Global Mass Spectrometry and Transcriptomics Array Based Drug Profiling Provides Novel Insight into Glucosamine Induced Endoplasmic Reticulum Stress*

Ana Sofia Carvalho‡§, Helena Ribeiro‡, Paula Voabil‡, Deborah Penque§, Ole N. Jensen¶, Henrik Molina¶, and Rune Matthiesen‡§**

We investigated the molecular effects of glucosamine supplements, a popular and safe alternative to nonsteroidal anti-inflammatory drugs, for decreasing pain, inflammation, and maintaining healthy joints. Numerous studies have reported an array of molecular effects after glucosamine treatment. We questioned whether the differences in the effects observed in previous studies were associated with the focus on a specific subproteome or with the use of specific cell lines or tissues. To address this question, global mass spectrometry- and transcription array-based glucosamine drug profiling was performed on malignant cell lines from different stages of lymphocyte development. We combined global label-free MS-based protein quantitation with an open search for modifications to obtain the best possible proteome coverage. Our data were largely consistent with previous studies in a variety of cellular models. We mainly observed glucosamine induced O-GlcNAcylation/O-GalNAcylation (O-HexNAcylation); however, we also observed global and local changes in acetylation, methylation, and phosphorylation. For example, our data provides two additional examples of “yin-yang” between phosphorylation and O-HexNAcylation. Furthermore, we mapped novel O-HexNAc sites on GLU2B and calnexin. GLU2B and calnexin are known to be located in the endoplasmic reticulum (ER) and involved in protein folding and quality control. The O-HexNAc sites were regulated by glucosamine treatment and correlated with the up-regulation of the ER stress marker GRP78. The occupancy of O-HexNAc on GLU2B and calnexin sites differed between the cytosolic and nuclear fractions with a higher occupancy in the cytosolic fraction. Based on our data we propose the hypothesis that O-HexNAc either inactivates calnexin and/or targets it to the cytosolic fraction. Further, we hypothesize that O-HexNAcylation induced by glucosamine treatment enhances protein trafficking. *Molecular & Cellular Proteomics 13: 10.1074/mcp.M113.034363, 3294–3307, 2014.

Glucosamine (GlcN)1 induces the posttranslational modification O-linked β-N-acetylglucosamine (O-GlcNAc) on serines and threonines (2, 3). Studies have demonstrated that GlcN treatment can lead to glucose intolerance and the death of pancreatic β-cells. For example, O-GlcNAc plays a role in the chronic complications of diabetes mellitus and insulin resistance (4, 5). The clinical relevance of GlcN, a dietary supplement used in osteoarthritis patients, is unclear (6). Previous investigations of long-term GlcN administration did not reveal risks or concerns (7), which is also consistent with recent studies (8). However, in recent clinical studies, GlcN oral administration (8) did not demonstrate any benefits, and the concentration used in vitro was not comparable with the levels observed in the plasma in vivo after the oral administration of GlcN, which raised skepticism (8, 9). Therefore,

1 The abbreviations used are: GlcN, Glucosamine; API5, apoptosis inhibitor 5; CANX, calnexin; CASP2, caspase-2; DEHP, Bis (2-Ethylhexyl)(Phthalate); FDR, false discovery rate; GEO, Gene Expression Omnibus; GLU2B, glucosidase 2 subunit beta; M, log ratios; MMP12, matrix metalloproteinase 12; mzXIC, mass, charge and isotope-dependent extracted ion chromatograms; O-GlcNAc, O-linked β-N-acetylglucosamine; OGA, O-GlcNAcase; OGT, O-GlcNAc transferase; PDMs, polycyclodi-methylsiloxane; PI, propidium iodide; PKC, protein kinase C; PLB, Phospholamban; PSMs, peptide spectral matches; qPCR, quantitative polymerase chain reaction; RIPA, radio-immune precipitation assay; RMA, robust multichip average; SERCA, sarco/endoplasmic reticulum calcium ATPase; UDP-GalNAz, Uridine 5′-diphospho-N-acidoacetylgalactosamine.
several studies have suggested caution for the use of GlcN in the treatment of osteoarthritis and other autoimmune diseases. In contrast, other studies claim that in addition to providing pain relief for osteoarthritis patients, GlcN is beneficial in ischemia/reperfusion injuries. For example, GlcN cardioprotection has been observed in perfused rat hearts (10), demonstrating the feasibility of obtaining an in vivo response to GlcN.

GlcN exhibits anticancer properties in vitro (11, 12), and a recent review by Slawson et al. (13) suggested several molecular mechanisms through which O-GlcNAcylation can play a regulatory role in cancer. In the case of cancer other administration strategies than oral intake are possible. Furthermore, the negative side effect from high drug dosages is more acceptable in the case of terminal cancers.

In general, elevated O-GlcNAc levels elicited by GlcN administration, utilizing in vitro cell lines and animal models, exhibit ER stress (14), changes in calcium signaling (15), modified transcriptional activity (16), and alterations in phospho-signaling cascades (17). Additionally, GlcN causes cell cycle arrest in G1 and/or G2 (11) accompanied by apoptosis (11). More recently, GlcN treatment has been shown to increase hyaluronan synthesis (18) and autophagy (19, 20). Autophagy is considered important in cancer because depending on the drug type, autophagy can either increase sensitivity to a drug or protect cancer cells from a drug (21).

Understanding the molecular mechanisms and pathways activated following GlcN intake could lead to the discovery of novel targets for the treatment of cardiovascular diseases, diabetes, and cancer. LC-MS-based proteomics for drug profiling demonstrates great potential for the elucidation of molecular mechanisms and provides detailed information about how specific cell types compensate for or become resistant to drugs. In this study, we combined, label-free MS-based quantitation with an open search strategy for modifications to obtain deepest possible proteome coverage in drug profiling experiments given the obtained Q-Exactive data.

To date, there is only indirect evidence to explain how GlcN affects calcium signaling and causes ER stress. PLB (phospholamban), a regulator of SERCA (sarco/endoplasmic reticulum calcium ATPase), contains an O-GlcNAc site; however, the regulation of this protein requires further investigation (22, 23). The link between GlcN treatment and ER stress might stem from the O-GlcNAc modification of heat shock proteins; however, this hypothesis needs experimental validation (24).

