Median based absolute quantification of proteins using **F**ully **U**nlabelled **G**eneric **I**nternal **S**tandard (**FUGIS)**

Bharath Kumar Raghuraman[1], Aliona Bogdanova[1], HongKee Moon[1], Ignacy Rzagalinski[1], Eric Geertsma[1], Lena Hersemann[1], Andrej Shevchenko [1, 2]

[1]Max Planck Institute of Molecular Cell Biology and Genetics, Pfotenhauerstr. 108, 01307, Dresden, Germany

[2]Corresponding author; e-mail: shevchenko@mpi-cbg.de

Abstract

By reporting molar abundances of proteins, absolute quantification determines their stoichiometry in complexes, pathways or networks and also relates them to abundances of non-protein biomolecules. Typically, absolute quantification relies either on protein- specific isotopically labelled peptide standards or on a semi-empirical calibration against the average abundance of peptides chosen from arbitrary selected standard proteins. Here we developed a generic protein standard FUGIS (Fully unlabelled Generic Internal Standard) that requires no isotopic labelling, synthesis of standards or external calibration and is applicable to proteins of any organismal origin. FUGIS is co-digested with analysed proteins and enables their absolute quantification in the same LC-MS/MS run. By using FUGIS, median based absolute quantification (MBAQ) workflow provides similar quantification accuracy compared to isotopically-labelled peptide standards and outperforms methods based on external calibration or selection of best ionized reporter peptides (*e.g.* Top3 quantification) with a median quantification error below 20%

Key Words: Absolute quantification of proteins, MS Western workflow, proteome-wide quantification

Introduction

Quantitative proteomics envelopes multiple workflows providing either relative or absolute quantification of individual proteins. Relative quantification determines how the abundance of the same protein changes across multiple conditions on a proteome-wide scale. As a complementary approach, absolute quantification determines the exact molar quantity of each protein in each condition. In this way, it is possible to relate the abundances of different proteins, estimate their expression level or determine stoichiometry within a variety of molecular constellations from stable complexes to organelles or metabolic pathways and interaction networks [1]–[10]. Absolute quantification holds an important promise to deliver reference values of individual proteins in liquid and solid biopsies, which is a pre-requisite for robust molecular diagnostics.

A palette of absolute quantification techniques for different analytical platforms, biological contexts and aims were developed. It is usually assumed that the abundance of a few representative peptides reflects the abundance of the corresponding endogenous protein of interest. In turn, peptides quantification relies either on protein- specific isotopically labelled peptide standards having exactly the same sequence as endogenous peptides or on a semi-empirical calibration against the abundances of selected (or, alternatively, of all detectable) peptides originating from arbitrary chosen standard proteins [11][12]. The former class of methods is accurate, yet targets a limited selection of proteins whose scope cannot be changed during the experiment. The latter methods are meant to cover the entire proteome, however they rely on some arbitrary assumptions and deliver experiment- and protein-

dependent accuracy. Hence, they are typically employed for mapping expression levels of proteins or similar experiments on a proteome-wide scope.

AQUA[13], an absolute quantification method uses a set of isotopically labelled synthetic peptide standards mimicking the proteotypic peptides from endogenous proteins. With advances in gene synthesis, techniques like QconCAT[14], PSAQ [15], PrEST [16], PCS [17], MEERCAT [18], DOSCAT [19], and GeLC-based MS Western [20] were developed which uses isotopically labelled protein chimeras that, upon proteolytic cleavage, produce the desired proteotypic standards. Unlike the aforementioned strategies, MS Western uses multiple proteotypic peptides per protein, which provides concordant quantification by controlling the ratio of abundances of peptide peaks. Common discrepancies in these ratios are due to the miscleaved peptides or unexpected modifications and are the reason to exclude corresponding pairs of endogenous and standard peptides from calculating the protein abundance.

In order to avoid isotopic labelling, label free internal standards like MIPA[21] and SCAR[22] were developed by using minimal sequence permutation or scrambling. The scrambled peptide acts as a close analogue of the endogenous peptide that retains its key physicochemical properties, yet alleviating the need for isotopic labelling [23][24]. While being sensitive, multiplexed and accurate, and expandable to quantifying PTMs, they were only intended to cover a few (typically, up to 50) target proteins.

Advances in robust and reproducible LC-MS/MS have led to a notion that generic, yet arbitrary measures of protein abundances could be deduced from raw peptide intensities that are not linked to pre-defined standards: emPAI [25], APEX [26], SCAMPI [27], [28], NSAF rely on the MS/MS fragmentation frequency of matched

peptides (otherwise known as spectral counts) and can be deduced from LC-MS/MS runs directly. Methods like Top3/Hi-3 [6], iBAQ [29],Proteomic Ruler [30] , xTop [31] and Pseudo-IS[32] use averaged MS1 intensities of selected or of all matched peptides assuming they are relatively independent of properties of individual peptides and the protein composition of the analysed sample. Not surprisingly, the reported concordance of abundance estimates was limited.

Hence there is a need in a method that combines accuracy and precision of the internal standard based targeted quantification with broad (potentially, proteome-wide) coverage and ease of use of untargeted methods. To this end, we developed an untargeted quantification workflow called median based absolute quantification (MBAQ) and designed a fully unlabelled generic internal standard (FUGIS) by considering common physicochemical properties of tryptic peptides which can be used for untargeted absolute quantification of proteins.

Materials and Methods

Protein extraction from HeLa cells

HeLa Kyoto cells were cultured in Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum and 1% penicillin-streptomycin (Gibco™ Life Technologies). HeLa cells were trypsinized, counted and washed 2x with PBS, before 1 $\times 10^6$ cells were lysed 30 mins on ice in either 1 mL or 0.5 mL RIPA buffer containing CLAAP protease inhibitors cocktail (10 µg/ml Aprotinin, 10 µg/ml Leupetin, 10 µg/ml Pepstatin, 10 µg/ml Antipain and 0.4 mM Phenylmethylsulfonyl Fluoride (PMSF)). Subsequently, cells were further lysed by passing them 10 times through a 25g syringe. A post-nuclear supernatant was obtained from a 15 mins centrifugation at

14.000 x g in an Eppendorf 5804R centrifuge. The supernatant was then used for the further analysis by GeLC-MS/MS workflow (**Supplementary methods**) with both MS-Western standard and FUGIS standard in separate experiments.

Absolute quantification of HeLa proteins using MS Western

Absolute protein quantification was performed using MS Western protocol [20] . The total protein content from HeLa cells from both dilutions were loaded on to a precast 4 to 20% gradient 1-mm thick polyacrylamide mini-gels were from Anamed Elektrophorese (Rodau, Germany) for 1D SDS PAGE. Separate gels were run for 1 pmol of BSA and isotopically labelled lysine (K) and arginine (R) incorporated chimeric standard containing 3-5 unique quantitypic peptides from target proteins The sample was cut into 3 gel fractions and each fraction was co digested with known amount of BSA and the chimeric standard using Trypsin Gold, mass spectrometry grade, (Promega,Madison). The digest was analysed using GeLC-MS/MS workflow (**Supplementary methods**). The peptide matching and chromatographic peak alignment from the raw files was carried out as described in **Supplementary methods** (Database **search and data processing**). The quantification was performed using the software developed in-house [9].

Absolute quantification of HeLa proteins using MBAQ and FUGIS

Similar to the MS Western experiments, the total HeLa cell lysate from both the dilutions were subjected to separation using 1D SDS PAGE. Similarly, separate gels were run for 1pmol of BSA and the fully unlabelled generic internal standard (FUGIS). The sample was cut into 3 gel fractions and each fraction was co-digested with the known amount of BSA and the FUGIS. The digests were again analysed using the GeLC-MS/MS workflow (**Supplementary methods**). The desired amount of FUGIS

on column was maintained between 200 fmol and 400 fmol. The peptide matching and chromatographic peak alignment from the raw files was carried out as described in **Supplementary methods** (Database **search and data processing**). The output .csv files with peptide XIC peak area and peptide identities were used as the input for GlobeQuant software.

GlobeQuant software for MBAQ quantification

GlobeQuant software was developed as a stand-alone Java script based application using in-memory SQL database (ref: https://github.com/agershun/alasql) for fast access and search in the CSV file. GlobeQuant runs on a Windows 7 workstation with 16 GB RAM and 4-cores processor. The .csv output from the Progenesis LC-MS v.4.1 (Nonlinear Dynamics, UK) with peptide ID's and their respective raw XIC peak areas were used by GlobeQuant software. A list of FUGIS peptides was provided as an input. The software first calculates the molar amount of the FUGIS standard by using the scrambled-native BSA peptide pairs. Then the median peak area for the FUGIS peptides are calculated by extracting the XIC peak areas of the FUGIS peptides. The calculated molar amount of the FUGIS standard is equated to the median peak area and used as a single point calibrant.

For BestN quantification peptides were chosen from a pool of Top3 peptides by calculating the coefficient of variation of all possible combination of Best2 and Best3 by default. If a protein does not contain Top3 peptides the Top2 peptides were taken as Best N peptides. Proteins identified with one peptide are excluded from quantification. The BestN combination with the lowest coefficient of variation (<20%)

was taken and averaged to provide the molar amounts of the protein. The software package is available at https://github.com/bharathkumar91/GlobeQuant.

Results and Discussion

MBAQ workflow for absolute quantification

MBAQ (for mean based absolute quantification) protocol relies on recombinant protein standard consisting of peptides whose sequences emulate the physicochemical properties of common proteotypic peptides, yet are not identical to any sequence in a protein database .

Tryptic digestion of a chimera protein consisting of concatenated proteotypic peptides produces them in exactly equimolar concentration [14][19][33][34]. We therefore propose to determine the median value of XIC area of their chromatographic peaks and then use it a single point calibrant to calculate the molar abundance of peptides from proteins of interest provided that the molar abundance of chimera protein is known. Effectively, we reasoned that peak areas of the equimolar amount of proteotypic peptides that were selected according to some common rules [20] could cluster around some median value irrespective of their sequence. If so, we only have to: i) provide a sufficient number of these peptides to compute the robust median value under given experimental conditions; ii) select suitable peptides from those matched to proteins of interest and iii) check if the quantification of target protein by individual peptides is concordant. In our institute we systematically produce large (40 to 270 kDa) protein chimeras comprising 40 to 250 proteotypic peptides from various proteins. We therefore used a few already available chimeras [20][9][35] to test the feasibility of MBAQ.

We first asked how areas of XIC peaks of proteotypic peptides chosen from different proteins and concatenated into a chimera are distributed around the median value. To this end, we digested 267 kDa chimeric protein (CP01) comprising 250 proteotypic peptides from 53 *Caenorhabditis elegans* proteins[4]. Despite their equimolar concentration, peak areas of individual peptides differed by almost an order of magnitude (**Figure 1A**). However the abundance of 48 % of peptides clustered near the median value (**Figure 1A**). In order to ascertain that clustering does not depend on their sequence (again, these were all pre-selected proteotypic peptides), we digested another 265 kDa chimera (CP02) harbouring proteotypic peptides of 48 proteins from *Drosophila melanogaster* [9]. We found that the abundance of 42 % of them was near the median value (**Figure 1A**). We concluded that, independently of peptide sequences, approximately half of proteotypic peptides would cluster around the same median, while others are scattered around. However, the commonality between peptide sequences within clustering and non-clustering groups was not immediately obvious.

