Current Insights in Mammalian Glycosylation: Implication of Glycoproteomics as Next-Generation Biomarkers in Diabetes Mellitus

Alok Raghav1*, Jamal Ahmad1, Zeeshan Ahmad Khan2, Brijesh Kumar Mishra3 and Rai Pawankumar Anirudh4

1Rajiv Gandhi Centre for Diabetes and Endocrinology, J.N Medical College, Aligarh Muslim University, Aligarh-202002, India
2Biological Rhythm Lab, Institute of Bioresources and Sustainable Development, Imphal, Manipur-795001, India
3Department of Endocrinology, Guru Teg Bahadur Hospital, Dilshad Garden, Delhi-110095, India
4ICAR-Department of Medicinal and Aromatic Plant Research, Ananad, Gujarat, India

Corresponding author: Alok Raghav, Rajiv Gandhi Centre for Diabetes and Endocrinology, J.N Medical College, Aligarh Muslim University, Aligarh-202002, UP, India, Tel: +91-571-2721173; E-mail: alokalig@gmail.com

Received date: February 19, 2017; Accepted date: March 14, 2017; Published date: March 21, 2017

Abstract

Glycosylation, an enzyme-directed, site-specific process is one of the major co-translational and post-translational modifications in proteins. So formed product, glycoproteins, mediate multiple functions of signal transductions, cell association, and inter and intra cell signalling. Glycoproteomics exploits all these biological functions and serves as the next generation diagnostic marker for analyzing proteins that are associated with micro- and macro-vascular complication of diabetes mellitus. This tool can unravel the secrets of the patterned expression of glycoproteins in healthy and patients with diabetes with secondary complications. The expression profile of several O-linked and N-linked mammalian glycoproteins delivers the knowledge and helps to create a database for expression in diabetes mellitus, for proposing a mechanistic approach to understanding the cellular pathway severely involved in the regulation of all glycoproteins. This future next generation glycopeptides tool will describe a novel strategy for the quantification of the several glycoproteins based on a specific chemical ligation described as reverse glycoblottting techniques, two-dimensional gel electrophoresis, western blotting, multiple reaction monitoring, LC-MS/MS analysis of glycosylated peptides along with lectin chromatography. These recent tools will reveal the expression pattern of several glycoproteins that will serve as a power tool precise diagnosis of diabetes associated complications.

Keywords: Glycoproteomics; Glycoproteins; Glycoproteins; Diabetes; Biomarkers; Post translational modifications

Introduction

Proteins are believed to be the building blocks of the body and have a keen involvement in the biological and physiological homeostasis of the organism [1]. Advancement in the proteomics in the past decades has made it an intensive and powerful diagnostic tool. One of the major areas of proteomics is glycoproteomics. Glycoproteomics relies on the study of glycoproteins that derived from glycosylation. The glycoproteomics identifies, characterizes and index proteins and peptides with carbohydrate moieties, including the post-translational modifications (PTMs) [2]. There are numerous PTMs that include ubiquitination, glycosylation, acetylation and phosphorylation [3]. It is more complicated and diverse class of protein modifications that plays a vital role in biological functions as well as disease progression and prognosis. It has been found recently that 80% mammalian proteins undergo PTMs [4] amongst them glycosylation constitute about 50% [5].

Glycosylation of proteins involves the addition of carbohydrate or glycans covalently to deliver protein function and activity. Glycans are the most abundant natural biopolymer and constitute the major component of the cells. The diverse function of glycans creates a new field of glycobiology to understand the structure, synthesis, regulation and metabolism of diverse glycans. Human blood groups, discovery gave birth to the importance and clinical applications of glycans function [6]. With numerous developments of the past decades in glycobiology, several parallel approaches evolve that includes genomics, proteomics, lipidomics, glycoproteomics, glycomics that interestingly imparts role of biomolecules in the disease process.

Glycoproteomics and glycomics are interrelated disciplines of glycobiology that dealt with knowing mechanism of glycoproteins and glycans respectively. Carbohydrate or glycans derived modifications significantly involved in the host-pathogen association, cell signalling, inflammation, growth, and development Expression profiling of cellular glycoproteins, glycoproteins library, with comparative between a normal homoeostatic and diabetes homeostatic, may reveal the fundamentals and underpinnings of the disease mechanism and prognosis. It also facilitates the determination of structurally and functionally modified proteins. Scheduled glycosylation maintains the homoeostatic environment in body, however if turn aberrant may result in abnormal and impaired changes in biological and physiological activity of protein that may, in turn, be the major culprit in molecular recognition of disease.

