Data in Brief 39 (2021) 107471

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Dataset of human EDEM2 melanoma cells proteomics, affinity proteomics and deglycoproteomics

Gabriela N. Chiriţoiu[a], Marioara Chiriţoiu[a], Cristian V.A. Munteanu[b,*]

[a] Department of Molecular Cell Biology, Institute of Biochemistry, Splaiul Independenţei 296, Bucharest 060031, Romania
[b] Department of Bioinformatics and Structural Biochemistry, Institute of Biochemistry, Splaiul Independenţei 296, Bucharest 060031, Romania

ARTICLE INFO

Article history:
Received 2 September 2021
Revised 2 October 2021
Accepted 4 October 2021
Available online 14 October 2021

Keywords:
Endoplasmic reticulum associated degradation (ERAD)
Stable isotope labeling by amino acids in cell culture (SILAC)
Cancer
Mass spectrometry
liquid chromatography
ITGA1
Tyrosinase
Proteasome

ABSTRACT

EDEM2 (Endoplasmic reticulum Degradation-Enhancing alpha-Mannosidase-like protein 2) is one of the key-proteins suggested to be involved in the selection and degradation of misfolded proteins from the endoplasmic reticulum. The datasets discussed in this article are related to experiments covering affinity proteomics, label-free quantitative proteomics, deglycoproteomics and SILAC (Stable Isotope Labeling by Amino Acids in Cell Culture) proteomics data of A375 melanoma cells with modified expression of EDEM2. Our first aim was to affinity-enrich EDEM2 alongside its potential interaction partners and analyse the obtained samples by nanoLC-MS/MS to identify novel EDEM2 associated proteins. The dataset was substantiated by SDF (Sucrose Density Fractionation)-nanoLC-MS/MS experiments, in an integrated workflow to validate EDEM2 identified partners and corroborate these with previous data. Our second aim was to delineate novel EDEM2 substrate candidates using a two-step strategy. The first one refers to the deglycoproteomics dataset, which covers nanoLC-MS/MS analysis of Concanavalin A enriched glycopeptides released by endoglycosidase digestion from A375 melanoma cell lysates.

* Corresponding author.
E-mail address: cristian.munteanu@biochim.ro (C.V.A. Munteanu).
Social media: (G.N. Chiriţoiu), (M. Chiriţoiu), (C.V.A. Munteanu)

https://doi.org/10.1016/j.dib.2021.107471

This allowed us to map the fraction of glycoproteins with non-matured N-glycans from A375 melanoma cells and find or validate *N*-glycosylation sites of proteins from the secretory pathway. The same dataset was also used to define glycoproteins altered by the down-regulation of endogenous EDEM2, which should contain its candidate-substrates. In a second step we delineate the degradation kinetics of some of these proteins using a pulse SILAC strategy (pSILAC) thus complementing our initial findings with a fourth dataset. Beside nanoLC-MS/MS analysis our findings were also validated by various biochemical experiments. All the data described are associated with a research article published in Molecular and Cellular Proteomics [1].