In search of explanations for the above observations, we investigated GlcN-treated malignant cells from different stages of lymphocyte development using mass spectrometry (Q-Exactive) and transcriptional arrays (Affymetrix). We observed the expected phenotypic mRNA and protein changes after GlcN treatment, such as ER stress, cytoplasmic vacuolization, altered regulation of proteins with immunological functions, and apoptosis. We identified novel O-HexNAc sites induced by GlcN treatment, which suggest a link between GlcN treatment and ER stress. Finally, the global up-regulation of O-HexNAcylation caused the expected global down-regulation of phosphorylation levels but also affected methylation and acetylation levels.

**Experimental Procedures**

**Short Summary of Methods—O-GlcNAc levels** (probed with antibodies CTD 110.6 and RL2) in four malignant hematopoietic cell lines, KMH2, RAMOS, HDML2, and Jurkat, before and after GlcN (Sigma) treatment, were assayed by Western blot over time and at different concentrations of GlcN (0.1, 2.5, 10, and 20 mM). The viability of the cell lines could be re-established by removing GlcN for all the tested conditions (Supplemental Fig. S1). The largest increment in the level of O-GlcNAcylation of proteins was observed in KMH2 cells following treatment with 20 mM GlcN for 24 h. The cell lines were characterized with respect to cell survival, cell cycle arrest, and apoptosis. KMH2 cell line was treated with different concentrations of GlcN and was assayed by blotting with anti-O-GlcNAc antibody at different times. The time point and concentration exhibiting the maximum O-GlcNAc response was globally profiled and compared with the untreated cells with respect to mRNA (microarray) and protein expression (MS-based proteomics, Q-Exactive). The quantitative proteomics results were validated using Western blots and three different MS-based protein quantitation techniques (Tandem Mass Tag labeling (TMT), stable-isotope dimethyl labeling, and label-free quantitation). KMH2 cells were fractionated into nuclear, mitochondrial, and cytosolic crude fractions. The main aim of the fractionation was to obtain a deeper sampling rather than determining the subcellular localization of proteins.

**Cell lines and Culture Conditions**—The human Hodgkin Lymphoma derived cell lines HDM-2 and KMH2 and the T cell leukemia-derived cell line Jurkat and Burkitt lymphoma-derived cell line RAMOS were obtained from the German Collection of Microorganisms and Cell Cultures, Department of Human and Animal Cell Cultures. All cell lines were cultured in Invitrogen RPMI medium 1640 Glutamax™ (Invitrogen, Carlsbad, CA) supplemented with 10% heat-inactivated FBS (Invitrogen) in a humid environment of 5% CO2 at 37 °C. For GlcN treatment cells were cultured for 24 h and replated at 5 × 10^4 cells/ml with or without GlcN at 20 mM for 6, 12, 24, and 48 h. For concentration effect cells were plated at 5 × 10^4 cells/ml with GlcN at 0.1, 2.5, 10, and 20 mM for 24 h.

For Bortezomib treatment, KMH2 cells were seeded into six well plates at a density of 5 × 10^5 cells/ml per well (2 ml) and treated with the following: vehicle control (DMSO), Bortezomib (PS-341) (Selleck Chemicals, US) at 1, 5, 10, and 100 nM and GlcN at 20 mM. The cells were treated for 24 h.

**Analytical and Biological Reliability—**All mRNA expression analysis such as qPCR and expression arrays was performed with three biological replicas. Western blots, cell-cycle, and apoptosis assays were performed with minimum three biological replicas. Each of the quantitative MS methods used were done with three technical replicas on independent biological samples.

**Cell-cycle and Apoptosis Assays—**Apoptosis was determined by annexin V-FITC and propidium iodide (PI) double staining according to the manufacturer’s instructions (BD Biosciences). Cell-cycle fractions were determined by propidium iodide nuclear staining. Briefly, cells were harvested, washed in PBS, fixed in 70% ethanol overnight at 4 °C, and incubated in propidium iodide solution (10 μg/ml propidium iodide, 0.1 mg/ml RNase A in PBS-Tween 20 (0.1% v/v) for 30 min at 37 °C). Data were collected on a FACS Canto II flow cytometer (BD Biosciences) and analyzed with FlowJo Version 7.6.5 software (TreeStar, US). Results represent the mean value of three independent experiments.
RNA Isolation, Reverse Transcription, and Quantitative PCR—Total RNA was isolated from the various cell lines at different conditions using the RNaseasy Plus Mini Kit (Qiagen, Stanford, CA) according to the manufacturer’s protocol. RNA yield and quality were determined spectrophotometrically.

Reverse transcription was performed using 1 μg total RNA, random oligonucleotides primers, and SuperScript II RT (Invitrogen) in a total volume of 20 μl as described by the manufacturer. For real-time PCR analysis, cDNA samples were diluted 10-fold with water and PCR amplified in triplicate using TaqMan® Gene Expression Assays (Hs00269228_m1-OGT,Hs00201970_m1-MGEAS), Hs03928985_g1-RN18S1, Hs01003267_m1-HRP1T in an ABI Prism 7500 Sequence Detection System (Applied Biosystems, Foster City, CA) according to manufacturer’s protocol. Expression of 185 and HRPT1 was used for normalization of target gene abundance.

Microarray Data Acquisition and Analysis—RNA expression profiling was performed using the Affymetrix GeneChip® technology, following the protocols recommended by the manufacturers. mRNA expression in the cell line RAMOS and three biological replicates of KMH2 were analyzed, before and after 24 h of GlcN treatment, by Affymetrix arrays (HuGENE-1_1-st-v1, Probe set annotation, release 32, 9/30/11). The data were collectively analyzed by using the R package “AFFYLMGUI” (http://www.bioconductor.org) (25). Background adjustment was done by using robust multichip average (RMA) (26). Correction for multiple testing was done by the method of Benjamini and Hochberg (27). Microarray data is publicly available at http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?&acc=GSE49185, Gene Expression Omnibus (GEO) accession number GSE49185.