Since the “near-median” (NM) peptides were evenly distributed across the retention time range (**Figure S1**), we checked if the median value could faithfully represent the molar abundance of the internal standard. We expect that, in this case, possible suppression of their ionization by a sample matrix would be likely randomized, compared to a hypothetical scenario if they would be eluting (almost) together. For this purpose, we used the CP01 to re-quantify 48 metabolic enzymes from *Caenorhabditis elegans* by MS Western protocol and, independently, by using a median value computed from the abundances of all peptides from the same 267 kDa chimera CP01. We underscore that in MS Western, each enzyme was quantified using isotopically labelled peptides that exactly matched sequences of corresponding native peptides [4]

with no recourse to other peptides, while in MBAQ all peptides were taken for calculating a single median value that was subsequently used for quantifying all proteins. The MBAQ was concordant with MS Western showing Pearson's correlation of 95 % (**Figure 1B**) and median quantification error of 18 % (**Figure 1C**) within 3 orders of magnitude of molar abundance difference.

In a separate experiment, we also quantified 30 proteins from the commercially available UPS2 protein standard (Sigma Aldrich, USA) using MBAQ and medians calculated from CP01 peptides. The Pearson's correlation was 96 % and the median quantification error was also less than 20 % (**Supplementary information, MBAQ vs UPS2**).

We therefore concluded that, if sufficient number of equimolar "near-median" peptides are detected by LC-MS/MS, their median abundance is invariant to their sequences and unaffected by other peptides included into the chimera. Use of the median abundance as a single point calibrant delivers good quantification accuracy that is close to the accuracy of the quantification relying on few peptide standards that are identical and unique to endogenous peptides from each target protein.

Though the median based absolute quantification was accurate the use of large isotopically labelled CP's looked unnecessary. Indeed, we only used about a half of its peptides and take no advantage of isotopic labelling, except for quantifying target proteins. In order to close this gap, we seek to design a generic (suitable for all proteins from all organisms) a Fully Unlabelled Generic Internal Standard (FUGIS).

Development of FUGIS.

FUGIS was conceived as a relatively short protein chimera composed of concatenated proteotypic-like tryptic peptides that, however, share no sequence identity to any known protein. It also comprises a few reference peptides with close similarity to some common protein standard *e.g.* BSA. Upon co-digestion with quantified proteins, FUGIS should produce equimolar mix of peptide standards whose median abundance would support one-point MBAQ quantification of co-detected target peptides as described above. The exact amount of FUGIS is quantified by comparison with the known amount of the reference protein (here, BSA) in the same LC-MS/MS experiment.

We first asked, what is the minimum number of peptides required to arrive to a consistent median value? For this purpose, we performed a bootstrapping experiment over the median abundance of tryptic peptides derived from CP01 and CP02 chimeras. Peptides were quantified by LC-MS/MS and median values calculated by repetitive selection of defined (3 to 120) number of peptide (**Figure 2**). The data collected by 100 bootstrap iterations suggested that, consistent median value can be reasonably projected by averaging XICs of as little as 5 to 10 peptides. However, the medians spread (which depends on “internal” peptide properties and “external” conditions of ionization) decreased with the number of averaged peptides and reached plateau at > 30 averaged peptides (**Figure 2A, B**). Also, bootstrapping confirmed that, irrespective of the selected peptide set, same peptides tend to cluster around the median. The abundance of 32 % of the total of 230 peptides were within 20 % of the median. Therefore, for further work we selected these peptide sequences (70 peptides) which showed concordant abundances with respect to the median value in multiple technical LC-MS/MS replicates.

Next, we altered sequences of these pre-selected “near-median” peptides such that they become different from any known sequence. Yet, we tried to preserve the similarity of their physicochemical properties, such as net charge, hydrophobicity index and location of polar (including C-terminal Arg (R) or Lys (K)) amino acid residues compared to corresponding “source” peptides. We noted that the examination of sequences of near-median and distant-median peptides did not reveal an unequivocal rationale behind clustering and therefore we tried several ways to change their sequences in parallel.

In our first strategy we selected a set of 40 peptides (from the 70 peptide pool and reversed their amino acid sequences (**Figure 3A**) except C-terminal lysine or arginine, assembled them in a chimeric protein (GCP01) that was expressed and metabolically labelled with $^{13}C^{15}N$-Arg and ^{13}C-Lys in *Escherichia coli* [20]. We analysed its band by GeLC-MS/MS[20], including co-digestion with the band with 1 pmol BSA. Similar to previously published strategy [36], the peptide abundances were normalised to the abundance of BSA peptides in the chimeric protein to see if the normalised median abundance (NMA) is close to unity (~1.0). A unit (~1.0) NMA means the median abundance truly represents the amount of the FUGIS standard while any deviation contributes to the error in quantification.

The NMA for the reversed sequences was 0.45, (**Figure 3B**) which was very far from the NMA of their native counterpart of 0.97. Thus, we concluded that reversing peptide sequences biases the median, increases the spread and should not be used for making a FUGIS chimera.

Next we scrambled the sequences by introducing point substitutions of amino acid residues. We allowed a maximum of two scrambling events per peptide by

following two intuitive rules. First, in each peptide two amino acid residues were swapped (**Figure 3A**). Second, to create a mass shift, an amino acid residue preferably located in the middle of the peptide sequence was substituted with another amino acid with the similar side chain (e.g. S to T or vice versa) (**Figure 3A**). To minimize the retention time shift, the aliphatic amino acids in the order of increasing hydrophobicity (G<A<V<L<I) were only substituted with amino acid with similar hydrophobicity (i.e. substitutions V by L were allowed, but G by I were not). Altogether, 20 scrambled sequences were assembled into a chimera together with corresponding 20 source "native" peptides (GCP02). The pairwise comparison of XIC peak areas of the native and the scrambled sequences suggested that they differed by less than 5 %. Similar to GCP01, we calculated the NMA for the peptides in GCP02. The scrambled peptides behaved similar to the native sequences with a NMA of 1.02 (**Figure 3B**). On average, the retention time difference between native and scrambled peptides was 3.21 (± 2.02) minutes. Therefore, these scrambled peptide sequences were chosen as a basis for further optimisation.

The isotopic labelling of GCP01 and GCP02 was unavoidable as their quantification was dependent on the reference BSA peptides. We next checked if these sequences could be also scrambled without affecting their ionisation behaviour to circumvent metabolic labelling. Scrambled BSA peptides behaved similar the native peptides with a retention time shift of 1.2 (± 0.5) minute. Also the relative abundances (peptide ratios)[20] of corresponding native and scrambled BSA peptide was almost the same (**Figure 3C**).

Taken together, we designed and produced FUGIS chimera with a molecular weight of 79.01 kDa (**Figure 3D**) (**Figure S2**) (**Supplementary information, FUGIS**) which harbours 43 sequence-scrambled NM peptides and 5 sequence of scrambled

peptides of BSA. All peptides showed no complete match to any protein sequence across organisms (**Supplementary information, FUGIS**).

MBAQ quantification using FUGIS

In order to assess the feasibility and accuracy of MBAQ quantification using FUGIS we quantified 4 proteins from 1 million HeLa cells at 2 dilutions and compared it with the quantities previously determined using MS Western [20]. Since MBAQ quantification is based on the median abundance, we wanted to access the accuracy of the median estimation in different matrix background. To this end we pre-fractionated the HeLa cells lysates from both dilutions by 1D-SDS PAGE gels and collected 3 gel bands from each. Each gel band was co digested with 1 pm of BSA and FUGIS.

In order to test effect of matrix background on the value of peak area median, we calculated NMA for the FUGIS with a HeLa digest background. The ratio was 0.98 irrespective of protein background with an error less than 10 % (**Figure 3E**).

We then proceeded to quantify the molar amounts of 4 proteins (PLK-1, TBA1A, CAT, G3P) from HeLa cells using MBAQ, MS Western[20] and by Top3/Hi-3 quantification[6] (**Figure 4**) (**Supplementary information, MBAQ vs MSW vs Hi3**). The quantities determined with MS Western was highly concordant with previously published molar amounts [20]. We noticed that, MBAQ provided molar amounts closer to MS Western than that of the Top-3/Hi-3 quantification. The MBAQ quantification was carried out by choosing peptides in the target protein, whose mean and median are similar with a variation of less than 15 % and provided concordant quantities. We called these peptides with low inter-peptide variability as “BestN” peptides – in contrast

to TopN peptides that correspond to N most abundant peptides from the quantified protein.

In order to assess whether using BestN peptides delivered better accuracy, we looked into the quantification of one of the four proteins (glyceraldyhyde-3 phosphate dehydrogenase (G3P Human P04406)) (**Figure 5A**) (**Supplementary information, MBAQ vs MSW vs Hi3**). We estimated the concordance of the molar amount independently calculated from multiple peptides by the coefficient of variation [9]. The coefficient of variation (%CV) of the molar amounts calculated using BestN peptides was 7 %, which was much lesser than that of the Hi-3 quantification (18 %) (**Figure 5B**) (**Figure S3**) (**Supplementary information, MBAQ vs MSW vs Hi3**).

In order to understand why the BestN peptides improved the quantification accuracy, we looked into the identities of the peptides. In the quantification of glyceraldyhyde-3 phosphate dehydrogenase (P04406) and tubulin 1 alpha (Q71U36), the top ionising peptide were excluded because they increased the inter-peptide variability. Excluding the best ionising peptide reduced the coefficient of variation down to less than 10 % (**Supplementary information, BestN Peptides**). In the case of human catalase protein (P04040) and serine/threonine protein kinase (P53350) the Top2 peptides were the BestN peptides, which also provided lower variability (**Supplementary information, BestN Peptides**). Taking together, we concluded that the BestN peptides that were chosen for the quantification were a subset of TopN peptides and a minimum of 2 peptides are required to provide reliable molar amounts.

Taking the molar amounts calculated with MS Western as a “true value”, we evaluated the accuracy of MBAQ quantification. MBAQ with BestN peptides provided the most accurate quantification with an accuracy of 96% (**Figure 5B**). We also

observed that MBAQ when used together with TopN also performed better than the conventional Top3 quantification with an accuracy of 94 % (**Figure 5B**). We note that here both TopN and BestN quantification relied on the same FUGIS chimera, whose abundance is link to the reference protein standard (BSA from Pierce that is used for BCA analysis) with exactly known concentration.