Diabetes has several biomarker but glycoprotein open new vistas in its diagnostics. It is a novel biomarker for an early detection of diabetes, its associated complications and also provides detailed information about risk factors and associated prognosis factors. Analysis of diabetes related glycoproteins’ expression (DRGE) may prove a novel next generation future biomarker in prognosis of diabetes and its associated micro and macro vascular complications. Diverse type of PTMs found in living organisms, including (1) N-linked glycosylation (2) O-linked glycosylation (3) C-linked...
glycosylation [7] and (4) S-linked glycosylation (occurs in bacteria) [8].

Furthermore, components of glycoproteins i.e., glycans plays important role in vast biological processes, including immune response, cell attachment and interaction, embryonic development that majorly involves sugar-sugar or sugar-protein interaction and recognition [9]. Mammalian glycans are resultant of an endogenous portfolio of several cellular metabolic enzymes and their substrate that have been conserved during the whole evolution spanning 1%-2% genome. Slight variation in the patterned expression of glycomes of organisms may be the molecular basis for interspecies recognition elements. Mammalian glycans are highly conserved along with species specific variations that may lead to the development of distinct trait [10]. Consequently, aberrant pattern of glycosylation of mammalian glycoproteins results in non-homeostatic expression of glycoproteins library in diseased individual compared to healthy. This has been implicated in several diseases, including immune deficiencies, genetic disorders, cardiovascular diseases, neurodegenerative disorders and cancer [11].

Several published literature reported implications of glycoproteins as targeted clinical biomarkers in complications, including Her2/neu (breast cancer), PSA (prostate specific antigen, marker for prostate cancer) and CA-125 (ovarian cancer) [12]. Indeed, impaired glycosylation in disease and associated complications showed a significant amount of biological information about mammalian glycome to be the clinical diagnostic marker [12]. In order to diagnose abrupt glycosylated protein expression in disease situations, responsive, specific, swift and strong methods are needed. Although, massive work is routinely going in various laboratories for identification of these glycopeptides markers, but the study of glycoproteins remains challenging. Frequent, comprehensive reviews related to this approach have been published in recent years covering identifications of glycoproteins [13]. This review focuses on the involvement of mammalian glycoproteome in diagnosis of diabetes and its related disorders, along with an emphasis on precise method development to create expression library of these glycopeptides in order to improve the glycoproteomics approach of human health.

Construction and Topology of Mammalian Glycosylation Mechanism

Monosaccharides play major role in mammalian glycosylation. Overall, nine monosaccharides participate in this enzyme mediated process of glycosylation. Ubiquitously precursor presents in the dietary components supply these nine monosaccharides to undergo processes of conserved biological synthesis.

Mammalian glycosylation involves primarily three diverse forms of modifications, including N-linked, O-linked and C-linked. Among these N-linked additions of sugars being the prime one. N-glycosylation in mammals occurs on asparagine’s amino residue at specific sequences of Asn-X-Ser/Thr occasionally on Cys residues with X being any of the amino acids except proline. The initiation of these modifications occurs on the cytoplasmic side of rough endoplasmic reticulum (RER), with binding of oligosaccharides Man5GlcNAc2 considered to be the precursor of dolichol pyrophosphate. Glc3Man9GlcNAc2-PP-dol is considered to be the appropriate core glycan precursor [14]. Depending upon the types and position of the monosaccharides encompass within chain, the N-glycans auxiliary grouped into high-mannose type, hybrid type and antennary type [14]. Components of high-mannose type of N-glycans are large mannose in the core structure. The Antennal region contains complex type of N-glycans Galβ1-3/4GlcNAc in N-terminal regime.

Furthermore, O-linked mammalian glycosylation occurs on serine or threonine residues with addition of N-acetyl glucosamine, N-acetylgalactosamine, fucose, mannose or xylose [14]. The most preferred O-linked glycosylation occurs with N-acetylgalactosamine linked with α-glycosidic bond to Ser/Thr.