Cell Fractionation—Cells were homogenized in ice-cold cell homogenization medium (10 mM Tris, pH 6.7, 150 mM MgCl₂, and 10 mM KC1) by passing through a 20G syringe. Cell breakage was examined under a phase-contrast microscope. Nuclei were pelleted by centrifuging for 5 min at 1000 × g at 4 °C after addition of cell homogenization medium containing 1 mM sucrose (final concentration 250 mM). Mitochondria were isolated by centrifuging the remaining supernatant for 10 min at 5000 × g at 4 °C and resuspending the pellet in ice-cold sucrose/ Mg²⁺ medium (10 mM Tris, pH 6.7, 150 mM MgCl₂, and 0.25 mM sucrose). The mitochondrial fraction is obtained by recentrifuging the suspension at 5000 × g for 10 min at 4 °C. The supernatant constituted the cytosolic fraction. All samples were stored at −80 °C until use.

SDS-PAGE and Western blot—Nuclei and mitochondria pellet obtained as described above and total cells were lysed using RIPA lysis buffer at 4 °C for 20 min and the lysate cleared by centrifuging at 20 min at 15,000 × g at 4 °C. Cytosolic fraction was analyzed without further processing. Samples were loaded into 10% acrylamide SDS-page gels and transferred overnight to PVDF membranes. Membranes were blocked and incubated with primary antibody and secondary HRP labeled antibody (Santa Cruz Biotechnology, Inc., Santa Cruz, CA) according to manufacturer’s recommendations. ECL-prime (GE Healthcare) was used for detection.

Immunoprecipitation—After 24 h incubation cells were treated with 20 μg GlcN for 24 h and then lysed with RIPA buffer. Dynabeads Protein G (50 μl of a 50% slurry, Invitrogen) were mixed with anti-calnexin, antibody (Thermo Scientific) or control mouse IgG and rotated for 1 h at room temperature. Lysates containing 800 μg proteins were added to the antibody-conjugated dynabeads and rotated overnight at 4 °C. The immunoprecipitates were collected, washed, and eluted before standard Western blot analysis using either anti-O-GlcNac (both CTD110.6 and RL2 were used) or calnexin antibodies.

In Vitro O-GlcNac Labeling Assay—Immunoprecipitated calnexin from treated and nontreated cells was labeled with mutated GalT(Y289L) and UDP-GalNAz using the Click-IT O-GlcNac enzymatic labeling kit, and biotin alkaline from the glycoprotein detection kit (Invitrogen) was used for detection by Western blot.

Peptide Sample Preparation—Protein solution containing SDS and DTT are loaded into filtering columns and washed exhaustively with 8 M urea in HEPES buffer. Proteins are then incubated overnight with trypsin sequencing grade (Promega, Madison, WI) after alkylation with iodoacetamide and reduction with DTT.

Chemical Labeling of Peptides with Stable Isotopes—For stable-isotope dimethyl labeling and Tandem Mass Tag labeling two portions of ~30 μg peptides from each of the two samples were prepared in 100 μM TEAB. For the stable-isotope dimethyl labeling experiment published protocols were followed (28). “Control Sample” was labeled with (CH₃)₂ and “Treated Sample” was labeled with (13C₂H₃)₂. After labeling the two samples were mixed.

For the Tandem Mass Tag labeling (TMT) labeling experiment a 6-plex TMT kit (cat# 90066, ThermoFisher, IL) was used. Following the manufacturer’s protocol “Control Sample” and “Treated Sample” was labeled with the tandem mass tags 126 and 127, respectively.

Both labeling experiments were desalted, concentrated and analyzed. Proper labeling was verified by LC-MS/MS. Hereafter the dimethyl and TMT samples were subjected to optional strong cation exchange fractionation and STAGE tip cleanup. Seven fractions, including flow through, were collected for each stable isotope labeled experiment (29).

Mass Spectrometry—Peptides generated as described above were desalted and concentrated (30) prior to analysis by nano LC-MS/MS using a Q-Exactive (Thermo, San Jose, CA) mass spectrometer coupled to a Dionex NCP3200RS HPLC setup (Thermo, Sunnyvale, CA). A 75 μm ID, 15 cm in length home build reversed phase column (Reprosil-pur C18 AQ 3 μm, Ammerbuch-Entringen, Germany) was used to separate peptides. The analytical gradient was generated at 200 nL/min increasing from 5% Buffer B (0.1% formic acid in acetonitrile)/95% Buffer A (0.1% formic acid) to 35% Buffer B/65% Buffer A over 110 min followed by an increase to 90% Buffer B/10% Buffer in 10 min. MS survey scans were scanned from m/z 350 to m/z 1400 at 70,000 resolution (AGC: 1e6 and Maximum IT: 120 ms). An upper limit of 20 most abundant ions was subjected to MS/MS and measured at a resolution of 35,000 (AGC: 5e4 and Maximum IT: 120 ms) with lowest mass set to m/z 100. All data are available in online repositories and supplemental files.

Preprocessing of MS Data—All Q-Exactive data were calibrated using polycyclo-dimethylsiloxane (PCMs—outgassed material from semiconductors) present in the ambient air and Bis(2-Ethylhexyl) (Phthalate) (DEHP—from plastic) (31, 32) using both MaxQuant version 1.3.0.5 (33) and modular VEMS, mVEMS v1.0 (34) (supplemental Fig. S2). mVEMS further allows alternative parent ion annotations for each MS/MS spectrum that is needed if two peptide elution profiles overlap in the m/z and retention time dimension. By allowing alternative parent ion annotation for each MS/MS spectrum, provides a space efficient data format. Furthermore these alternative parent ion annotations were taken into account during the database dependent search of MSMS data.

Database Dependent Search of MS Data—All data were searched with mVEMS (35) and MaxQuant version 1.3.0.5 (33). Mass accuracy was set to 10 ppm for peptides and 10 Da for peptide fragments. Four missed cleavages were specified and the database UniProtKB/TrEMBL (Release 2013_02) were used including permutated protein repositories and supplemental files. In mVEMS a search of MSMS data.