Having established the accuracy of MBAQ quantification using BestN, we expanded the quantification landscape of MBAQ to all HeLa proteins identified with a minimum of 2 peptides. GlobeQuant software supports MBAQ quantification workflow (**Figure 6A**) by choosing BestN peptides for quantification and using FUGIS as a single generic protein standards. We quantified in total 1450 proteins in both dilutions of HeLa lysate and plotted it as a ranked cumulative abundance (**Figure 6B**) (**Supplementary information, GlobeQuant HeLa**)**.** The most and the least abundant protein in was human actin B (ACTB) and DnaJ homolog (DNAJC11). Since we analysed samples with 2 fold dilution, we looked if MBAQ can faithfully recapitulate the fold change difference. The accuracy of fold change estimation was 92 % which shows that MBAQ can capture 2-fold differences in protein amounts very efficiently. These results reinforced that MBAQ together with BestN provides accurate molar quantities and can faithfully capture 2 fold differences in the absolute abundance.

Conclusion and perspectives

MBAQ quantification relying upon FUGIS, a generic protein standard, provided accurate molar quantities because it was composed of curated and optimised generic peptide sequences which provided concordant median abundance per mole value of the protein. High expression level, full solubility of FUGIS makes it preferred internal

standard for any label-free experiment aiming at the absolute and relative quantification. Upon tryptic digestion, it produces 43 peptides in exactly known equimolar amount covering common range of peptide retention times. The versatile use of FUGIS as a single point calibrant and normalisation standard will accelerate routine discovery based proteomics of any detectable protein. . Though the current workflow involves GeLC-MS/MS strategy, it can be easily adjusted for in-solution protocols with no need in the purification of the FUGIS.

It has long been noticed that the abundance of proteins could be projected from the abundances top ionised (TopN) peptides, as in Hi-3 quantification[6], however the method suffered from outliers hampered the quantification[31][36][37]. In MBAQ, the use of BestN instead of TopN peptides circumvented this problem by providing 2-3 equimolar calibrant peptides for robust measurement of the molar amount. For each protein NM peptides can be selected according to simple and intuitive requirement of concordant mean and median areas of XIC peaks. This, effectively, alleviates the quantification bias caused by selection of best-ionized peptides whose ionization capacity is strongly sequence-dependent. Furthermore, using a single generic protein whose abundance is referenced to a recognized commercial quantitative standard is an important move towards establishing diagnostically relevant protein values in clinical applications. MBAQ can be expanded to quantify any protein which can be detected with multiple peptides in the same experiment, including data-independent acquisition (DIA).

Having the proteome expressed in molar quantities will facilitate our understanding of multiple pathway modules and their interaction paving way for the molar modular proteome. We propose that MBAQ will help to digitalize a substantial part of the

proteome in one experiment and help in discovering proteotype-phenotype relationships.

Author Contributions

BKR and AS conceptualized and designed the FUGIS and MBAQ and BKR performed the experiment and interpreted the data. AB expressed the FUGIS standard. HKM and LH conceptualized and developed the software. BKR and AS wrote the manuscript. IR provided critical and expert advice for the project.

Acknowledgements

Work in AS group was supported by Max Planck Society. We would like to thank Dr. Sandra Segeletz for providing the HeLa cells. The figures were assembled using Bio render tool (Created with BioRender.com)

Data availability

Raw data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD026886

References

[1] K. Kito, K. Ota, T. Fujita, T. Ito, *J. Proteome Res.* **2007**, *6*, 792.

[2] I. Wohlgemuth, C. Lenz, H. Urlaub, *Proteomics* **2015**, *15*, 862.

[3] J.R. Wiśniewski, D. Rakus, *J. Proteomics* **2014**, *109*, 322.

[4] S. Penkov, B.K. Raghuraman, C. Erkut, J. Oertel, R. Galli, E.J.M. Ackerman, D. Vorkel, J.M. Verbavatz, E. Koch, K. Fahmy, A. Shevchenko, T. V. Kurzchalia, *BMC Biol.* **2020**, *18*, 31.

[5] J. Nielsen, J.T. Pronk, *FEMS Yeast Res.* **2011**, *12*, 2628.

[6] J.C. Silva, M. V Gorenstein, G.-Z. Li, J.P.C. Vissers, S.J. Geromanos, *Mol. Cell. Proteomics* **2006**, *5*, 144.

[7] K. Kito, T. Ito, *Curr. Genomics* **2008**, *9*, 263.

[8] A. Hammel, D. Zimmer, F. Sommer, T. Mühlhaus, M. Schroda, *Front. Plant Sci.* **2018**, *9*, 1265.

[9] B.K. Raghuraman, S. Hebbar, M. Kumar, H. Moon, I. Henry, E. Knust, A. Shevchenko, *Proteomics* **2020**, *20*, 1900049.

[10] S.K. Stelder, C. Benito De Moya, H.C.J. Hoefsloot, L.J. De Koning, S. Brul, C.G. De Koster, *J. Proteome Res.* **2018**, *17*, 903.

[11] E. Ahrné, L. Molzahn, T. Glatter, A. Schmidt, *Proteomics* **2013**, *13*, 2567.

[12] F. Calderón-Celis, J.R. Encinar, A. Sanz-Medel, *Mass Spectrom. Rev.* **2018**, *37*, 715.

[13] S.A. Gerber, J. Rush, O. Stemman, M.W. Kirschner, S.P. Gygi, *Proc. Natl.*

Acad. Sci. U. S. A. **2003**, *100*, 6940.

[14] R.J. Beynon, M.K. Doherty, J.M. Pratt, S.J. Gaskell, *Nat. Methods* **2005**, *2*, 587.

[15] S.E. Kaiser, B.E. Riley, T.A. Shaler, R.S. Trevino, C.H. Becker, H. Schulman, R.R. Kopito, *Nat. Methods* **2011**, *8*, 691.

[16] M. Zeiler, W.L. Straube, E. Lundberg, M. Uhlen, M. Mann, *Mol. Cell. Proteomics* **2012**, *11*, O111.009613.

[17] K. Kito, M. Okada, Y. Ishibashi, S. Okada, T. Ito, *Proteomics* **2016**, *16*, 1457.

[18] N. Takemori, A. Takemori, Y. Tanaka, Y. Endo, J.L. Hurst, G. Gómez-Baena, V.M. Harman, R.J. Beynon, *Mol. Cell. Proteomics* **2017**, *16*, 2169.

[19] R.J. Bennett, D.M. Simpson, S.W. Holman, S. Ryan, P. Brownridge, C.E. Eyers, J. Colyer, R.J. Beynon, *Sci. Rep.* **2017**, *7*, 45570.

[20] M. Kumar, S.R. Joseph, M. Augsburg, A. Bogdanova, D. Drechsel, N.L. Vastenhouw, F. Buchholz, M. Gentzel, A. Shevchenko, *Mol. Cell. Proteomics* **2018**, *17*, 384.

[21] D. Winter, J. Seidler, D. Kugelstadt, B. Derrer, B. Kappes, W.D. Lehmann, *Proteomics* **2010**, *10*, 1510.

[22] E.R. Remily-wood, J.M. Koomen, **2012**, 188.

[23] T. Yoneyama, S. Ohtsuki, M. Tachikawa, Y. Uchida, T. Terasaki, *J. Proteome Res.* **2017**, *16*, 1556.

[24] Y. Li, H. Ji, L. Shen, X. Shi, J. Wang, *Anal. Biochem.* **2019**, 113451.

[25] Y. Ishihama, Y. Oda, T. Tabata, T. Sato, T. Nagasu, J. Rappsilber, M. Mann, *Mol. Cell. Proteomics* **2005**, *4*, 1265.

[26] P. Lu, C. Vogel, R. Wang, X. Yao, E.M. Marcotte, *Nat. Biotechnol.* **2007**, *25*, 117.

[27] G. Rosenberger, C. Ludwig, H.L. Röst, R. Aebersold, L. Malmström, *Bioinformatics* **2014**, *30*, 2511.

[28] S. Gerster, T. Kwon, C. Ludwig, M. Matondo, C. Vogel, E.M. Marcotte, R.

Aebersold, P. Bühlmann, *Mol. Cell. Proteomics* **2014**, *13*, 666.

[29] B. Schwanhüusser, D. Busse, N. Li, G. Dittmar, J. Schuchhardt, J. Wolf, W. Chen, M. Selbach, *Nature* **2011**, *473*, 337.

[30] J.R. Wiśniewski, M.Y. Hein, J. Cox, M. Mann, *Mol. Cell. Proteomics* **2014**, *13*, 3497.

[31] M. Mori, Z. Zhang, A. Banaei-Esfahani, J. Lalanne, H. Okano, B.C. Collins, A. Schmidt, O.T. Schubert, D. Lee, G. Li, R. Aebersold, T. Hwa, C. Ludwig, *Mol. Syst. Biol.* **2021**, *17*, e9536.

[32] T. Tabata, T. Sato, J. Kuromitsu, Y. Oda, *Anal. Chem.* **2007**, *79*, 8440.

[33] K.M. Carroll, D.M. Simpson, C.E. Eyers, C.G. Knight, P. Brownridge, W.B. Dunn, C.L. Winder, K. Lanthaler, P. Pir, N. Malys, D.B. Kell, S.G. Oliver, S.J. Gaskell, R.J. Beynon, *Mol. Cell. Proteomics* **2011**, *10*, M111.007633.

[34] Z. Shi, H. Yu, Y. Sun, C. Yang, H. Lian, P. Cai, **2015**.

[35] S. Penkov, B.K. Raghuraman, C. Erkut, J. Oertel, R. Galli, E.J.M. Ackerman, D. Vorkel, J.-M. Verbavatz, E. Koch, K. Fahmy, *BMC Biol.* **2020**, *18*, 1.

[36] D.G.S. Smith, G. Gingras, Y. Aubin, T.D. Cyr, *J. Proteomics* **2016**, *146*, 133.

[37] M.H.D.R. Al Shweiki, S. Mönchgesang, P. Majovsky, D. Thieme, D. Trutschel, W. Hoehenwarter, *J. Proteome Res.* **2017**, *16*, 1410.

Figure 1. MBAQ Quantification A) Clustering of MS1 peak area of peptides of CP01 and CP02 in 3 independent chromatographic runs. B) Comparison of molar quantities of 48 metabolic enzymes from *Caenorhabditis elegans* derived using MBAQ quantification and MS Western, C) %Error of quantification for MBAQ using MS Western derived molar amounts as the "True value".

Figure 2. Minimum number of peptides for median estimation. Bootstrapping of MS1 peak area of peptides in A) CP01 and B) CP02 over a fixed number of peptides ranging from 3 to 120. Median per iteration is represented as ◆

A total 100 iteration per peptide number was performed. The green bars represent the peptide number with stable median.

