched), and critical for regulating the biological functions of E-cad in cancer (18, 34). For instance, the site-specific glycosylation at N558 elongate the deleterious branching structures and induces the tumor progression in gastric cancer (34). In our results, the presence of L amino acids at 2, N at -4, and A at 3, -5 collectively promote the site-specific glycosylation at N558 and direct the elongation of deleterious branching structures due to high electron density and polar environment which may induce the increased tumor progression in gastric cancer. In addition, the presence of N622 and N637 in close proximity of N558 (Figure 4) may further induce the elongation of glycan structures on E-cad.

Overall, the combination of polar and non-polar components, and acidic and basic groups of both penultimate and vicinal amino acids surrounding the *N*-glycosylation site, favor the normal *N*-glycosylation process. However, the mutational changes at *N*-glycosylation site and/or vicinal preferred sites may impact the oncogenicity of proteins (Figure 5) by modifying their polar chemistry.

## Conclusion and Future Prospects

Glycans remodel the protein backbone by reducing its conformational freedom and the loss of configurational entropy upon folding. To unravel the phenomenon of *N*-glycan remodeling in both ER and Golgi, charge and polarity of penultimate and vicinal amino acids were found important, and alteration in charge and polarity environment of Asn-containing epitopes results the heavy glycosylation (both on preferred and non-preferred sites) and reinforces the neoplastic phenotype. Thereby, the genetic alterations in many transmembrane receptors and adhesion proteins are important to determine the success or failure of glycosylation in these transmembrane proteins. This is particularly critical in neoplastic cells where disease-associated risks need to be assessed for many transmembrane oncogenic TKs.

In our analysis, cumulative affinity of both L amino acids at position 2 (with polar-uncharged R-groups) and O amino acids at position 1 (with non-polar aliphatic R-groups) was found significantly preferred for *N*-glycosylation. Moreover, remote parts of the protein chain rich in aromatic amino acids (A) support protein folding and promote the glycan–protein hydrophobic and nucleophilic interactions at positions 3/-5/-1 (Tables 3 and 4), in the presence of L amino acids at position 2, as in EGFR, TGF $\beta$ 1, and E-cad (Figures 3 and 4). The N amino acids at positions -4, 3, and 5 help maintaining the hydrophilic and other ionic interactions, as highlighted in EGFR and vitamin K-dependent protein C. Therefore, we suggest that the charge and polarity of R-groups in both penultimate and vicinal residues, notably the non-polar R-groups containing residues (O) at position 1, and polar R-groups containing L residues at position 2 along with the aromatic residues at positions 3/-1/-2/-5, acidic residues at 3/5/-4, and basic residues at -3/9/10 positions significantly contribute to normal *N*-glycosylation, and mutational alterations at these positions will significantly change the charge and polarity environment and cause constitutive activation of the glycoprotein by aberrant glycosylation leading to aggregation and intracellular phosphorylation in several carcinomas.



## AUTHOR NOTE

This publication is dedicated to the memory of Professor Dr. Nasir-ud-Din (1937–2016), founder and chairman of Institute of Molecular Sciences and Bioinformatics and Fellow of Pakistan Academy of Sciences.

## AUTHOR CONTRIBUTIONS

MM designed the basic research theme, and contributed for data collection, analysis and validation. He also contributed significantly in manuscript writing. ZI defined and designed the research theme, methods, and generated various results. WQ contributed to generating results and participated in the improvement of the results interpretations. DH provided suggestions for improving the structure of the basic theme of

the research and writing of the manuscript. He also helped to improve the scope of the study.

## ACKNOWLEDGMENTS

The ZI, WQ, and DH thank Dr. Uzma Yaqoob (Gynecologist at Services Institute of Medical Sciences, Lahore), Dr. Muhammad Akbar Saeed (Ex. Director General, PCSIR and acting chairman, Institute of Molecular Sciences and Bioinformatics) and PureChem Pvt. Ltd. for their financial support. We would also like to thank Dr. Muhammad Akbar Saeed for his scholastic support to the Institute of Molecular Sciences and Bioinformatics.