Database Dependent Search of MS Data—All data were searched with mVEMS (35) and MaxQuant version 1.3.0.5 (33). Mass accuracy was set to 10 ppm for peptides and 10 Da for peptide fragments. Four missed cleavages were specified and the database UniProtKB/TrEMBL (Release 2013_02) were used including permutated protein repositories and supplemental files. In mVEMS a search of MSMS data.
specified for mVEMS searches because mVEMS by default checks N-terminal Met-loss. The modifications Phosphate-ribosylation (UNIMOD 1356), ADP-ribose (UNIMOD 213), Hydroxylation (UNIMOD 35), and Myristoylation (UNIMOD 45) were not frequent at a 1% false discovery rate (FDR) threshold. We conclude that these few counts are likely to be false positives and the search was therefore repeated excluding these modifications (see supplemental Table S1 for the final list of modifications that was used for the analysis). For stable-isotope dimethyl labeled data the additional variable modifications were obtained by mzXIC (mass, charge, and isotope-dependent extracted ion chromatograms) (34) profiles extracted from label free, stable-isotope dimethyl labeling and, and TMT LC-MS data and used for peptide quantitation. Proteins were quantified by spectral counting (38). To determine significant changes in protein expression or modification upon glucosamine treatment, ANOVA tests on log transformed quantitative values were employed using the R package DanteR (39). To determine significant changes in protein expression or modification upon glucosamine treatment, ANOVA tests on log transformed quantitative values were employed using the R package DanteR (39), p values were corrected for multiple testing by the method of Benjamin and Hochberg (BH) (27) and no imputation for missing values were used.

Overall changes of PTMs in subcellular fractions were estimated by summing all PSMs for specific modifications and fraction before and after GlcN treatment. The relative change was calculated as: 

\[
\frac{\#\text{PSMs}_{\text{GlcN}}}{\#\text{PSMs}_{\text{Control}}} - 1
\]

Venn diagrams were made with R package “VennDiagram.” To access the level of protein trafficking enrichment analysis of proteins position in the Venn diagram was calculated using the R function dhyper. Heatmaps were generated by using the R package “heatmap.3.” Significant regulated protein isoforms were collapsed to the genes that encode them to facilitate visualization of regulated proteins.

RESULTS

Lymphocyte-derived Cell Lines Respond Differently to GlcN—To define the effects of GlcN at the molecular level, we reasoned that a strong increase in the O-GlcNAcylation of proteins would be needed. Given that the O-GlcNAcylation of proteins is abundant and plays an important role in the differentiation of bone marrow-derived cells, we investigated four malignant cell lines from different stages of lymphocyte development for their response to GlcN treatment: KMH2, RAMOS, Jurkat, and HDML2. KMH2 cells exhibited the strongest response in terms of the O-GlcNAcylation of proteins (supplemental Fig. S1A). The optimal GlcN concentration and time point in terms of an increase in O-GlcNAcylated protein levels for KMH2 cell line was 20 mM and 24 h of treatment (supplemental Fig. 1C).

Experimental Outline of the Global GlcN Profiling of Proteins and mRNA—To investigate the mechanism of the GlcN effects, we examined GlcN-treated malignant cells from different stages of lymphocyte development using mass spectrometry and transcriptional arrays. Total mRNA extracts were used for transcriptional arrays. For quantitative proteome analysis cell fractionation was used to identify localization-dependent responses and to obtain deeper sampling. Nuclear, mitochondrial, and cytosolic fractions were analyzed using a label-free MS approach. The nuclear fraction from nontreated and GlcN-treated cells were analyzed using alternative quantitative methods, such as isobaric tandem mass tags (TMT) and stable-isotope dimethyl labeling providing additional validation for the quantitative findings. The label-free MS data for the nuclear proteins from each condition were analyzed using an open search approach to avoid cumbersome enrichment protocols and to profile several posttranslational modifications simultaneously in a single search.

We used tandem mass tags (TMT), stable-isotope dimethyl labeling and label-free MS-based quantitative methods to analyze the nuclear fraction before and after challenging KMH2 cells with GlcN (supplemental Fig. S3). Label-free quantitation based on spectral counts and ion counts in survey scans provided the deepest profiling (supplemental Fig. S3). The label-free quantitation identified 3397 proteins at 1% FDR threshold after collapsing the proteins to the encoded genes in the nuclear fraction. The label-free approach identified all proteins previously found by alternative quantitative approaches and identified ~1000 additional proteins. At the peptide level, the label-free method resulted in the identification of twice as many peptide spectral matches (PSMs) per protein, leading to more reliable protein quantitation. High protein sequence coverage generates improved protein quantitation because of problems caused by modifications and partial tryptic cleavages. Therefore, we continued using label-free quantitation for the mitochondrial and cytosolic fractions. However, the TMT and stable-isotope dimethyl labeling methods are still used to validate the differential regulation observed from the label free data. The database-dependent search of the MS/MS spectra from the nuclear, mitochondrial, and cytosolic crude fractions resulted in a set of 18,816 proteins using a 1% FDR cutoff. Of these, 17,745 proteins were placed in evidence groups one to three as defined by Matthiesen et al. (37). Collapsing the proteins into the corresponding coding genes yielded 5181 proteins, 5065 of which were not common contaminants.

GlcN Increases the O-GlcNAcylation of Primarily Cyttoplasmic and Nuclear Proteins—Western blotting of total cell lysate, nuclear, mitochondrial, and cytosolic crude fractions demonstrated abundant O-GlcNAcylation of proteins in the cytosolic and nuclear fractions. Furthermore, the strongest increase in O-GlcNAcylation after the GlcN treatment occurred in the nuclear and cytosolic fractions (Fig. 1A). Indeed, the Western blot analysis using the mitochondrial fractions purified by differential centrifugation were negative for lysosome and peroxisome markers but positive for mitochondrial markers and demonstrated no detectable level of O-GlcNAc-
lated proteins in KMH2 cells (data not shown). Furthermore, OGA (O-GlcNAcase) was not detected in the crude mitochondrial fraction using MS and Western blot analysis (Fig. 1A).