Figure 3. Development of FUGIS. A) a. Example representation of reversing of peptide sequence used in GCP01, b. Example representation of scrambling of peptide sequence used in GCP02. # represents a swap and @ represents substitution. B) Normalised median abundance (NMA) of reverse, native and scrambled peptide sequences. C) Relative peptide abundance distribution (Peptide ratio) of native and scrambled peptides. * represents a scrambled sequence. a/a* (HLVDEPQNLIK/HLVEEPNQLIK), b/b* (LGEYFGQNALIVR/LGDYGFNNALIVR), c/c* (YLYEIAR/YLYDVAR), d/d* (DAFLGSFLYEYSR/DAFIGTFLYEYSR). D) Peptide sequence of the 80 kDa FUGIS (Supplementary information, FUGIS). The red sequences are the scrambled BSA peptides. The grey sequences represent the native protein sequence used for scrambling. – represents the position of swap or substitution

(represented by green colour amino acid) (Full sequence is available in Figure S2A). E) NMA of FUGIS peptides in a HeLa cell digest background.

Figure 4. Comparison between molar quantities of MBAQ, MS Western and Hi3 quantification. MBAQ vs MS Western vs Top3/Hi3 quantification of 4 (PLK-1, CAT, G3P, TBA1A) HeLa cell proteins from 2 different dilutions. The error bars represent +/- SEM from technical replicates.

Figure 5. Quantification of G3P protein using multiple strategies A) Schematic representation of multiple quantification strategies. The XIC peak areas of the peptides are provides in Figure S3 B) % Coefficient of variation of molar amounts calculated using multiple peptides in every strategy with the % Error of quantification per strategy using MS Western amounts as true value.

Figure 6. MBAQ quantification of HeLa proteome using GlobeQuant A) Schematic representation MBAQ –GlobeQuant workflow B) Ranked cumulative protein abundance of 1450 HeLa proteins from both the dilutions with least abundant protein on the right. ACTB is the most abundant protein.

Figure S1. Retention distribution of near median peptides with the gradient elution profile in grey.

Figure S2. A) Full length sequence of FUGIS standard. B) Expressed FUGIS standard separated using 1D SDS PAGE. C) Retention time distribution of peptides in FUGIS standard with the gradient elution profile in grey.

Figure S3. XIC peak area of Top4 peptides of GAPDH (P04406) with rational behind the selection of BestN peptides. %CV is the coefficient of variation of multiple selections.

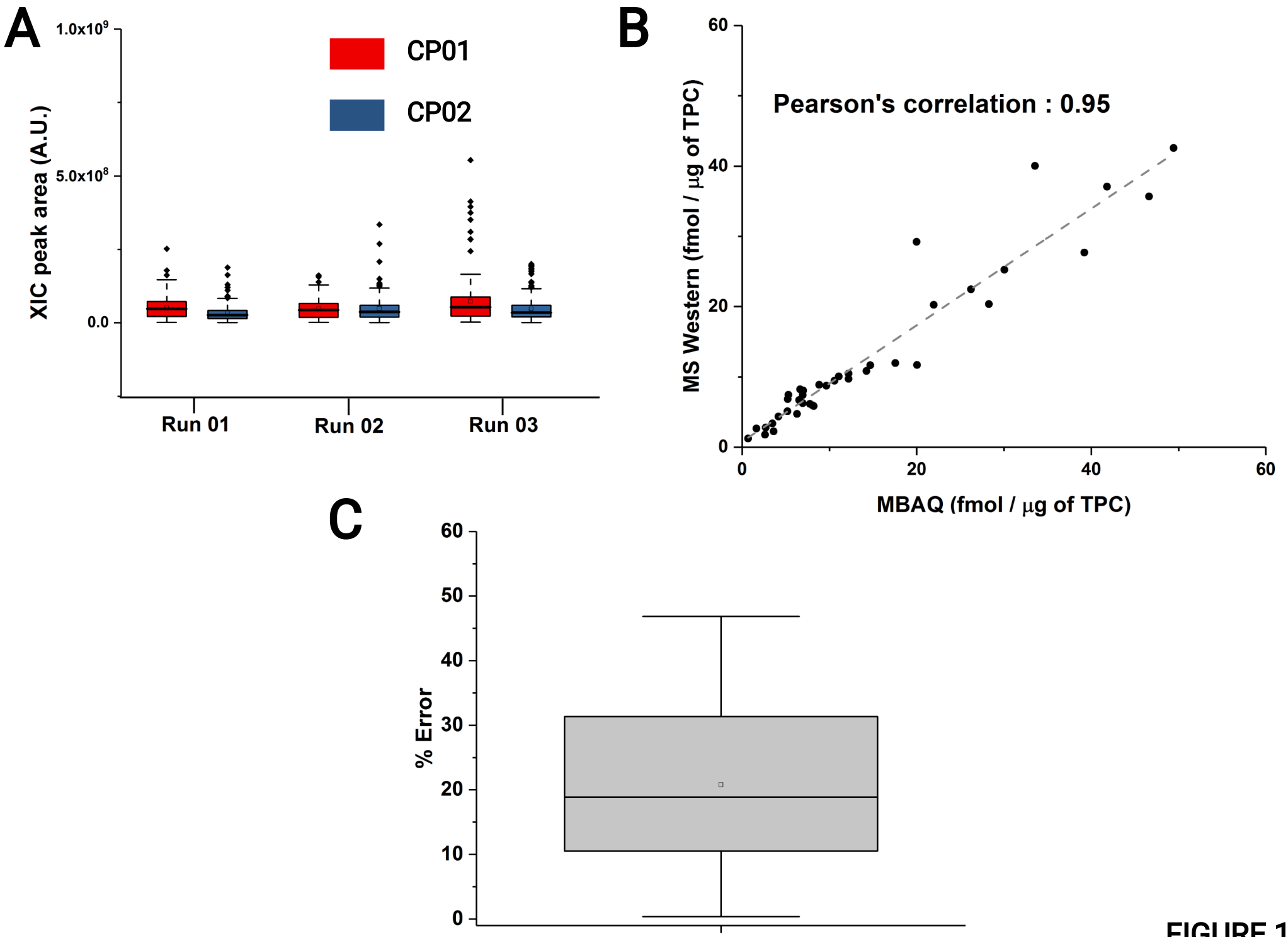
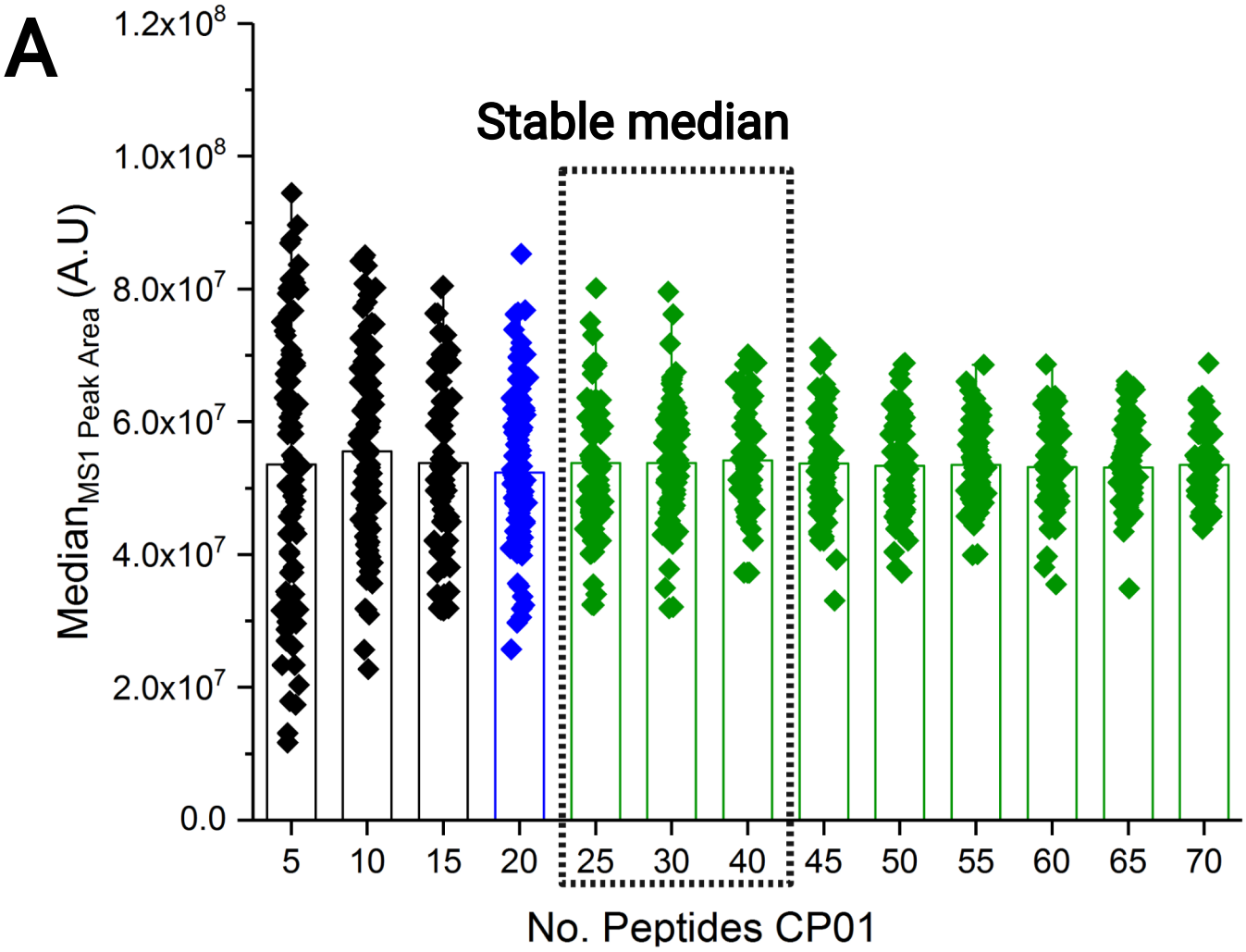

A
1.0x10^9
5.0x10^8
0.0
XIC peak area (A.U.)
CP01
CP02
Run 01
Run 02
Run 03
B
Pearson's correlation : 0.95
MS Western (fmol / µg of TPC)
MBAQ (fmol / µg of TPC)
0
20
40
60
C
% Error
0
10
20
30
40
50
60

FIGURE 1


A
Stable median
1.2x10^8
1.0x10^8
8.0x10^7
6.0x10^7
4.0x10^7
2.0x10^7
0.0
Median MS1 Peak Area (A.U)
5
10
15
20
25
30
40
45
50
55
60
65
70
No. Peptides CP01

