



HAL
open science

Data-Independent Acquisition Coupled to Visible Laser-Induced Dissociation at 473 nm (DIA-LID) for Peptide-Centric Specific Analysis of Cysteine-Containing Peptide Subset

Lény Garcia, Marion Girod, Magali Rompais, Philippe Dugourd, Christine Carapito, Jérôme Lemoine

► To cite this version:

Lény Garcia, Marion Girod, Magali Rompais, Philippe Dugourd, Christine Carapito, et al.. Data-Independent Acquisition Coupled to Visible Laser-Induced Dissociation at 473 nm (DIA-LID) for Peptide-Centric Specific Analysis of Cysteine-Containing Peptide Subset. *Analytical Chemistry*, 2018, 90 (6), pp.3928 - 3935. 10.1021/acs.analchem.7b04821 . hal-01767310

HAL Id: hal-01767310

<https://hal.science/hal-01767310v1>

Submitted on 4 Sep 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **Data Independent Acquisition Coupled to Visible Laser-Induced**
2 **Dissociation at 473 nm (DIA-LID) for Peptide-Centric Specific**
3 **Analysis of Cysteine-Containing Peptide Subset.**

4
5
6 Lény Garcia^a, Marion Girod^a, Magali Rompais^b, Philippe Dugourd^c, Christine Carapito^b and
7 Jérôme Lemoine^{a*}

8
9
10 ^a Université de Lyon, CNRS, Université Claude Bernard Lyon 1, Ecole Normale Supérieure
11 de Lyon, Institut des Sciences Analytiques, UMR 5280, 5 rue de la Doua, F-69100
12 Villeurbanne, France

13
14 ^b Laboratoire de Spectrométrie de Masse Bio-Organique (LSMBO), IPHC, Université de
15 Strasbourg, CNRS, UMR 7178, 25 rue Becquerel, 67087 STRASBOURG, France

16
17 ^c Université de Lyon, Université Claude Bernard Lyon 1, CNRS, Institut Lumière Matière, F-
18 69622, VILLEURBANNE, France

19
20 ***To whom correspondence should be addressed:**

21 Jérôme Lemoine: jerome.lemoine@univ-lyon1.fr

23 **ABSTRACT**

24 Thanks to comprehensive and unbiased sampling of all precursor ions, the interest to move
25 towards bottom-up proteomic with Data Independent Acquisition (DIA) is continuously
26 growing. DIA offers precision and reproducibility performances comparable to true targeted
27 methods, but has the advantage of enabling retrospective data testing with the hypothetical
28 presence of new proteins of interest. Nonetheless the chimeric nature of DIA MS/MS spectra
29 inherent to concomitant transmission of a multiplicity of precursor ions makes the confident
30 identification of peptides often challenging, even with spectral library-based extraction
31 strategy. The introduction of specificity at the fragmentation step upon ultraviolet or visible
32 laser-induced dissociation (LID) range targeting only the subset of cysteine-containing
33 peptides (Cys-peptide) has been proposed as an option to streamline and reduce the search
34 space. Here, we describe the first coupling between DIA and visible LID at 473 nm to test for
35 the presence of Cys-peptides with a peptide-centric approach. As a test run, a spectral library
36 was built for a pool of Cys-synthetic peptides used as surrogates of human kinases (1 peptide
37 per protein). By extracting ion chromatograms of query standard and kinase peptides spiked at
38 different concentration levels in an *Escherichia coli* proteome lysate, DIA-LID demonstrates
39 a dynamic range of detection of at least 3 decades and coefficients of precision better than 20
40 %. Finally, the spectral library was used to search for endogenous kinases in human cellular
41 extract.

42

43 **Keywords**

44 Laser-induced dissociation, data-independent acquisition, spectral library extraction, cysteine-
45 containing peptides, chromophore derivatization

46

47

48 **Introduction**

49 As an alternative to Data Dependent Acquisition (DDA), the Data Independent
50 Acquisition (DIA) technique pioneered by Yates and coll.¹, then popularized by Aebersold
51 and coll.² under the acronym SWATH (Sequential Window Acquisition of all Theoretical
52 fragment-ion spectra), has gained a great interest in proteomics in the past few years. In DIA
53 mode, a mass range of typically 400 to 1200 Th is segmented in small consecutive windows
54 of 8 to 25 Th through which all simultaneously transmitted precursor ions are fragmented to
55 form a composite MS/MS spectrum. When the duty cycle of concatenated windows is
56 optimized to ensure at least 8 data points per peak, DIA combines comprehensive and
57 unbiased peptide sampling with quantification metrics nearing those achieved with pure
58 targeted modes^{3,4}. By the use of peptide spectral libraries built either from DDA experiments
59 or synthetic banks and indexed with their respective relative retention times, hundreds to few
60 thousands of proteins⁵ may thus be confidently and reproducibly quantified by DIA at a
61 relatively high throughput level⁶⁻⁹. Moreover, DIA allows reanalysis of old datasets for testing
62 new biological hypothesis.