According to the Western blot analysis, O-GlcNAc transferase (OGT) was down-regulated and OGA was up-regulated in all fractions at the protein level after the GlcN treatment (Fig. 1A and supplemental Fig. S1B). Based on the MS data OGT was detected and significantly down-regulated (pBH value < 0.05, BH: Benjamini and Hochberg correction) in the cytosolic (more than fourfold), and nuclear fraction (more than 3-fold) and OGA was only detected significantly up-regulated (more than 9-fold) by MS in the cytosolic fraction thereby confirming the Western blots in cases where OGT and OGA were identified by MS (Fig. 2A–2C). OGT and OGA were not detected in mitochondrial fraction by MS. The microarray data showed mRNA up-regulation of OGT with a ratio of 1.14 and a marginal significant p value, whereas OGA (MGEA5) was significantly up-regulated with small effect size (ratio of 1.35) at the mRNA level that was confirmed by qPCR (Fig. 1B, t test followed by Benjamini and Hochberg correction).

Global Regulation of Proteins, PTMs, and mRNA Expression in Cells Challenged with GlcN—The combined MS-based quantitative proteomics and transcriptomics data allowed the calculation of the ratios between untreated and treated cells, the p values and the abundance values for the mRNA and proteins in the nuclear, mitochondrial, and cytosolic fractions. Fig. 2A–2C summarizes Z normalized spectral count data for the significant changed proteins observed in the three fractions and correlated the proteomics data with the observed mRNA log ratios (M values). The data presented in Fig. 2A–2C were filtered to include proteins that are not common contaminants seen by MS and with evidence groups 1 to 3. This means that proteins not required to explain the MS data, that is, evidence groups 4 to 5, were filtered out and only peptide spectra assignments with maximum scores were taken into account for protein inference (37). Furthermore, to provide a more compact view and because we have incomplete knowledge of how specific protein isoforms function, the protein isoforms were collapsed into the corresponding coding genes for the presented heatmaps. Most of the significantly regulated proteins were in the cytosolic fraction (Fig. 2A–2C). The percentage of significantly regulated proteins that were up-regulated was considerably higher for the mitochondrial fraction (71% compared to 40 and 45% for the nuclear and cytosolic fractions, respectively).

To identify PTM regulation, statistical analysis of mzIXIC (mass, charge, and isotope-dependent extracted ion chromatograms) values were performed as previously described (34) (Fig. 2D–2F). Based on the mzIXIC values, a large number of unmodified peptides were observed significantly regulated; however, only significantly regulated peptides exhibiting lysine acetylation, lysine methylation, phosphorylation, or O-HexNAcylation modifications are shown in Fig. 2D–2F (after BH correction of p values). Figures of all raw spectra annotations of modified and significant regulated peptides are provided in supplementary file regulated.pdf. Annotation of raw spectra from all modified peptides and extracted mzIXIC values are provided in supplementary file PTMpep.zip as obtained directly from the VEMS modules v1.0. File HexNAC.zip
FIG. 2. Global protein and PTM regulation observed in the three subcellular fractions and corresponding transcriptional regulation. Z normalized spectral count data for significant regulated proteins in A, nuclear, B, cytosolic, and C, mitochondrial fractions. Z normalized mziXIC values of significantly regulated PTMs in D, nuclear, E, cytosolic, and F, mitochondrial fractions. “C” indicates control and “T” treated. “SC” is the range of total number of spectral counts. “M” is the log ratio for mRNA expression. “E” is the evidence group: For a given protein “E/H11005^1” means that the protein has at least one peptide spectra assignment matching uniquely, “E/H11005^2” means that the protein belongs to a group of proteins that cannot be distinguished based on the MS data and have at least one peptide spectra assignment matching uniquely to the group, and “E/H11005^3” means a protein that belongs to a group of proteins that share peptide spectra assignments of which none belongs to evidence groups one or two.
contains raw spectra annotation and quantitative values of all O-HexNAc peptides with ambiguous modifications sites specified. It is unrealistic to determine sites of O-HexNAc based on HCD data because of the intense neutral loss of O-HexNAc.

**GlcN Causes ER Stress and Vacuolization Affecting Cell Cycle Progression**—GlcN caused an increase in sub-G1 KMH2 population, G1 RAMOS population and G2/M KMH2 and RAMOS populations (Fig. 3A). KMH2 cells doubling time are longer than RAMOS cells that can account for the smallest decrease in S phase population for KMH2 compared with RAMOS and no significant increase in G0/G1 population for KMH2. GlcN hindered KMH2 and RAMOS cell cycle progression and furthermore a 10% decreased in cell viability was observed for KMH2 compared to 6% for RAMOS cells (data not shown). Increased apoptosis was observed in KMH2 compared with the RAMOS cells, and in both cases, the level of apoptosis could not fully explain the decrease in viable cells (Fig. 3B). In conclusion, the inhibition of cell progression together with apoptosis caused an overall decrease in total number of viable cells upon GlcN treatment for KMH2 and RAMOS. KMH2 (Fig. 3C) and RAMOS (data not shown) cells treated with GlcN accumulated vacuoles in the cytoplasm. The GlcN treatment caused an approximately twofold up-regulation of SQSTM1 (a protein described to interact with MAP1 LC3 family members) in the three subcellular fractions based on spectral counting (Fig. 3D). SQSTM1 in the nuclear fraction (see Fig. 2) and its transcriptional level (Fig. 3D) appeared to be significantly regulated ($p < 0.05$, t test followed by Benjamini and Hochberg).

**Fig. 3.** GlcN causes cell cycle arrest, apoptosis, ER stress and increased autophagy. A, Cell cycle phase distribution for KMH2 and RAMOS cells (black: sub-G1, dark gray: G0/G1, white: S-phase, and light gray: G2/M). B, Apoptosis assay for KMH2 and RAMOS cells. C, Increased vacuolization of KMH2 cells upon GlcN treatment. D, SQSTM1 was up-regulated after GlcN treatment (green depicts untreated and red depicts after the GlcN treatment) at both the (a) protein and (b) mRNA level after GlcN treatment. E, GRP78/HSPA5 was up-regulated following the GlcN treatment (green depicts untreated and red depicts after the GlcN treatment) at both the (a) protein and (b) mRNA level after GlcN treatment.
GlcN also led to the up-regulation of GRP78/HSPA5 (a marker for ER stress) at both the protein and mRNA level (Fig. 3E) based on a t test followed by Benjamini and Hochberg correction \((p < 0.05)\).