## FUNDING

The funding bodies have no roles in the study design, data collection and analysis, and manuscript writing.

## REFERENCES

- Pereira NA, Pu HX, Goh H, Song Z. Golgi phosphoprotein 3 mediates the Golgi localization and function of protein O-linked mannose beta-1,2-N-acetylglucosaminyltransferase. *J Biol Chem* (2014) 289:14762–70. doi:10.1074/jbc.M114.548305
- Liu Y, Xia B, Gleason TJ, Castaneda U, He M, Berry GT, et al. N- and O-linked glycosylation of total plasma glycoproteins in galactosemia. *Mol Genet Metab* (2012) 106(4):442–54. doi:10.1016/j.ymgme.2012.05.025
- Hussain MR, Nasir J, Al-Aama JY. Clinically significant missense variants in human GALNT3, GALNT8, GALNT12, and GALNT13 genes: intriguing in silico findings. *J Cell Biochem* (2014) 115(2):313–27. doi:10.1002/jcb.24666
- Wolfert MA, Boons GJ. Adaptive immune activation: glycosylation does matter. *Nat Chem Biol* (2013) 9(12):776–84. doi:10.1038/nchembio.1403
- Hussain MRM, Din N, Hassan M, Razaq A, Iqbal Z. Physiological significance of Fuc and Sialic acid containing glycans in the body. *Arabian J Chem* (2016) 9(Suppl 1):S9–20. doi:10.1016/j.arabjc.2011.06.028
- Freeze HH. Genetic defects in the human glycome. *Nat Rev Genet* (2006) 7(7):537–51. doi:10.1038/nrg1894
- Pinho SS, Reis CA. Glycosylation in cancer: mechanisms and clinical implications. *Nat Rev Cancer* (2015) 15(9):540–55. doi:10.1038/nrc3982
- Kaleem A, Ahmad I, Hoessli DC, Walker-Nasir E, Saleem M, Shakoori AR, et al. Epidermal growth factor receptors: function modulation by phosphorylation and glycosylation interplay. *Mol Biol Rep* (2009) 36(4):631–9. doi:10.1007/s11033-008-9223-6
- O'Connor SE, Imperiali B. Modulation of protein structure and function by asparagine-linked glycosylation. *Chem Biol* (1996) 3(10):803–12. doi:10.1016/S1074-5521(96)90064-2
- Ahmad I, Qazi WM, Khurshid A, Ahmad M, Hoessli DC, Khawaja I, et al. MAPRes: mining association patterns among preferred amino acid residues in the vicinity of amino acids targeted for post-translational modifications. *Proteomics* (2008) 8(10):1954–8. doi:10.1002/pmic.200700657
- Nagae M, Yamaguchi Y. Function and 3D structure of the N-glycans on glycoproteins. *Int J Mol Sci* (2012) 13(7):8398–429. doi:10.3390/ijms13078398
- Ahmad I, Hoessli DC, Qazi WM, Khurshid A, Mehmood A, Walker-Nasir E, et al. MAPRes: an efficient method to analyze protein sequence around post-translational modification sites. *J Cell Biochem* (2008) 104(4):1220–31. doi:10.1002/jcb.21699
- Banerjee DK. N-glycans in cell survival and death: cross-talk between glycosyltransferases. *Biochim Biophys Acta* (2012) 1820(9):1338–46. doi:10.1016/j.bbagen.2012.01.013
- UniProt Consortium. UniProt: a hub for protein information. *Nucleic Acids Res* (2014) 43(D1):D204–12. doi:10.1093/nar/gku989
- Boutet E, Lieberherr D, Tognolli M, Schneider M, Bantsal P, Bridge AJ, et al. UniProtKB/Swiss-prot, the manually annotated section of the UniProt knowledge base: how to use the entry view. *Methods Mol Biol* (2016) 1374:23–54. doi:10.1007/978-1-4939-3167-5\_2