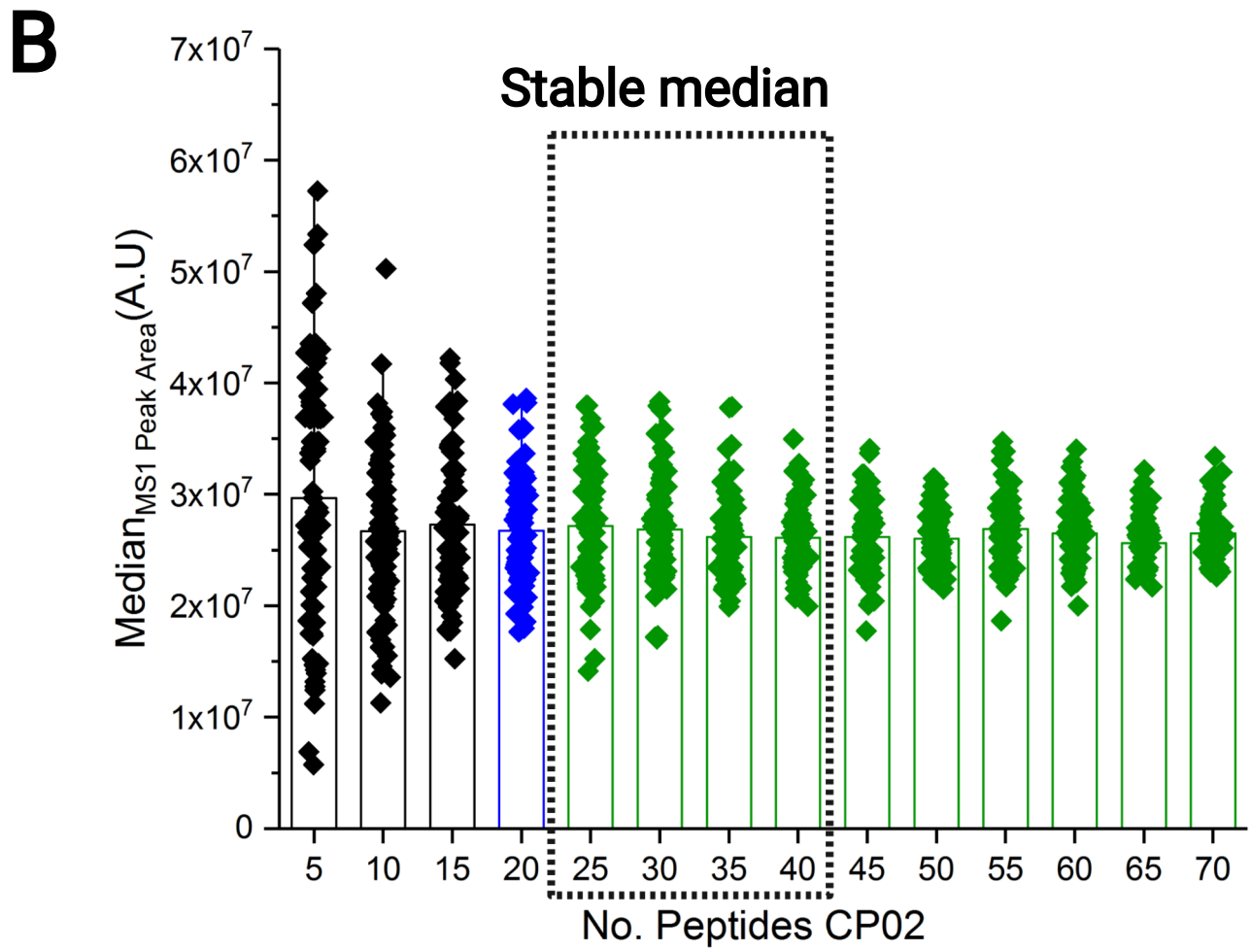

B
Stable median
7x10^7
6x10^7
5x10^7
4x10^7
3x10^7
2x10^7
1x10^7
0
Median MS1 Peak Area (A.U)
5
10
15
20
25
30
35
40
45
50
55
60
65
70
No. Peptides CP02

FIGURE 2

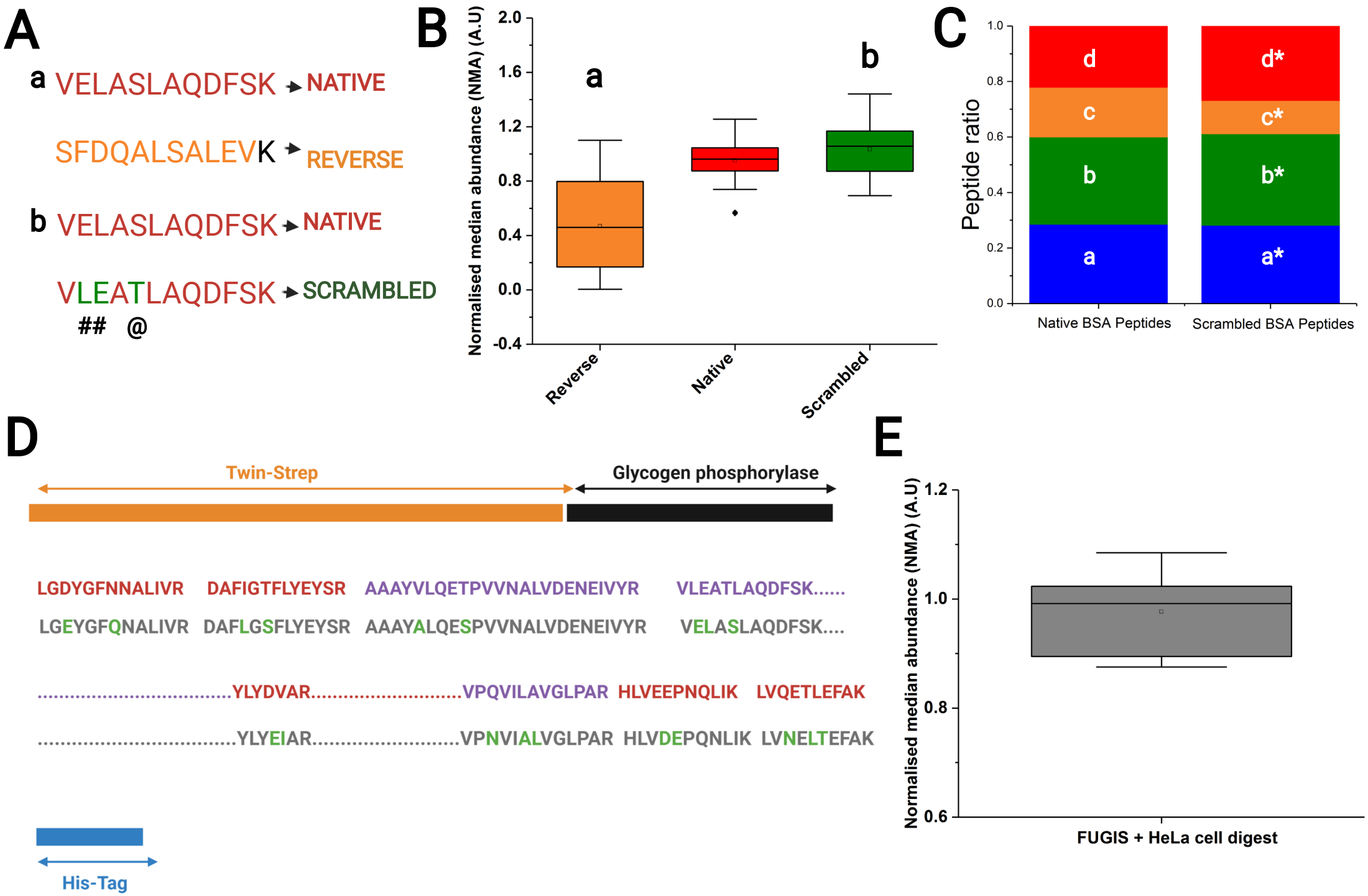

A
a VELASLAQDFSK NATIVE
SFDQALSALEVK REVERSE
b VELASLAQDFSK NATIVE
VLEATLAQDFSK SCRAMBLED
@
B
Normalised median abundance (NMA) (A.U)
2.0
1.6
1.2
0.8
0.4
0.0
-0.4
a
b
Reverse
Native
Scrambled
C
Peptide ratio
1.0
0.8
0.6
0.4
0.2
0.0
d
c
b
a
d*
c*
b*
a*
Native BSA Peptides
Scrambled BSA Peptides
D
Twin-Strep
Glycogen phosphorylase
LGDYGFNNALIVR DAFIGTFLYEYSR AAAYVLQETPVVNALVDENEIVYR VLEATLAQDFSK.....
LGEYGFQNALIVR DAFLGSFLYEYSR AAAYALQESPVVNALVDENEIVYR VELASLAQDFSK....
.....YLYDVAR.....VPQVILAVGLPAR HLVEEPNQLIK LVQETLEFAK
.....YLYEIAR.....VPNVIALVGLPAR HLVDEPQNLIK LVNELTEFAK
His-Tag
E
Normalised median abundance (NMA) (A.U)
1.2
1.0
0.8
0.6
FUGIS + HeLa cell digest