**GlcN Increases the O-HexNAcylation of Calnexin and Decreases Calnexin Phosphorylation**—The MS data presented here is unable to distinguish between O-GlcNAC and O-GalNAC and therefore “O-HexNAc” is used for MS spectra annotation. The analysis described in Fig. 2D–E2 suggested the novel regulation of O-HexNAc on calnexin. The O-HexNAc-modified peptide “APVPTGEVFADSFDR” mapped only to calnexin isoforms encoded by a single gene, CANX (supplemental Fig. S4). The label-free (Fig. 4A and supplemental Fig. S5), TMT, and stable-isotope dimethyl labeling data (supplemental Fig. S6A and S6B) confidently identified both the unmodified and O-HexNAc-modified peptide. Annotation of the raw spectrum pinpointed the O-HexNAc site to be Thr-66 or Ser-74 in calnexin (UniProtKB P27824, Version 153 and supplemental Fig. S4) with approximately equal probability for the two sites. GalNAc-O-linked as a single residue in the ER has not been previously reported in nonengineered cells because it is normally extended. However, a nonextended GalNAc (the authors assume O-GalNAc because of the ER location of calnexin) at Thr-66 has been described in engineered cells in which the core extending glycosyltransferase responsible for “mucin-type” (GalNAc-type) O-linked glycosylation or its essential chaperone has been knocked out (40). Compared with O-GalNAc, GlcN is a more direct precursor for O-GlcNAC. Furthermore, Sakaidani and coworkers recently reported an alternative O-GlcNAC transferase (EGOT) that localizes to the lumen of the endoplasmic reticulum and transfers GlcNAC to epidermal growth factor-like domains in an OGT-independent manner (41). However, to provide further evidence for the O-GlcNAC modification of calnexin, calnexin IP (immunoprecipitation) from KM92 cell extracts was performed before and after the GlcN treatment (supplemental Fig. S7A and S7B), followed by Gal-T1 and Click-IT™ Biotin Alkyne Detection (Fig. 4B). The enzyme Gal-T1 is described to be specific for O-GlcNAC, and an increased Biotin Alkyne Detection signal was observed in the treated cells. The negative controls lacking Gal-T1 were blank, and the positive control using alpha-crystallin was positive (Fig. 4B). Positive detection of O-GlcNAC on Western blot of calnexin IP with RL2 and CTD110.6 could not be confirmed (supplemental Fig. S7B). This does, however, not exclude the possibility that calnexin is O-GlcNAC modified because antibodies targeting glycan modified peptides may partially recognize part of the protein as well.

Ser-74 in calnexin is required for Ca\(^{2+}\) binding. The Ca\(^{2+}\)-binding residues are highly conserved, and calnexin undergoes Ca\(^{2+}\)-dependent conformational changes (42). The peptide “APVPTGEVFADSFDR HexNAc” in calnexin was up-regulated more than 73-fold compared with the control (supplemental Fig. S8), providing a potential explanation for how GlcN can cause ER stress directly (Fig. 4D).

GlcN decreased the calnexin phosphorylation from 26% to 8% (ratio of modified/unmodified protein in the nuclear fraction, supplemental Fig. S9 and S10) on Ser-583. The modification of Ser-583 is known to regulate calnexin activity levels during the unfolding protein response (43) and is a potential target of PKC (protein kinase C) (44, 45). The ratio between the unmodified and the O-HexNAc-modified “APVPTGEVFADSFDR” form of calnexin changed from ~0% to ~12% in the nuclear fraction after the GlcN treatment (Fig. 4D). Interestingly, the ratio between the unmodified and O-HexNAc-modified “APVPTGEVFADSFDR” form of calnexin changed from ~0% to ~37% in the cytosolic fraction (Fig. 4D).

The ratio between unmodified and O-HexNAc-modified “SEALPTDLPAPsAPDLTEPK” forms of GLU2B (glucosidase 2 subunit beta), an ER heterodimeric enzyme, increased from 38 to 177% (increase of ~fivefold compared with the canonical peptide) in the nuclear fraction after GlcN treatment (supplemental Fig. S11 and Fig. 5D). In the cytosolic fraction, the ratio increased from 32% to 118%, which represents a ~fourfold increase compared with the canonical peptide.

**Link between GlcN Metabolism and Posttranslational Modifications**—The GlcN treatment lead to alterations in acetylation, methylation, and phosphorylation in addition to O-HexNAcylation, as evaluated by the relative change in the number of assigned PSMs on a 1% FDR threshold (Fig. 5). The largest change in O-HexNAcylation in terms of the assigned PSMs occurred in the cytosolic fraction (Fig. 5A). Interestingly only phosphorylation in the nucleus decreased overall following the GlcN treatment. The GlcN caused acetylation to increase in the nucleus and decrease in the cytosolic and mitochondrial fractions. Although monomethylation in the cytosolic and mitochondrial fractions and dimethylation in the mitochondrial fraction decreased, trimethylation increased in all fractions (Fig. 5A). N-terminal protein acetylation was prominent as the posttranslational modification detected most often (Fig. 5B). In this analysis, we disregarded modifications, such as deamidation and oxidation, which were more abundant but difficult to interpret because they are often the result of an artifact of sample handling.