Specifications Table

Subject	Omics: Proteomics
Specific subject area	Endoplasmic reticulum associated degradation (ERAD) Proteomics
Type of data	Tables Figures Raw data
How data were acquired	The data was acquired using nanoLC-MS/MS analysis: Easy-nanoLC II using a reversed-phased analytical column (RP C18 Acclaim PepMap 100, ThermoFisher Scientific), coupled online to an LTQ-Orbitrap Velos Pro hybrid mass spectrometer (ThermoFisher Scientific).
Data format	Raw Analyzed Filtered
Parameters for data collection	The peptides were separated using a 120/240 min gradient of 2–30% solvent B (0.06% formic acid – FA, 80% acetonitrile- ACN) and detected using data-dependent acquisition, in which a survey scan acquired in the orbital trap was followed by the fragmentation of top 5/10 most abundant ions and detection of the fragments in the linear trap (LTQ). For glycodataset the fragmentation was performed in the higher collisional dissociation (HCD) cell with Orbitrap detection.
Description of data collection	The raw data were analysed using either the Andromeda algorithm integrated into the Maxquant environment (v1.5.3.17/v1.6.0.1) or SEQUEST integrated into the Proteome Discoverer v1.4 (ThermoFisher Scientific) by searching the data against the human version of the UniProtKB database.
Data source location	The raw data has been collected in the mass spectrometry (Protein Chemistry Facility –PCF) laboratory from the Department of Bioinformatics and Structural Biochemistry (DBSB) from the Institute of Biochemistry of the Romanian Academy (IBRA). The data have been deposited to the ProteomeXchange Consotium through the PRIDE repository with the dataset identifier PXD025881.
Data accessibility	Repository name: proteomexchange Data identification number: PXD025881 Direct URL to data: http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD025881
Related research article	Munteanu CVA, Chiriţoiu GN, Chiriţoiu M, Ghenea S, Petrescu AJ, Petrescu ŞM. Affinity proteomics and deglycoproteomics uncover novel EDEM2 endogenous substrates and an integrative ERAD network. Mol Cell Proteomics. 2021 Jul 28:100125. doi:10.1016/j.mcpro.2021.100125. Epub ahead of print. PMID: 34332121.

Value of the Data

- The datasets provide relevant information regarding the secretory pathway associated proteome under various conditions (up-/down-regulated EDEM2 expression) but also about the glycosylation status of numerous proteins from this pathway in melanoma cells.
- Researchers which are interested in protein folding, quality control, degradation (ERQC, ERAD) and deglycoproteomics, particularly those studying these mechanisms in tumor cells will find these datasets valuable.
- The data could be reused to study additional molecular mechanisms regarding the complex interlink between protein misfolding, degradation and cancer cell response to stress.

1. Data Description

The deposited raw data encloses four proteomics datasets of A375 melanoma cells with altered expression of a key-protein involved in the degradation of misfolded glycoproteins (ER degradation-enhancing alpha-mannosidase-like protein 2 – EDEM2). EDEM2 was suggested to be one the first mannosidases involved in the selection and degradation of misfolded glycoproteins [2], thus deciphering its impact at the proteome and glycoproteome level is an essential step in the aim of finding its endogenous substrates. The acquired data focused on two aspects: first, mapping proteins associated with EDEM2 in melanoma cells to denote the most important factors in misfolded protein degradation and second, generation of a short candidate list of proteins disturbed by the EDEM2 overexpression or down-regulation. For the first aim we generated two main datasets comprising LC-MS/MS raw data acquisition of affinity enriched EDEM2 from A375 melanoma cell lysates stably expressing EDEM2 and a second dataset comprising LC-MS/MS analysis of sucrose density separated fractions (SDF) from similar cell lysates. Experiments were performed in biological replicates and the SEQUEST XCorr distributions were compared between replicates and different fractions to validate the SDF dataset (Fig. 1). Analysis of the obtained data depicted the most important factors associated with EDEM2, which could participate in the degradation of misfolded substrates as was also shown by our alternative biochemical methods. The second aspect of the dataset focused on the analysis of changes at the proteome and deglycoproteome level following either short-term inducible expression of EDEM2 or its silencing. The deglycoproteomics data set was obtained following melanoma cells transfection with siRNA targeting the endogenous human EDEM2, glycopeptides isolation using lectin-affinity and nanoLC-MS/MS analysis following EndoH (Endoglycosidase H) digestion, thus providing one the earliest datasets regarding the EndoH deglycoproteome of melanoma cells. For this dataset experiments were performed in biological replicates and to assess up-regulated proteins, which would constitute substrate candidates, quality control plots were generated first to validate the dataset (Fig. 2). To confirm some candidates a second dataset was generated, in which cells were SILAC labelled with heavy arginine (R6/R10) and light lysine (K0). Following short term-induction of EDEM2 with tetracycline (TET) the media was switched to light arginine (R0) and heavy lysine (K4/K8) and cells were harvested at different time-points. The experiment was performed in replicates with label-swap to assess proteins with up- or down-regulated patterns following EDEM2 induction (Fig. 3). Thus, the results from the proteomic and deglycoproteomic datasets would contain potential candidates of the ER degradation machinery from melanoma cells and possible proteins generating HLA-associated peptides. The repository data files located at https://www.ebi.ac.uk/pride/archive/projects/PXD025881 contains: LC-MS/MS raw files generated for this dataset, msf. files generated by Proteome Discoverer analysis, MaxQuant output files as zipped archives and single peptide IDs as a pdf file.