A recent literature demonstrated that O-linked glycosylation in mucin may involve in immune response, cell adhesion and invasion that may be served as biomarker for diagnosis of cancer [12]. The nascent protein undergoes this post-translational modification after secreted from the classical secretory pathway from Golgi body. Hydroxyl (OH) group present on the Ser/Thr actively participates in O-linked glycosylation by addition of galactose/N-acetylgalcosamine. Post addition of GalNac followed by further elongation and termination (sialylation) confirm a large number of O-linked glycopeptides synthesized by glycosyltransferases [15]. Compared to N-linked glycosylation, the quantification of O-linked modifications has proven to be tough. The probable hurdles may due to the presence of readily available predicted sequences (Asn-Xaa-Ser/Thr or rare - Asn-Xaa-Cys/ Xaa is nay amino acid except Proline in N-linked glycosylation. In disparity, Ser/Thr in O-linked glycosylation can be predicted only in the beginning. Secondly, N-linked glycopeptides are well separated, while the O-linked glycopeptides are located in clusters thereby making it more difficult to identify.

A final and unexpected example of mammalian glycosylation is C-linked. This modification involves α-mannose group of certain tryptophan residues in proteins which have been previously described by Hofsteenge and their colleagues. C-mannosylation is known to occur on the Dol-P-Man consensus sequence along with Trp-Xaa-Xaa Trp sequence. However, present knowledge for C-linked glycosylation is limited. The degree of complexity, diversity, structural changes is subject to further research and investigations.

Regulation of Mammalian Glycosylation

Regulation of mammalian glycosylation is also complex as their structure and conformations. Glycans are specific targets of glycosyltransferase and glycosidase enzymes. Glycans chains are synthetized by glycosyltransferase while hydrolyzed by glycosidases enzyme class. Being anabolic components of glycosylation, glycosyltransferase in collaboration with glycosidases determine the structural and functional outcome of the biosynthesis of glycans [16]. Glicosaminoglycans and O-glycans biosynthetic pathways involve single enzyme reaction that completely depends on the topology of glycans, create by earlier enzymes for being the substrate for next [17].

Mammalian glycosylation mechanism of secretory proteins is a dynamic mechanism that involves alterations in the activity and accessibility of glycosyltransferase and glycosidase towards its substrate. Structural, topological variety in the glycan repertoire in the biosynthetic pathways generates numerous clinical biomarkers; some of these may correlate with the disease and its associated complications in every species to distinguish it from normal homeostasis. A previously supported literature proves elevated truncated O-glycans served as cancer biomarkers [18]. Transcription process modulates RNA expression of biosynthetic pathways of glycans formation governed by glycosyltransferases and glycosidases. The microarray analysis of genes involved in the biosynthetic pathways of glycans plays
a key role in the cellular expression pattern of glycans that may perhaps be predictive in providing vast information as biomarkers [19]. Similarly, gene transcription shows that core 2 GlcNAc-T-I is induced by transcription factor T-bet in T helper type 1 lymphocytes [20]. Although, numerous transcription factors regulate the glycosylation pathways gene expression, the concise effect is still unclear.

The existent evidence reflects that regulation of mammalian glycans occurs on post-transcriptional and post-translational stages too. Foremost alterations in glycoproteins are induced by loss of multiprotein complexes along with few chaperones that impair the glycosyltransferase mediated trafficking via endoplasmic reticulum and Golgi machinery [21]. A potential regulatory approach of glycoyltransferases and glycosidases is through phosphorylation event of their cytoplasmic tails, which might contribute to the intracellular trafficking and intermolecular rearrangements. Another potential regulatory factor in glycan formation is the disintegration of glycosyltransferases and glycosidases from their locations via proteolysis. The mammalian biosynthetic pathway enzymes are type 2 transmembrane glycoproteins containing luminal catalytic domain along with luminal proximal stem domain. Specific cleavage at stem domain may cleave catalytic domain separately, thereby may be the probable cause of inflammation as reported earlier [10]. This cleavage mechanism is still unclear and their ability to catalyze glycan formation among extracellular compartments has vanished. Lysosome mediated degradation of mammalian cell surface glycans is a key feature of infection causing pathogens like influenza virus [22]. Mammalian glycan formation is broadly governed by biosynthesis and availability of nucleotide sugar donors. Obstruction in the donor may lead to functional loss of transporters residing in endoplasmic reticulum and Golgi membrane and may abolish cellular glycans productions [23]. The Previously published study showed that extra supplements of glucosamine to hexosamine pathway elevate the glycans in mammalian cells [24]. The controlled regulation of this mammalian glycosylation is not yet known and careful investigations are required to study modulating mammalian glycan expression.