FIGURE 3

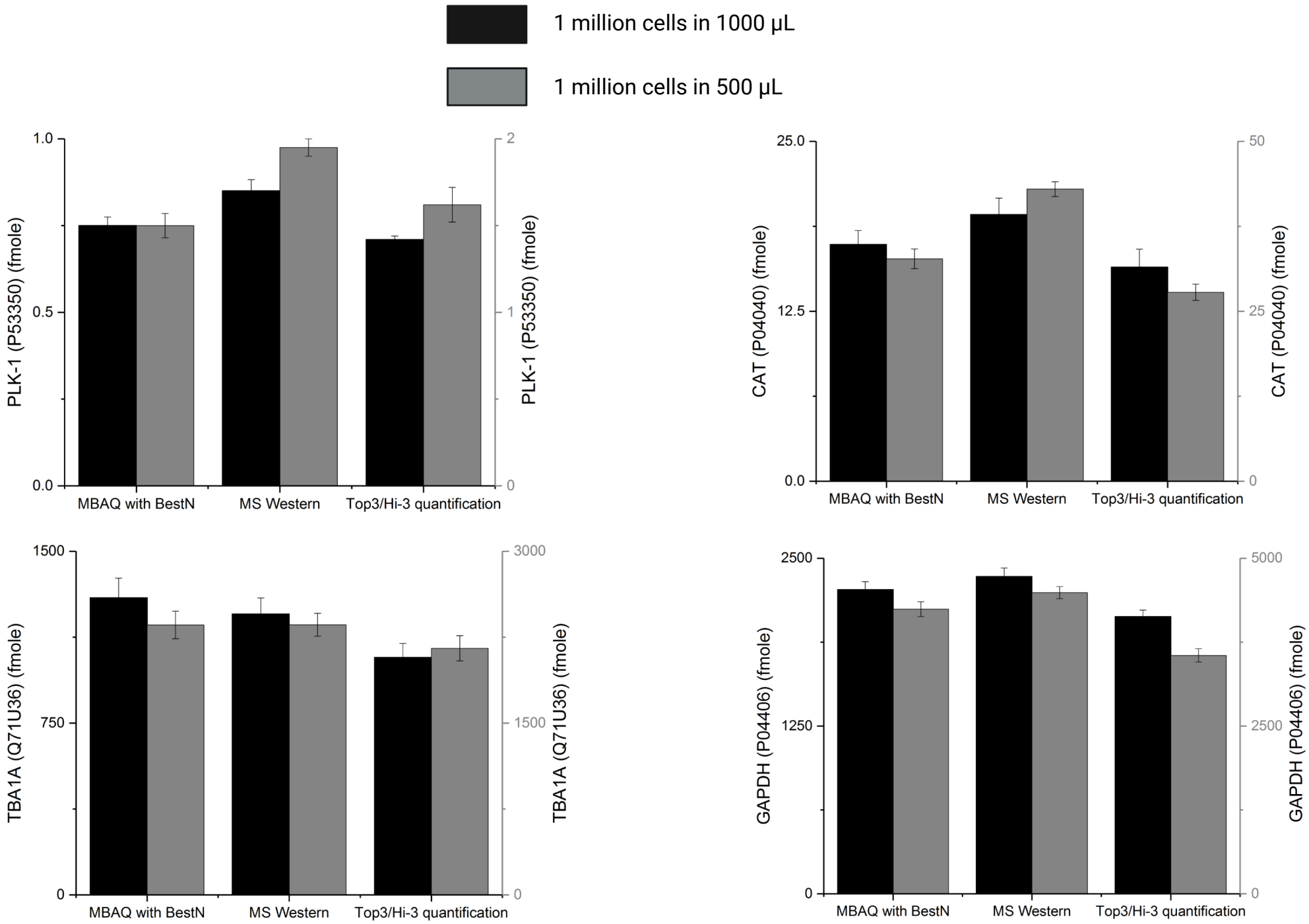

1 million cells in 1000 µL
1 million cells in 500 µL
PLK-1 (P53350) (fmole)
1.0
0.5
0.0
2
1
0
MBAQ with BestN
MS Western
Top3/Hi-3 quantification
CAT (P04040) (fmole)
25.0
12.5
0.0
50
25
0
TBA1A (Q71U36) (fmole)
1500
750
0
3000
GAPDH (P04406) (fmole)
2500
1250
5000

FIGURE 4

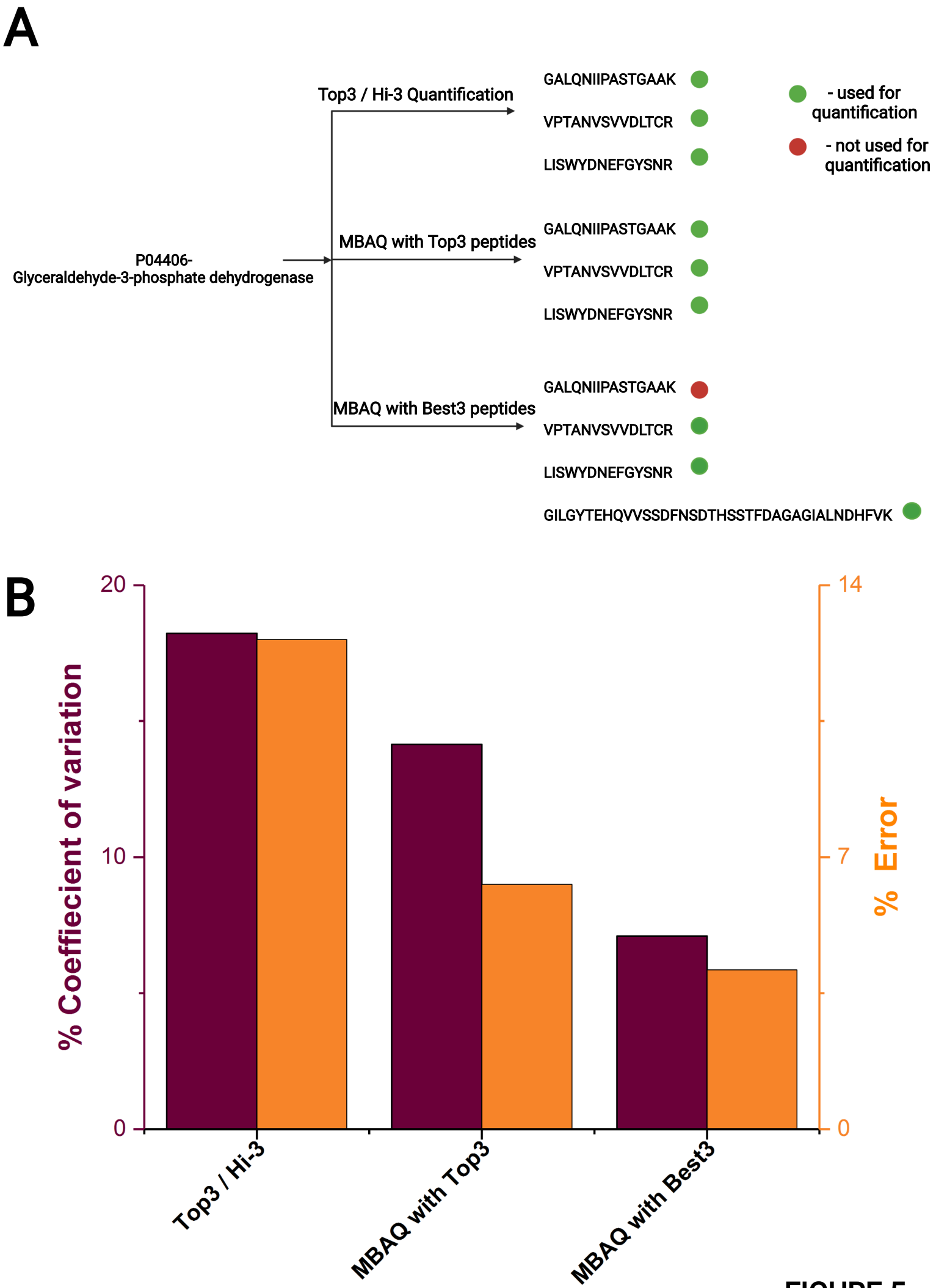
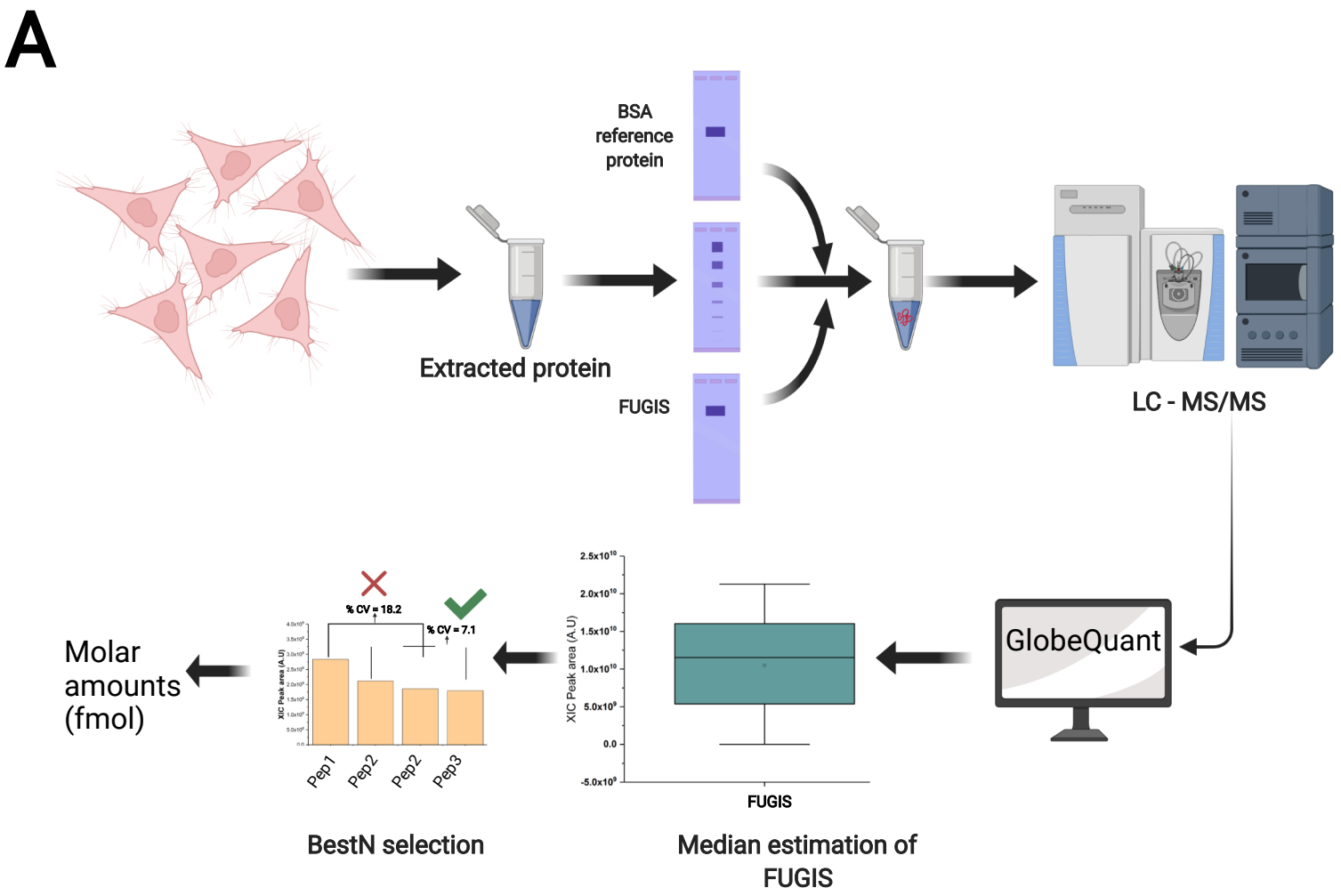

A
P04406-
Glyceraldehyde-3-phosphate dehydrogenase
Top3 / Hi-3 Quantification
MBAQ with Top3 peptides
MBAQ with Best3 peptides
GALQNIIPASTGAAK
VPTANVSVVDLTCR
LISWYDNEFGYSNR
GALQNIIPASTGAAK
VPTANVSVVDLTCR
LISWYDNEFGYSNR
GALQNIIPASTGAAK
VPTANVSVVDLTCR
LISWYDNEFGYSNR
GILGYTEHQVVSSDFNSDTHSSTFDAGAGIALNDHFVK
- used for quantification
- not used for quantification
B
% Coeffiecient of variation
% Error
20
10
0
14
7
0
Top3 / Hi-3
MBAQ with Top3
MBAQ with Best3

FIGURE 5


A
BSA reference protein
Extracted protein
FUGIS
LC - MS/MS
GlobeQuant
Molar amounts (fmol)
% CV = 18.2
% CV = 7.1
Pep1
Pep2
Pep2
Pep3
XIC Peak area (A.U)
FUGIS
BestN selection
Median estimation of FUGIS