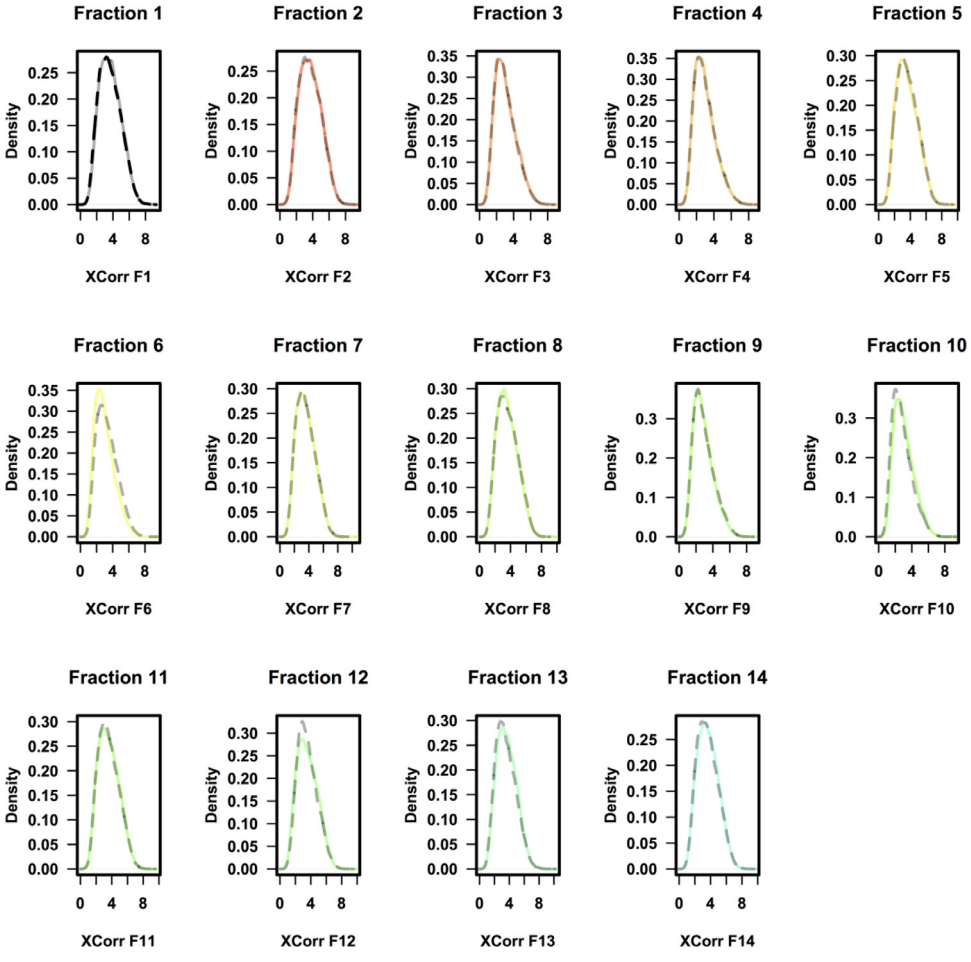

Fig. 1. Comparison of the SEQUEST XCorr score distribution across fractions and replicates for the SDF dataset. Solid and dashed lines denote the two distinct replicates.