Glycoproteomic Development of Biomarker

The discovery of novel glycoproteins as novel biomarker has proven its clinical application in diabetes screening, diagnosis, and prognosis. The construction of glycoproteins expression library focuses on translational research in diabetes to fight against its associated complications. Mass spectroscopic (MS) glycoproteomics approach may be used for early diagnostic biomarkers to differentiate disease from healthy individuals. Although the conventional proteomics approach in biomarker discovery, some recent advance MS technologies have been introduced for identifications, characterization and quantifications of glycoproteins expression (Figure 1).

Matrix assisted laser desorption/ionization-time of flight (MALDI-TOF), electron spray ionization-mass spectroscopy (ESI-MS) and surface enhanced laser desorption/ionization (SELDI) in combination with liquid chromatography mass spectroscopy (LC-MS) may contribute in the validation and identification of novel glycopeptides biomarkers in diabetes and its associated complications. A recent study demonstrated the application of the MS approach to identify Sp1 as novel biomarker for hyperglycemia development due to activation of hexosamine pathway that has been easily implicated in diabetes and its associated pathogenesis [25].

Figure 1: Showing workflow of glycoproteomics identification and characterization of all glycosylated proteins.

Recently, due to generation of innovative novel MS-based glycoproteomics technologies the diabetes glycoproteomics area is explored too much for the human welfare and day by day novel biomarkers are evolving. Additionally, high throughput capacity of MS-based approach coupled with LC-MS for identification of biomarker in diabetes [26] is a new hope for development of glycoproteomics expression library. Indeed, large efficient efforts were made to sincerely investigate the diabetes glycoproteomics with various human specimens including adipose tissues, cell lines, serum, plasma, urine and saliva for creating a database of an expression library of glycosylated proteins in diabetes associated complications. The MS-based glycoproteomics research incorporates two principle approaches (1) glycoprotein identifications (2) pattern expression and recognition. Both approaches require high throughput automated computing and bioinformatics systems to create database for generation of expression library of glycoproteome associated with diabetes and its complications.

Specifically, glycoproteomics identifications involve glycoprotein isolation, enrichment, proteolytic digestion and identifications with advance MS based approaches. Table 1 showed various enrichment techniques of glycoproteins. Since, various biological samples of serum, plasma constitute mixtures of glycoproteins, therefore, extensive chromatographic separations is required that uses ion-exchange, affinity column separation, size exclusion and hydrophobic separation to minimize the complexity of proteins and enhance accuracy [27]. Recently develops a bottom-up approach has been developed for analysis of glycoprotein [28]. In this approach, glycoproteins are digested and undergo deglycosylation event with peptide N-glycosidase F (PNGase F) that cleaves glycosidic bonds between Asn and innermost GlcNAc of N-linked glycoproteins with few exceptions [29]. Digested and deglycosylated peptides were subjected to MS analysis coupled with MS tool, bioinformatics with numerous algorithm and glycoproteomic database analyzing software that identifies the specific expression of glycans and glycopeptides.
Complications as Biomarkers

Role of Glycopeptides in Diabetes and Its Associated Complications as Biomarkers

The beta cells are unique to contains large amount of OGT that enables them to respond to physiological increase in the glucose concentration by transforming the glucose to the OGT substrate UDP-GlcNAc thereby disruption in the O-linked glycosylation pathway [34]. A recent research in relation to diabetes associated hyperglycaemia reveals that increased sugar level in subjects with diabetes induce mitochondrial superoxide overproduction that in turn activate the hexosamine pathway and induces the plasminogen activator inhibitor-1 expression due to increased production of Sp1 glycosylation [32]. This study reveals the alterations in the pattern of glycoprotein expression in diabetes complications. A specific study on type 2 diabetes mellitus patients states that hyperglycemia induced endothelial glycoproteome changes lead to alterations in homeostatic endothelial signalling pathways [35]. In another recent study, glycoproteins are found to be the major components of the cardiac extracellular matrix that plays important role in cardiovascular disease progression [36]. Thrombospondin, tenascin-C and periostin are some of the extracellular matrix glycoproteins that modulate cardiac remodelling post myocardial infarctions [37]. Hexosamine biosynthetic pathway flux and OGlcnAc level increased in hyperglycaemic conditions. A recent study examined the modification of OGlcnAc alters Ca^{2+} binding proteins that may lead to cardiac dysfunction in patients with diabetes [38]. Subsequently Hu et al. reported increased OGT expression and OGlcnAc in heart from STZ induced diabetic mice that may lead to cardiomyocyte dysfunction [38]. Table 2 clearly demonstrates glycomarkers proteins and transcription factors express during diabetes and its associated complications.