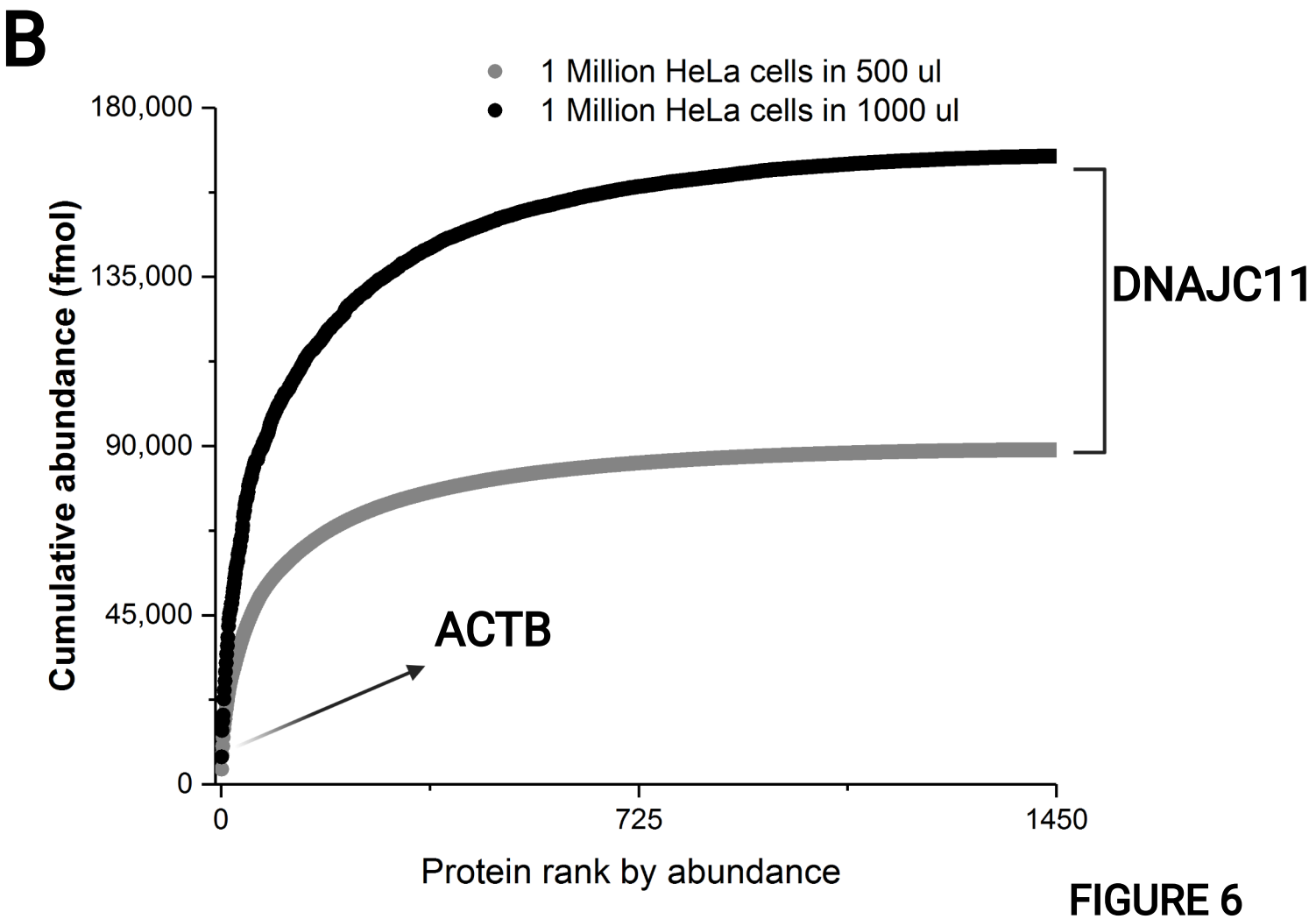

B
1 Million HeLa cells in 500 ul
1 Million HeLa cells in 1000 ul
180,000
135,000
90,000
45,000
0
Cumulative abundance (fmol)
DNAJC11
ACTB
0
725
1450
Protein rank by abundance

FIGURE 6

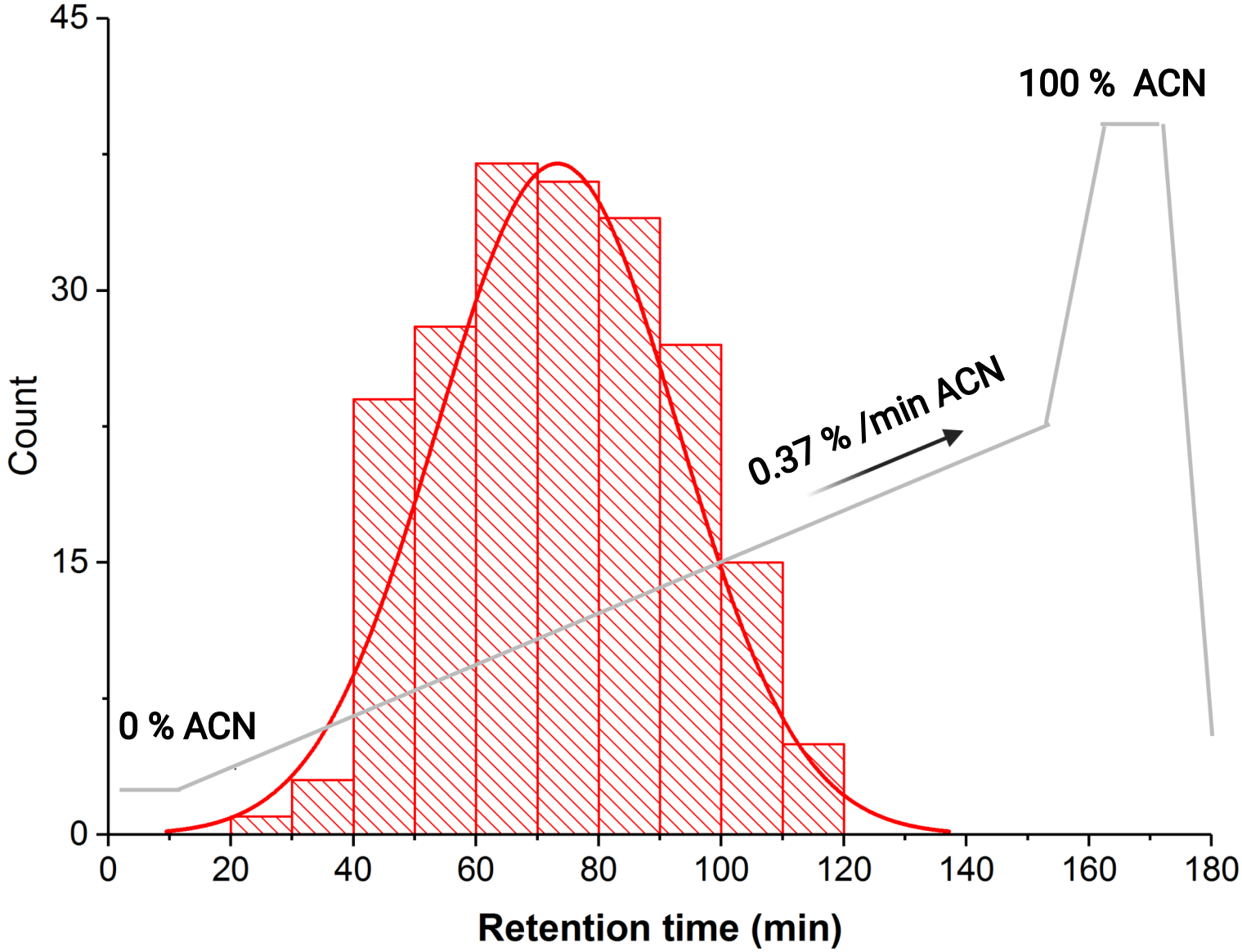

45
30
15
0
Count
0
20
40
60
80
100
120
140
160
180
Retention time (min)
100 % ACN
0.37 % /min ACN
0 % ACN