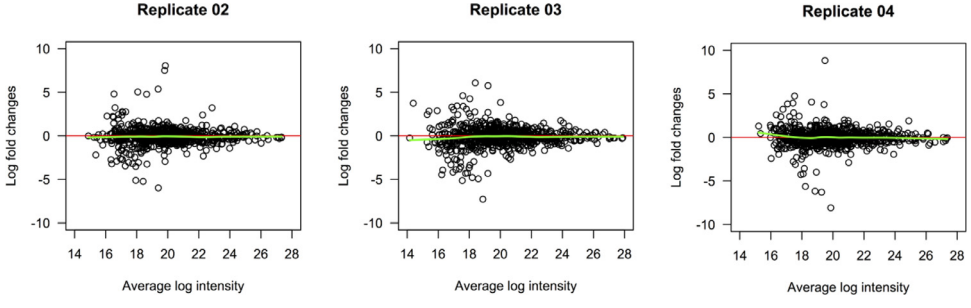

Fig. 2. Analysis of the EndoH deglycoproteome dataset. MA plots of three biological replicates. Shown is the average versus fold change in log scale for each replicate. Red line denotes no change ('zero' fold change) and green line displays a local regression.

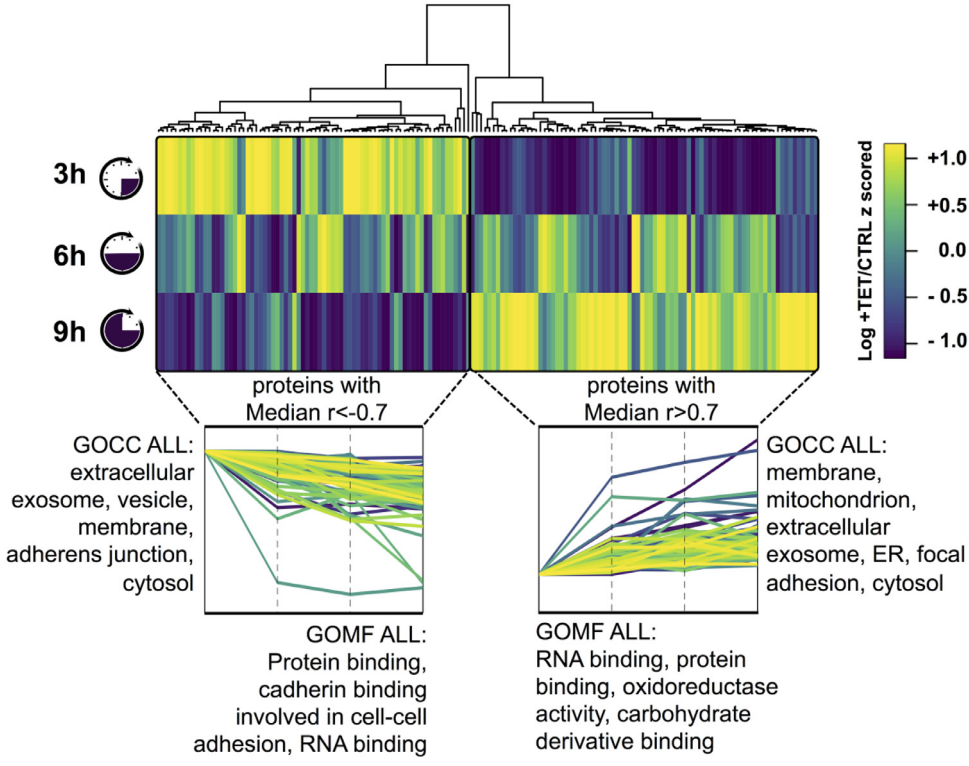

Fig. 3. Gene Ontology (GO) analysis of proteins identified from the SILAC proteomics dataset with a negative or positive expression level trend following EDEM2 induction. EDEM2 expression was induced following tetracycline (TET) treatment of cells. Shown are the *Z* scores of log ratios (TET/CTRL) from proteins with negative r (Pearson correlation below -0.7) or positive r (above 0.7). The lower part mentions some GO key-terms associated with the identified proteins.