**Table 1: Glycoproteome enrichment approach coupled with mass spectrometry tool for construction, identification and characterization of expression library of glycoproteins and glycos.

**Table 2: List of proteins and transcription factors contributed in diabetic associated complication that are modified and impaired directly or indirectly by O-GlcNAc. These proteins/transcription factors may contribute to the diagnostic biomarkers in the patients with altered glycosylation having diabetes or related complications.
Biomarker techniques have been evolving. Glycans and complete approach to glycoproteome library of few glycoproteins. In a Study by Hiromi Ito norvegicus, 3T3-L1: Cell Line; SVG: Human Fetal Glial Cell Line; sialylated or mucin type O-linked glycoproteins with the help of proteome. Murrell et al. further made glycans profiling: Future Hope of Glycomics

Concise investigations of transcripts and proteins, the key of genomics and proteomics, have enhanced our knowledge of cell's biochemical machinery. The close relation between cellular metabolism and transcription of genes help to understand the biological complexity of each system. Sequencing and accessibility of whole genome offer blueprint of biochemical makeup of individual. Extensive efforts have been made to relate the genome with its proteome. Murrell et al. further made efforts to relate genome and proteome with the complete repertoire of glycans of an organism to understand glycomic approach [39]. Considering inventory efforts in the field of glycomics and glycoproteomics, new high-throughput techniques have been evolving. Glycans profiling approach can be implicated at several hierarchical levels of organism's structural database content. At first hierarchical level, the aim is to identify the glycoproteins bearing monosaccharide that shows certain modifications with change in physiological homeostasis. A convenient and complete approach to profiling numerous cytosolic and nuclear proteins incorporates the bio orthogonal strategy. A recent work also showed identification of β-O-GlcNAc modification with affinity capture and MS. Another identical profiling study recognized sialylated or mucin type O-linked glycoproteins with the help of azidosugars as substrate. The higher level of profiling includes the identification of specific glycan linkage patterns. Subtle alterations in the monosaccharide units and linkage may have profound metabolic consequences, which further may relate to the development of complications and pathogenesis of diseases. The range of glycans profiling is aided by antibodies and lectins that may use in microarrays along with fluorescently labeled samples to provide detailed information of bound specific class of glycans in the glycoconjugate. Lectin and antibody approach for determination of glycans in the glyco-conjugates. Lectin and antibody approach for determination of glycan profiling, however, limitations of glycan epitopes binding with their complementary antibodies or lectins. Thus, glycan profiling can be effectively being performed with mass spectrometry [40]. Considerable expertise in profiling of glycans is required. Microchips labelled with synthetic glycan are used for high throughput profiling of mammalian cells [41].

Significance of Glycoproteomics Biomarker Expression Library

Glycoproteomic approach for generating expression functional proteome of all glycoproteins in healthy and subjects with diabetes will deliver the database for further analysis of the glycoproteins. The library database will enhance the ability to understand the mechanistic pathway that primarily involves in the complications arises due to these glycoproteins. These libraries will also help to discover a novel therapeutic approach for prevention of diabetes associated micro- and macro-vascular complications.

Another important significance is the discovery of next generation future diagnostic marker of diabetes. Several glycoproteins alter in its structural and functional proteome whose properties can be exploited for diagnosing chronic disease. Currently, investigations of diabetes diagnosis limit up to HbA1c and fructosamine only, but due to several limitations and shortcomings, these parameters are not fulfilling the scope completely. Apart from these conventional markers, the glycoproteins may be the possible next generation diagnosis marker of diabetes for deciphering the precise and accurate investigations in diabetes.

Conclusion

Glycoproteomics is an important part of proteomics as glycosylation of proteins reveals biological impairments in disease conditions. Extensive research is needed to study glycoproteins and glycans structure and expression altered pattern in diabetes and its associated complication for generation of new diagnostic future marker. Incorporation of various identifying and characterizing approach based on mass spectroscopy coupled with chromatography may enhance the precise of the biomarkers. Still, there is an extensive data processing is needed for studying glycans and its related database with bioinformatics coupled tool with algorithms. These glycoproteins expression pattern in healthy and diabetes patients will add knowledge to the international database so that one can exploit these libraries for proposing a novel therapeutic approach in diabetes mellitus. The research of glycoproteomics is never ending and technically challenged yet very promising to discover novel next generation future diagnostic markers for early diagnosis of diabetes and its associated complications.

Conflict of Interest

The authors declare no conflict of interest.
References