- Williams R, Ma X, Schott RK, Mohammad N, Ho CY, Li CF, et al. Encoding asymmetry of the N-glycosylation motif facilitates glycoprotein evolution. *PLoS One* (2014) 9(1):e86088. doi:10.1371/journal.pone.0086088
- Hofherr A, Wagner C, Fedeles S, Somlo S, Kottgen M. N-glycosylation determines the abundance of the transient receptor potential TRPP2. *J Biol Chem* (2014) 289(21):14854–67. doi:10.1074/jbc.M114.562264
- Carvalho S, Catarino T, Dias A, Kato M, Almeida A, Hessling B, et al. Preventing E-cadherin aberrant N-glycosylation at Asn-554 improves its critical function in gastric cancer. *Oncogene* (2015) 35:1619–31. doi:10.1038/onc.2015.225
- Chugh S, Gnanapragassam VS, Jain M, Rachagani S, Ponnusamy MP, Batra SK. Pathobiological implications of mucin glycans in cancer: sweet poison and novel targets. *Biochim Biophys Acta* (2015) 1856(2):211–25. doi:10.1016/j.bbcan.2015.08.003
- Freire-de-Lima L, Previato JO, Mendonça-Previato L. Editorial: glycosylation changes in cancer: an innovative frontier at the interface of cancer and glycobiology. *Front Oncol* (2016) 6:254. doi:10.3389/fonc.2016.00254
- Nardy AFFR, Freire-de-Lima L, Freire-de-Lima CG, Morrot A. The sweet side of immune evasion: role of glycans in the mechanisms of cancer progression. *Front Oncol* (2016) 6:54. doi:10.3389/fonc.2016.00054
- Nagarajan U, Pakkiriswami S, Pillai AB. Sugar tags and tumorigenesis. *Front Cell Dev Biol* (2015) 3:69. doi:10.3389/fcell.2015.00069
- El Bairi K, Kandhro AH, Gouri A, Mahfoud W, Louanjli N, Saadani B, et al. Emerging diagnostic, prognostic and therapeutic biomarkers for ovarian cancer. *Cell Oncol* (2017) 40(2):105–18. doi:10.1007/s13402-016-0309-1
- Natoni A, Macauley MS, O'Dwyer ME. Targeting selectins and their ligands in cancer. *Front Oncol* (2016) 6:93. doi:10.3389/fonc.2016.00093
- Pakkiriswami S, Couto A, Nagarajan U, Georgiou M. Glycosylated notch and cancer. *Front Oncol* (2016) 6:37. doi:10.3389/fonc.2016.00037
- Cox OT, O'Shea S, Tresse E, Bustamante-Garrido M, Kiran-Deevi R, O'Connor R. IGF-1 receptor and adhesion signaling: an important axis in determining cancer cell phenotype and therapy resistance. *Front Endocrinol* (2015) 6:106. doi:10.3389/fendo.2015.00106
- Vasconcelos-dos-Santos A, Oliveira IA, Lucena MC, Mantuano NR, Whelan SA, Dias WB, et al. Biosynthetic machinery involved in aberrant glycosylation: promising targets for developing of drugs against cancer. *Front Oncol* (2015) 5:138. doi:10.3389/fonc.2015.00138
- Ahmad I, Mehmood A, Khurshid A, Qazi WM, Hoessli DC, Walker-Nasir E, et al. Phosphoproteome sequence analysis and significance: mining association patterns around phosphorylation sites utilizing MAPRes. *J Cell Biochem* (2009) 108(1):64–74. doi:10.1002/jcb.22220
- Blom N, Sicheritz-Ponten T, Gupta R, Gammeltoft S, Brunak S. Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics* (2004) 4(6):1633–49. doi:10.1002/pmic.200300771
- Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER suite: protein structure and function prediction. *Nat Methods* (2015) 12(1):7–8. doi:10.1038/nmeth.3213
- Hussain MRM, Hoessli DC, Fang M. N-acetylgalactosaminyltransferases in cancer. *Oncotarget* (2016) 7(33):54067–81. doi:10.18632/oncotarget.10042
- Pinho SS, Seruca R, Gärtner F, Yamaguchi Y, Gu J, Taniguchi N, et al. Modulation of E-cadherin function and dysfunction by N-glycosylation. *Cell Mol Life Sci* (2011) 68(6):1011–20. doi:10.1007/s00018-010-0595-0
- Wang Y, Gao J, Guo X, Tong T, Shi X, Li L, et al. Regulation of EGFR nanocluster formation by ionic protein-lipid interaction. *Cell Res* (2014) 24(8):959–76. doi:10.1038/cr.2014.89
- Carvalho S, Oliveira T, Bartels ME, Miyoshi E, Pierce M, Taniguchi N, et al. O-mannosylation and N-glycosylation: two coordinated mechanisms regulating the tumour suppressor functions of E-cadherin in cancer. *Oncotarget* (2016) 7(40):65231–46. doi:10.18632/oncotarget.11245
- Ohtsubo K, Marth JD. Glycosylation in cellular mechanisms of health and disease. *Cell* (2006) 126(5):855–67. doi:10.1016/j.cell.2006.08.019
- Palade G. Intracellular aspects of the process of protein synthesis. *Science* (1975) 189(4206):867. doi:10.1126/science.189.4206.867-b
- Hamby SE, Hirst JD. Prediction of glycosylation sites using random forests. *BMC Bioinformatics* (2008) 9:500. doi:10.1186/1471-2105-9-500
- Iqbal Z, Hoessli DC, Kaleem A, Munir J, Saleem M, Afzal I, et al. Influence of the sequence environment and properties of neighboring amino acids on amino-acetylation: relevance for structure-function analysis. *J Cell Biochem* (2013) 114(4):874–87. doi:10.1002/jcb.24426
- Senger RS, Karim MN. Variable site-occupancy classification of N-linked glycosylation using artificial neural networks. *Biotechnol Prog* (2005) 21(6):1653–62. doi:10.1021/bp0502375
- Kasturi L, Chen H, Shakin-Eshleman SH. Regulation of N-linked core glycosylation: use of a site-directed mutagenesis approach to identify Asn-Xaa-Ser/Thr sequons that are poor oligosaccharide acceptors. *Biochem J* (1997) 323(Pt 2):415–9. doi:10.1042/bj3230415
- Petrescu AJ, Milac AL, Petrescu SM, Dwek RA, Wormald MR. Statistical analysis of the protein environment of N-glycosylation sites: implications for occupancy, structure, and folding. *Glycobiology* (2004) 14(2):103–14. doi:10.1093/glycob/cwh008
- Zielinska DF, Gnad F, Wisniewski JR, Mann M. Precision mapping of an in vivo N-glycoproteome reveals rigid topological and sequence constraints. *Cell* (2010) 141(5):897–907. doi:10.1016/j.cell.2010.04.012
- Jitsuahara Y, Toyoda T, Itai T, Yamaguchi H. Chaperone-like functions of high-mannose type and complex-type N-glycans and their molecular basis. *J Biochem* (2002) 132(5):803–11. doi:10.1093/oxfordjournals.jbchem.a003290
- Sato C, Kim JH, Abe Y, Saito K, Yokoyama S, Kohda D. Characterization of the N-oligosaccharides attached to the atypical Asn-X-Cys sequence of