FIGURE S1

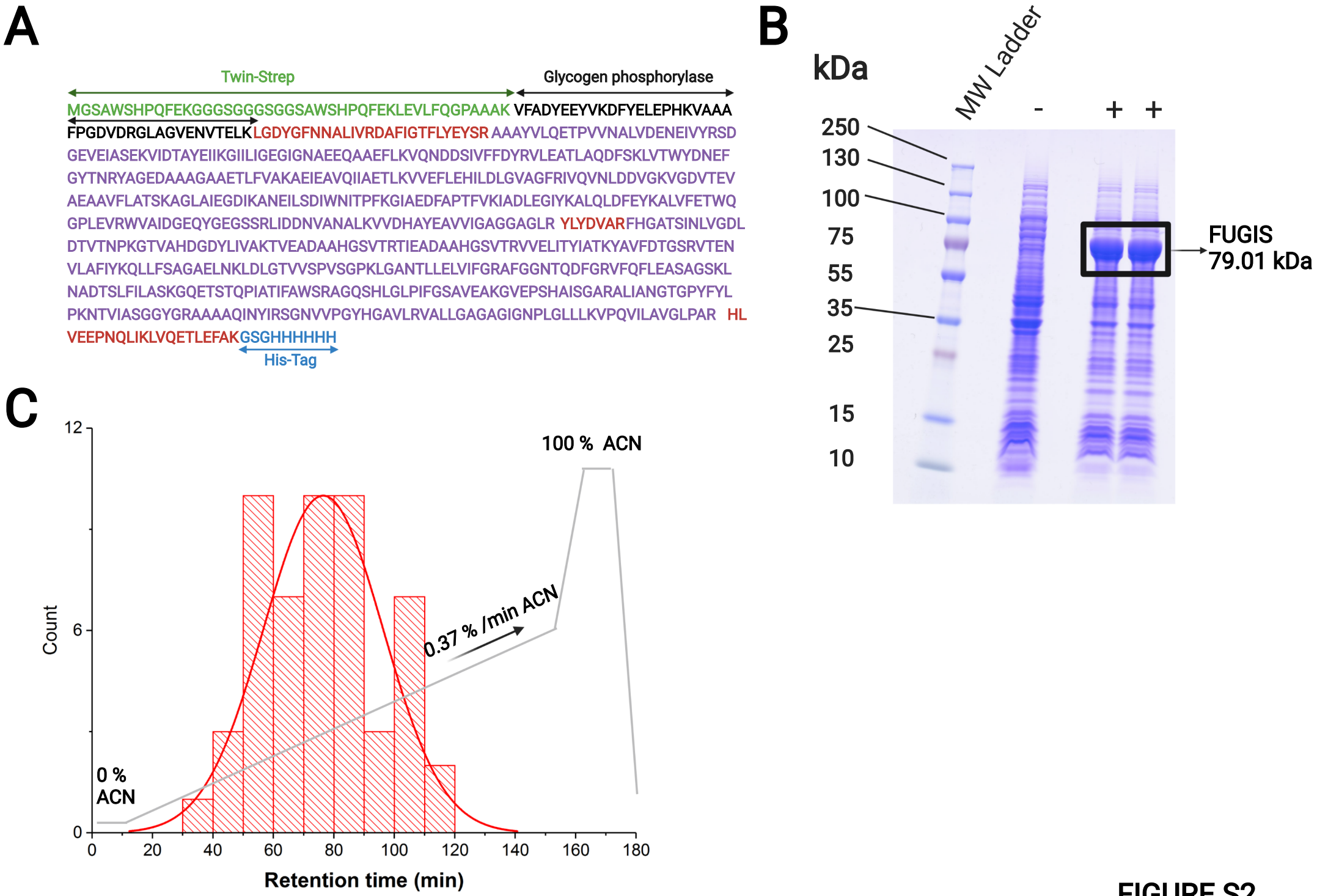

A
Twin-Strep
Glycogen phosphorylase
MGSAWSHPQFEKGGGSGGGSGGSAWSHPQFEKLEVLFQGPAAAK VFADYEEYVKDFYELEPHKVAAA
FPGDVDRGLAGVENVTELKLGDYGFNNALIVRDAFIGTFLYEYSR AAAYVLQETPVVNALVDENEIVYRSD
GEVEIASEKVIDTAYEIIKGIILIGEGIGNAEEQAAEFLKVQNDDSIVFFDYRVLEATLAQDFSKLVTWYDNEF
GYTNRYAGEDAAAGAAETLFVAKAEIEAVQIIAETLKVVEFLEHILDLGVAGFRIVQVNLDDVGKVGDVTEV
AEAAVFLATSKAGLAIEGDIKANEILSDIWNITPFKGIAEDFAPTFVKIADLEGIYKALQLDFEYKALVFETWQ
GPLEVRWVAIDGEQYGEGSSRLIDDNVANALKVVDHAYEAVVIGAGGAGLR YLYDVARFHGATSINLVGDL
DTVTNPKGTVAHDGDYLIVAKTVEADAAHGSVTRTIEADAAHGSVTRVVELITYIATKYAVFDTGSRVTEN
VLAFIYKQLLFSAGAELNKLDLGTVVSPVSGPKLGANTLLELVIFGRAFGGNTQDFGRVFQFLEASAGSKL
NADTSLFILASKGQETSTQPIATIFAWSRAGQSHLGLPIFGSAVEAKGVEPSHAISGARALIANGTGPYFYL
PKNTVIASGGYGRAAAAQINYIRSGNVVPGYHGAVLRVALLGAGAGIGNPLGLLLKVPQVILAVGLPAR HL
VEEPNQLIKLVQETLEFAKGSGHHHHHH
His-Tag
B
kDa
MW Ladder
-
+
+
250
130
100
75
55
35
25
15
10
FUGIS
79.01 kDa
C
100 % ACN
0.37 % /min ACN
0 %
ACN
Count
Retention time (min)
0
6
12
0
20
40
60
80
100
120
140
160
180

FIGURE S2

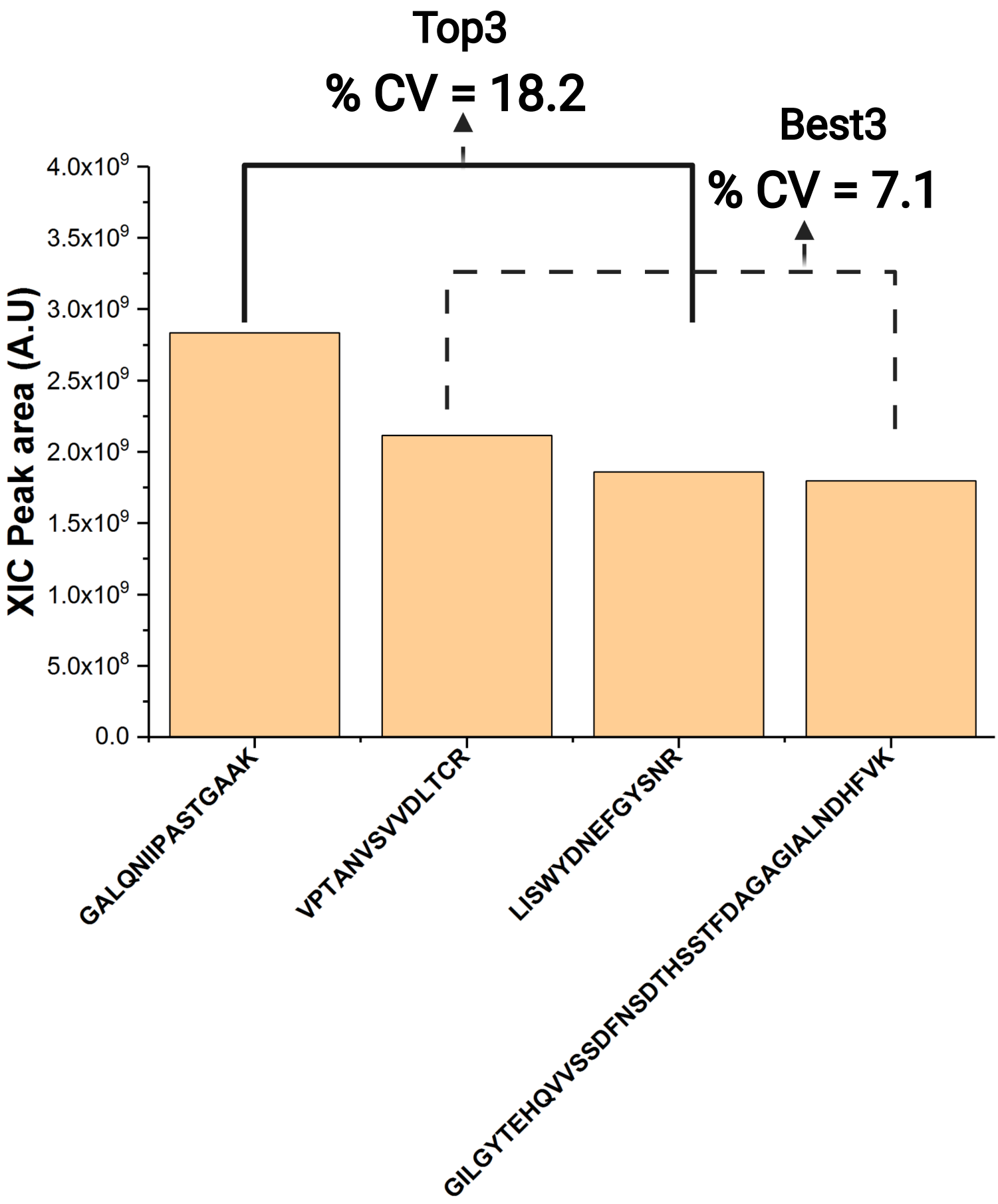

Top3
% CV = 18.2
Best3
% CV = 7.1
XIC Peak area (A.U)
4.0x10^9
3.5x10^9
3.0x10^9
2.5x10^9
2.0x10^9
1.5x10^9
1.0x10^9
5.0x10^8
0.0
GALQNIIPASTGAAK
VPTANVSVVDLTCR
LISWYDNEFGYSNR
GILGYTEHQVVSSDFNSDTHSSTFDAGAGIALNDHFVK

FIGURE S3

Supplementary methods

GeLC-MS

In-gel digestion workflow was adapted from the protocol described by Shevchenko *et al* [33]. After electrophoresis the gel slab is stained with Coomassie brilliant blue R-250 for 15 min at room temperature. After staining the gels are destained using 5:4:1 (v/v) of Water:Methanol:Acetic acid. The gel slice were cut to approximately 1 mm size or any gel slice size of interest. and transferred to a 1.5 ml LoBind Eppendorf tubes for further processing.

The gel pieces were destained completely by ACN/Water, and subsequently reduced by incubating the gels with 10 mM DTT at 56°C for 45 minutes. After reduction the proteins were alkylated using 55 mM of Iodoacetamide for 30 minutes in dark at room temperature. The reduced and alkylated samples were digested overnight with trypsin (10ng/µl) in 10 mM ammonium bicarbonate. After digestion the tryptic peptides were extracted using water/ACN/FA, dried under vacuum and stored at -20°C for further analysis.

The dried extracts were reconstituted in 5% FA and 5 µl was injected using the auto sampler into a Dionex Ultimate 3000 Nano-HPLC system equipped with a two column setup comprising of a trap column (5 mm × 300 µm i.d) and an analytical column (Acclaim PepMap100 C18 15 cm × 75 µm). Water with 0.1% (solvent A) and ACN with 0.1% (solvent B) were used as mobile phase. The samples were loaded in the trap column with a flow rate of 20 µl/min. The trap column was then switched to the analytical column with a flow rate of 200 nL/min for separation. The separation was carried out using Dionex Ultimate 3000- HPLC system (Thermo Scientific,

Bremen, Germany) running gradient elution program for 180 min and the output was hyphenated to Q-Exactive HF (Thermo Scientific, Bremen, Germany) and the mass spectra were acquired in data-dependent acquisition mode. The acquisition parameters are provided in (**Supplemental information, MS parameters**)

Expression and metabolic labelling of protein standards

We adapted the same expression workflow as described in [20]. Synthetic genes produced by GenScript (Piscataway NJ) were sub-cloned into pET expression vector and transformed into an *E.coli* strain that was dual auxotroph for arginine and lysine (ΔArgΔLysBL21 (DE3) T1 pRARE). The successful transformants were then diluted and sub cultured in MDAG-135 media [34]. They were induced by 0.2mM isopropyl β-d-1-thiogalactopyranoside (IPTG). After 4 to 6 hr post induction cells were pelleted, re-suspended in 2x phosphate-buffered saline (PBS). The suspended cells were then aliquoted, snap frozen in liquid nitrogen and stored at - 80°C until used for quantification.

Database search and data processing

Peptide matching was carried out using Mascot v.2.2.04 software (Matrix Science, London, UK) against *Homo sapiens* (August 2020) proteome downloaded from Uniprot. A precursor mass tolerance of 5ppm and fragment mass tolerance of 0.03 Da was applied, fixed modification: carbamidomethyl (C); variable modifications: acetyl (protein N terminus), oxidation (M); labels: $^{13}C(6)$ (K) and $^{13}C(6)^{15}N(4)$ (R) (only for MS Western the labelling was used); cleavage specificity: trypsin, with up to 2 missed cleavages allowed. Peptides having the ions score above 15 were accepted (significance threshold $p < 0.05$). The chromatographic alignment and feature

detection were carried using Progenesis LC-MS v.4.1 (Nonlinear Dynamics, UK). The absolute quantification was performed by calculating the abundances for the labelled and the unlabelled peptide using an in-house software as previously published [9]. The statistical analysis were carried out in OriginLab (2017) (OriginLab Corp., Northampton, Massachusetts, USA).