GlcN treatment increased the levels of EF1A1 methylation in the nuclear fraction (Fig. 5C). The label-free MS data set identified the unmodified, mono-, di-, and tri-methylated versions of the peptide MDSTEPPYSQkR (lowercase indicates site of methylation). Furthermore, these peptides were both observed as double- and triple-charged. A standard statistical analysis (t test in combination with Benjamini and Hochberg correction) identified only the mono-methylated peptide as significantly down-regulated, giving the general impression that methylation decreased. However, a careful analysis of the label-free quantitative data indicated that di- and tri-methylated versions of the peptide were significantly up-regulated before the Benjamini and Hochberg correction. In addition, if the combined up-regulation of the di- and trimethylated peptides were considered in the calculation of the
**FIG. 4.** GlcN increases the O-GlcNAcylation of calnexin. A, MS/MS raw spectrum of the assigned peptide “APVPtGEVYFADsFDR”._H11001_. B, Purified calnexin from untreated and treated cells labeled using the Click-iT™ O-GlcNAc Enzymatic Labeling System. Detection was performed using the Click-iT™ Biotin Alkyne Detection Reagent (_H11001_). Purified calnexin (_vide Supplemental Fig. S7_) was subjected to the same procedure without the addition of Gal-T1 (Y289L) enzyme (–). C, GlcN increased the levels of EF1A1 methylation in the nuclear fraction. The modified serine residues are indicated by a white arrow, and the backbone carbonyl group is involved in the metal coordination of Ca²⁺. D, Regulation of ER quality control under protein misfolding conditions (schematic modified from (44)). Subcellular quantitative levels of O-HexNAc and phosphorylation on calnexin. % indicates the ratio between modified and unmodified residues (in percentage); black and yellow circles represent mannose residues, and the triangle represents a glucose residue; H, O-HexNAc; P, phosphorylation; ERAF, ER-assisted folding; ERAD, ER-associated protein degradation; Glc II, glucosidase II; UGGT, UDP-glucose:glycoprotein glucosyltransferase.
Global Glucosamine Profiling

We observed little overlap of significantly regulated proteins among the subcellular fractions; however, METAP2 was up-regulated in both the nuclear and cytosolic fraction, OG1 was down-regulated in both the nuclear and cytosolic fraction, and COPG2 was up-regulated in the nuclear and mitochondrial fractions. These results are consistent with the analysis of identified proteins in each of the subcellular fractions before and after the GlcN treatment (Fig. 5E). After the GlcN treatment, the altered proteins (1% FDR and a minimum of five peptide spectra matches per protein) were located predominantly in the parts of the Venn diagram that were unique to specific subcellular fractions (Fig. 5E). For example, 140 new proteins were observed in cytosolic fraction after GlcN treatment (p value = 2e-6, using the hypergeometric distribution). This analysis was repeated with 1% FDR and a minimum of 0, 5, and 10 peptide spectra matches per protein, and in all cases the pattern persisted. Using all the proteins identified at 1% FDR resulted in more protein overlap between fractions.

Furthermore, the analysis suggested a significant trafficking of proteins between the cytoplasm and the mitochondrion. For example, 104 proteins that were uniquely identified in the cytosolic fraction without GlcN treatment were observed in both the mitochondrial and cytosolic fraction upon GlcN treatment (p value = 1e-39 using the hypergeometric distribution). This finding is consistent with our hypothesis that O-GlcNAc is involved in protein trafficking. It will be interesting to determine in future studies whether this is a general trend for drug-affected proteomes. A number of the observed regulated proteins are not in the anticipated canonical subcellular fraction (Fig. 2A, 2B, and 2C). For example, IGFBP1 and IGBP2, known to be secreted, end up in the mitochondrial fraction after GlcN treatment. This might be the result of the gross changes in cellular architecture as indicated by the formation of vacuoles and cells undergoing apoptosis. Furthermore, subtypes of KMH2 cells are multinucleated, which potentially can influence proteins canonical subcellular location.
The identification of significantly regulated novel key protein markers obtained by MS-based drug profiling can be used to define innovative combined treatment strategies. Furthermore, these data provide novel evidence for how GlcN affects cells at the molecular level. MS-based drug profiling experiments, demonstrated in this study using GlcN, provide detailed information regarding the effects of drugs at a cost similar to mRNA expression arrays.

**Glucosamine, ER Stress, and Autophagy**—To date, no mechanisms linking GlcN to ER stress and autophagy have been described. The O-GlcNAcylation of heat shock proteins is a viable hypothesis; however, there is no direct evidence to support this hypothesis. We propose that the O-HexNAcylaton of Ser-74 in calnexin (located on the luminal side of the ER), a residue involved in the Ca\(^{2+}\) binding pocket, inactivates the protein or disturbs ion coordination (Fig. 4C). Ca\(^{2+}\) regulates the activity of calnexin, and in protein data bank (PDB) structures, Ca\(^{2+}\)-binding pockets in general exhibit a hydrophobic environment. Placing O-HexNAcylation on Ser-74 increases the hydrophilicity of the local environment. Therefore, a possible hypothesis is that Ser-74 O-HexNAc lowers calnexin’s affinity for Ca\(^{2+}\). Furthermore, the MS data support this model. The GlcN decreased calnexin phosphorylation on Ser-583 (from 26% to 8% modified/unmodified), a modification that should increase following ER stress (44), and based on the increase of the ER marker GRP78, ER stress was clearly generated. The decrease in Ser-583 phosphorylation (in the cytoplasmic domain of calnexin), together with the increased ratio of O-HexNAc-modified to unmodified calnexin in the cytosolic fraction (presumably from autophagy vesicles) compared with the nuclear fraction (presumably of ER origin), suggests that autophagy clears the impaired calnexin (with O-HexNAcylation on Ser-74) from the ER. However, O-HexNAcylation of calnexin Ser-74 or Thr-66 could be involved in regulating vesicle transport, providing an alternative hypothesis in which the O-HexNAcylaton constitutes a consequence instead of a cause of ER stress.

**GLU2B** (glucosidase 2, subunit beta), an ER luminal protein involved in protein quality control, and the ratio between the O-HexNAc-modified and unmodified GLU2B in the nuclear fraction (presumably of ER origin) increased from 38% to 177% after GlcN treatment. The GlcN down-regulated PKC (Fig. 2 and 5D-e, PRKCG) and the level of O-HexNAcylaton of GLU2B, a known substrate of PKC, increased. This observation is consistent with the “yin-yang” hypothesis for phosphorylation and O-GlcNAcylaton proposed by Hart and co-authors (53).