- recombinant human epidermal growth factor receptor. *J Biochem* (2000) 127(1):65–72. doi:10.1093/oxfordjournals.jbchem.a022585
45. Zhou F, Su J, Fu L, Yang Y, Zhang L, Wang L, et al. Unglycosylation at Asn-633 made extracellular domain of E-cadherin folded incorrectly and arrested in endoplasmic reticulum, then sequentially degraded by ERAD. *Glycoconj J* (2008) 25(8):727–40. doi:10.1007/s10719-008-9133-9
  46. Lemmon MA, Schlessinger J, Ferguson KM. The EGFR family: not so prototypical receptor tyrosine kinases. *Cold Spring Harb Perspect Biol* (2014) 6(4):a020768. doi:10.1101/cshperspect.a020768
  47. Lin WL, Lin YS, Shi GY, Chang CF, Wu HL. Lewisy promotes migration of oral cancer cells by glycosylation of epidermal growth factor receptor. *PLoS One* (2015) 10(3):e0120162. doi:10.1371/journal.pone.0120162
  48. Huang M-J, Hu R-H, Chou C-H, Hsu C-L, Liu Y-W, Huang J, et al. Knockdown of GALNT1 suppresses malignant phenotype of hepatocellular carcinoma by suppressing EGFR signaling. *Oncotarget* (2015) 6(8):5650–65. doi:10.18632/oncotarget.3117
  49. Ling YH, Li T, Perez-Soler R, Haigentz M Jr. Activation of ER stress and inhibition of EGFR N-glycosylation by tunicamycin enhances susceptibility of human non-small cell lung cancer cells to erlotinib. *Cancer Chemother Pharmacol* (2009) 64(3):539–48. doi:10.1007/s00280-008-0902-8
  50. Wu S-L, Kim J, Hancock WS, Karger B. Extended range proteomic analysis (ERPA): a new and sensitive LC-MS platform for high sequence coverage of complex proteins with extensive post-translational modifications comprehensive analysis of beta-casein and epidermal growth factor receptor (EGFR). *J Proteome Res* (2005) 4(4):1155–70. doi:10.1021/pr050113n
  51. Zhen Y, Caprioli RM, Staros JV. Characterization of glycosylation sites of the epidermal growth factor receptor. *Biochemistry* (2003) 42(18):5478–92. doi:10.1021/bi027101p
  52. Ogiso H, Ishitani R, Nureki O, Fukai S, Yamanaka M, Kim J-H, et al. Crystal structure of the complex of human epidermal growth factor and receptor extracellular domains. *Cell* (2002) 110(6):775–87. doi:10.1016/S0092-8674(02)00963-7
  53. Chen R, Jiang X, Sun D, Han G, Wang F, Ye M, et al. Glycoproteomics analysis of human liver tissue by combination of multiple enzyme digestion and hydrazide chemistry. *J Proteome Res* (2009) 8(2):651–61. doi:10.1021/pr8008012
  54. Huang Y, Bharill S, Karandur D, Peterson SM, Marita M, Shi X, et al. Molecular basis for multimerization in the activation of the epidermal growth factor receptor. *Elife* (2016) 5:e14107. doi:10.7554/eLife.14107
  55. Lajoie P, Partridge EA, Guay G, Goetz JG, Pawling J, Lagana A, et al. Plasma membrane domain organisation regulates EGFR signaling in tumor cells. *J Cell Biol* (2007) 179:341–56. doi:10.1083/jcb.200611106
  56. Hang Q, Isaji T, Hou S, Im S, Fukuda T, Gu J. Integrin  $\alpha 5$  suppresses the phosphorylation of epidermal growth factor receptor and its cellular signaling of cell proliferation via N-glycosylation. *J Biol Chem* (2015) 290(49):29345–60. doi:10.1074/jbc.M115.682229
  57. Hang Q, Isaji T, Hou S, Zhou Y, Fukuda T, Gu J. N-glycosylation of integrin  $\alpha 5$  acts as a switch for EGFR-mediated complex formation of integrin  $\alpha 5\beta 1$  to  $\alpha 6\beta 4$ . *Sci Rep* (2016) 6:33507. doi:10.1038/srep33507

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Manwar Hussain, Iqbal, Qazi and Hoessli. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.