2. Experimental Design, Materials and Methods

2.1. Experimental design

Dataset acquisition using nanoLC-MS/MS focused around two main goals: characterization of EDEM2 associated proteins and identification of novel EDEM2 potential substrates. Thus, to characterize the interaction partners we generated two datasets, one covering affinity proteomics data which includes biological triplicates with six fractions and a second one including the sucrose density fractionation (SDF) data with biological duplicates covering 14 fractions with technical duplicates. For EDEM2 candidate substrate investigation we generated another two datasets the first one covering a N-deglycoproteomics experiment performed in triplicate with 2–4 technical replicates and a second one covering a pSILAC proteomics experiment performed in biological duplicates with label swap.

2.2. Materials and methods

2.2.1. Cell lines and biological material

The cell lines A375-pLPCX (A375-C), A375 HA tagged EDEM2-pLPCX (A375-E2), used for the affinity proteomics and SDF datasets, were obtained using a retroviral system as previously de-

scribed [3], A375-ST-TYR-E2i, used for the pSILAC dataset, were obtained using a lentiviral system as described in the related article [1].

2.2.2. Data acquisition

Sample analysis was performed using an Easy-nanoLC II coupled online to an LTQ-Orbitrap Velos Pro instrument (Thermo Scientific) as described below and elsewhere [1]. Samples reconstituted in solvent A (0.06% FA and 2% ACN) were injected and the peptides were first concentrated on a RP C18 Acclaim™ PepMap™ 100 trap column (ThermoFisher Scientific), followed by their separation on a RP C18 Acclaim™ PepMap™ 100 analytical column (ThermoFisher Scientific) using a 2–30% solvent B (0.06% FA and 80% ACN) for 120 or 240 min, depending on sample complexity. A top five or ten data-dependent method was used for data acquisition in which an Orbitrap initial scan at 60,000 resolution (m/z 400) was followed by the fragmentation and detection of the corresponding fragments in the linear trap portion of the instrument. Only charge states of 2 and higher were considered for fragmentation. The PTM dataset was acquired by a preliminary scan in the orbital trap at 30,000 resolution (m/z 400) followed by the fragmentation of the selected ions in the higher collisional dissociation (HCD) cell and fragment detection at a resolution of 7500 (m/z 400) in the orbital trap. The lock mass option was enabled for 445.120026 during data acquisition with a spray voltage of ∼1.8–2.0 kV and a transfer capillary temperature of 275 or 300 °C. To reduce redundancy the option for dynamic exclusion was enabled with duration of 30/60 s, a repeat count of 1 and a maximum list size of 500. In general, for samples with multiple injections an exclusion list was created to keep redundancy to a minimum.

2.3. Data analysis

Raw data were analysed using either the Andromeda algorithm from MaxQuant [4] or the SEQUEST module from Proteome Discoverer v1.4 (ThermoFisher Scientific). For both cases the human version of the UniProtKB database was used with trypsin as the protease (additionally GluC for the endoglycosidase samples with multiple digestions) and considering Cys carbamidomethylation as a fixed/static modification and Met oxidation as a dynamic/variable modification. The Andromeda analysis also considered protein *N*-terminus acetylation as a variable modification. Furthermore for the PTM dataset, additionally, HexNAc and dHexHexNAc modifications on Asn residues were considered and for the SILAC data heavy and medium Arg and Lys modifications of the corresponding peptides. To reduce the false positives, the results were filtered at 1% FDR, estimated by the target-decoy method [5]. Label free quantification was performed either using the label free quan (LFQ) option from MaxQuant [6] with default values, or using the spectral counts values (PSMs) reported by Proteome Discoverer. For MaxQuant analysis the option of match between runs was activated under a match time window alignment of 20 min and a match time of 0.7 min. Results were further analysed using GrapPad Prism v6.00, R v3.5.3 [7], SaintExpress [8,9] or Perseus [10], after establishing quality thresholds for the data as previously described [1]. Quantification at the level of ion current was performed using the MaxQuant reported LFQ values. Values were log transformed, identifications from reversed database and contaminants were removed and only samples with all values in at least one of the conditions from the biological replicates were further kept. Missing values were replaced by values with a similar standard deviation and a median close to the instrument detection limit. To assess significance a two-sample t-test with permutation-based FDR correction was applied for the affinity dataset and the Significance analysis of microarray (SAM) [11] was used for the deglycoproteomics data. Quantification was also performed at the spectral count level for the affinity enrichment dataset and the SDF data. To evaluate significance, spectral counts were exported and SaintExpress [9] was used to denote statistically enriched proteins. For SDF the normalized spectral count abundance score (NSC) was calculated as previously descried [12].