Previous studies have demonstrated functional roles for O-GlcNAc in the nucleus and cytoplasm. The identification of two regulated O-HexNAc (potentially O-GlcNAc) sites on proteins known to be ER residents suggests that O-GlcNAc may play a functional role in the ER in addition to O-GalNAc.

**Anticancer Properties of GlcN**—Our data indicate that the induced overload of internal UDP-GlcNAc, probably because of some cancer cells inability to regulate GlcN transporters in cell membrane, leads to saturation of O-glycans pathways in the ER. We speculate that improper and uncontrolled initiation of O-glycosylation lead to ER stress, which is partly resolved by transporting the affected proteins to internal vacuoles rather than the normal secretory pathways. Specifically targeting the proteome of vacuoles formed upon GlcN treatment by MS-based methods could resolve this issue in the future. We believe this is important because one possibility is that increased short O-glycans on the cell surface potentially can lead to increase migration potential and invasiveness of the cancer cells (54). Cell migration assays could address this question.

Other issue to consider in terms of anticancer properties is the observation that KMH2 cells become more resistant to Bortezomib treatment (Supplemental Fig. S14), if combined with GlcN treatment. We observe that GlcN has an additive effect in terms of causing cell death at low Bortezomib concentration. However, at higher Bortezomib concentrations GlcN protects the cells. This suggests that a clinical study on the combined effect of bortezomib treatment and GlcN intake or diet is relevant.

Analyzing the global regulation of mRNA and proteins in KMH2 and RAMOS cell lines shows a clear trend that mitochondrial genes and proteins are in general down-regulated (supplemental Fig. S15). This could indicate a shift to glucose metabolism as a response to the metabolic burden.

The ER stress activates signaling pathways that leads to cell cycle arrest as indicated by the decrease in S-phase, replication dependent histones (both mRNA and protein level), and deactivation of STAT3 (pTyr705 and pSer727). STAT3 activation can be restored by releasing the cells from GlcN media (see supplemental Fig. S1). The STAT3 dephosphorylation is unlikely to be caused by steric hindrance from STAT3 O-GlcNAcylaton because STAT3 IP followed by Western blot for O-GlcNAc displayed no signal for O-GlcNAcylation but positive for STAT3 enrichment (Click-IT, CTD110.6, and RL2 assays for O-GlcNAc were attempted). We therefore speculate that the “yin-yang” relationship between phosphorylation and O-GlcNAcylation that we observe occurs not by steric hindrance but rather by regulation of kinases and phosphatase activity.

**GlcN-induced O-GlcNAcylaton Affects Modifications other than the Global Levels of Phosphorylation**—Several reports support the hypothesis of a “yin-yang” relationship between phosphorylation and O-GlcNAcylation (53, 55, 56). These reports include examples ranging from a global, overall “yin-yang” relationship between phosphorylation and O-GlcNAcylation to a “yin-yang” on specific protein targets. In this study, we identified two additional examples of a “yin-yang” relationship between O-HexNAcylaton and phosphorylation. For example, the relationship between O-HexNAcylaton and phosphorylation on calnexin, observed in our study, in which the modification sites were in two different subcellular compart-
ments. Interestingly, our subcellular fractionation in combination with GlcN-induced O-HexNAcylaton down-regulated the phosphorylation level of primarily nuclear proteins (Fig. 5A). This example supports further the hypothesis of global "yin-yang" between O-GlcNAcylation and phosphorylation. The "yin-yang" between O-GlcNAcylation and phosphorylation observed upon GlcN treatment could provide an alternative hypothesis to how GlcN cause cell cycle arrest and displays anticancer properties.

Finally, GlcN caused an overall shift from monomethylation to di- and tri-methylation and altered the acetylation pattern of proteins (Fig. 5A), suggesting the need for studies examining PTMs in combination.

CONCLUSION

The data presented in this study demonstrate that two important protein factors, calnexin and GLU2B, which function in the protein folding and quality control pathway in ER, are O-HexNAc-regulated after GlcN treatment. Furthermore, different ratios of the O-HexNAc modified and canonical peptide from GLU2B and calnexin were observed in different cellular compartments. Because GLU2B and calnexin are involved in ER stress this demonstrates a strong evidence for a direct link between O-HexNAc and GlcN-induced ER stress as either a cause or a consequence and suggests that O-HexNAc potentially in the O-GlcNAc form plays a functional role in the ER in terms of subcellular localization.

Acknowledgments—We thank Dr. Jorg Becker and Joao Sobral (Instituto Gulbenkian de Ciencia, Gene Expression Unit, Oeiras) for microarray service. We thank Catarina Leitao (Advanced Flow Cytometry Unit at Institute for Molecular and Cell Biology, Porto) for flow cytometry technical assistance.

* Funding: RM is supported by PTDC/EIA-EIA/099458/2008 Fundaçça para a Ciência e a Tecnologia (FCT), program FCT investigator 2012, ASC is supported by grant SFRH / BPD / 85569 / 2012 funded by Fundaçã para a Ciência e Tecnologia. IPATIMUP is an Associate Laboratory of the Portuguese Ministry of Science, Technology and Higher Education and is partially supported by FCT. HR, PV, ASC, and RM are further supported by FCT grants (PTDC/QUI-BIQ/099457/2008).

This article contains Supplemental Figs. S1 to S16 and Table S1.

** To whom correspondence should be addressed: Human Genetics Department, National Institute of Health Dr Ricardo Jorge, Av. Padre Cruz, 1649-016 Lisboa, Portugal. Tel.: 351-21-7508145; Fax: 351-21-7526400; E-mail: rumen2009@gmail.com.

Data and materials availability: Mass spectrometry raw data including help files, processed spectra and tables containing the identified peptides, raw spectra number, processed spectra number, peptide scores, FDR statistics, extracted ion counts for peptides, spectra counts for proteins and statistical analysis have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository (1) with the dataset identifier PXD000380. Microarray data is publicly available at http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?&acc=GSE49185, Gene Expression Omnibus (GEO) accession number GSE49185.

REFERENCES


Global Glucosamine Profiling

Molecular & Cellular Proteomics 13.12

3307