Ethics Statement

There is no ethical issue for this study as no animals or patients were involved in data acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships, which have or could be perceived to have influenced the work reported in this article.

CRediT Author Statement

Gabriela N. Chiriţoiu: Investigation, Methodology, Writing – original draft; **Marioara Chiriţoiu:** Investigation; **Cristian V.A. Munteanu:** Conceptualization, Methodology, Investigation, Writing – original draft.

Acknowledgments

This work was supported by the Romanian Ministry of Education and Research grants, CNCS - UEFISCDI, project numbers PN-III-P1-1.1-PD-2019-1242, PN-III-P1-1.1-PD-2019-1278, PN-III-P2-2.1-PED-2019-3297 within PNCDI III, grant PNCDI-III-PCCDI-2018-1 and the Romanian Academy.

References

[1] C.V.A. Munteanu, G.N. Chiriţoiu, M. Chiriţoiu, et al., Affinity proteomics and deglycoproteomics uncover novel EDEM2 endogenous substrates and an integrative ERAD network, Mol. Cell. Proteom. (2021) 1–21, doi:10.1016/j.mcpro.2021.100125.
[2] S. Ninagawa, T. Okada, Y. Sumitomo, et al., EDEM2 initiates mammalian glycoprotein ERAD by catalyzing the first mannose trimming step, J. Cell Biol. 206 (3) (2014) 347–356.
[3] C.V.A. Munteanu, G.N. Chiriţoiu, A.J. Petrescu, et al., Profiling optimal conditions for capturing EDEM proteins complexes in melanoma using mass spectrometry, in: A.G. Woods, C.C. Darie (Eds.), Advancements of Mass Spectrometry in Biomedical Research, Springer International Publishing, Cham, 2019, pp. 155–167. Editors.
[4] J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification, Nat. Biotechnol. 26 (12) (2008) 1367–1372.
[5] J.E. Elias, S.P. Gygi, Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry, Nat. Methods 4 (3) (2007) 207–214.
[6] J. Cox, M.Y. Hein, C.A. Luber, et al., Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ*, Mol. Cell. Proteom. 13 (9) (2014) 2513–2526.
[7] R Development Core Team, R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, 2014.
[8] H. Choi, B. Larsen, Z.Y. Lin, et al., SAINT: probabilistic scoring of affinity purification–mass spectrometry data, Nat. Methods 8 (1) (2011) 70–73.
[9] G. Teo, G. Liu, J. Zhang, et al., SAINTexpress: Improvements and additional features in significance analysis of interactome software, J. Proteom. 100 (2014) 37–43.
[10] S. Tyanova, T. Temu, P. Sinitcyn, et al., The Perseus computational platform for comprehensive analysis of (prote)omics data, Nat. Methods 13 (9) (2016) 731–740.
[11] V.G. Tusher, R. Tibshirani, G. Chu, Significance analysis of microarrays applied to the ionizing radiation response, Proc. Natl. Acad. Sci. 98 (9) (2001) 5116.
[12] K.J. Kirkwood, Y. Ahmad, M. Larance, et al., Characterization of native protein complexes and protein isoform variation using size-fractionation-based quantitative proteomics*, Mol. Cell. Proteom. 12 (12) (2013) 3851–3873.