63 Despite aforementioned recognized figures of merit, the composite nature of
64 fragmentation spectra may also lead to false positive or negative identifications. This feature
65 is particularly exacerbated when large spectral libraries, i.e. the whole human spectra library⁵
66 are used to extract peptide/protein signals from data acquired on a particular biological
67 compartment such as urine for instance¹⁰. With the aim of reducing both DIA data complexity
68 and search space, laser-induced photodissociation (LID) could be an exquisite alternative to
69 introduce the lacking dimension of specificity at the fragmentation level. Interestingly, LID
70 implementation in commercial instruments is rather simple as it does not require any
71 hardware modification other than endowing the collision cell with a window. As an
72 illustration of LID specificity, infrared multiphoton photodissociation (IRMPD) at 10.6 μm

73 has shown interesting propensity to differentiate phospho- and non-phosphorylated peptides¹¹
74 due to increased photon absorption of the phosphate group. Moreover, IRMPD shows
75 specificity towards sulfonated peptides^{12,13}. The fragmentation specificity of peptides has been
76 similarly illustrated in the ultraviolet range (UVPD) by leveraging natural absorption of
77 intrinsic chromophores such as disulfide bonds or tyrosine residue¹⁴⁻¹⁶. Nonetheless, the most
78 obvious rationale to introduce fragmentation specificity lies in targeting only the subset of
79 peptides after grafting them with a proper chromophore^{17,18}. Among the 20 natural amino-
80 acids, cysteine is by far the best candidate for the development of such a targeted approach.
81 Indeed, if cysteine is present in almost all human proteins (91 % contain at least one), its
82 occurrence is very weak (3.3 %), which translates into low proportion (24%) of tryptic
83 peptides containing one or more cysteine residues¹⁹.

84 Furthermore, the sulfhydryl side chain group of cysteine is a highly reactive nucleophilic
85 group that condenses with alkyl halide or maleimide reagents, usually in a stoichiometric
86 proportion. Hence, many studies already reported site-selective peptide bond cleavages in the
87 vicinity of cysteine derivatized with one of the two radical initiators, quinone through UVPD
88 at 266 nm²⁰. The same chemistry operates with the thiol alkylating reagent N-
89 (phenylseleno)phtalimide, which initiates a characteristic neutral loss of 156 (benzeneselenol
90 moiety) only from Cys-peptides under selective 266 nm excitation. Then, laser-induced
91 neutral loss triggering CID MS/MS may afford selective sequencing of the cysteine peptide
92 subset (UVPDnLossCID) for streamlining the search space in a proteomic database²¹. Another
93 work from the same group describes the on-line LC-MS coupling with an excimer laser
94 emitting at 351 nm to pinpoint in an antibody trypsin digest only the Cys-peptides selectively
95 tagged with Alexa Fluor 350 on the basis of a reporter ion at m/z 296 and extensive series of b
96 and y fragment ions²². We previously reported such an informative pattern of dissociation
97 when irradiating Cys-peptides modified by a Dabcyl group in the visible range at 473 nm^{23,24},

98 which shown promising applications in the context of Multiple Reaction Monitoring
99 detection. Together, these studies illustrate the ability of carrying on UV or visible LID on a
100 chromatographic time scale and to increase the detection specificity in targeted or in
101 spectrum-centric approaches.

102 Considering the above mentioned merits of DIA but also its inherent limitation arising from
103 the composite nature of fragmentation spectra, we thus sought to implement and investigate
104 the performances of the coupling between LID at 473 nm and DIA mode following a peptide-
105 centric strategy for streamlining the detection of Cys-peptide surrogates in a spectral library.
106 In order to test this new methodology, a LID spectral library was first built from a synthetic
107 bank of 446 Cys-peptides derived from human kinases and that were previously identified as
108 MS-friendly candidates (all peptides had already been identified in the PeptideAtlas
109 repository). We next used this spectral library to identify the synthetic peptides spiked into an
110 *Escherichia coli* background proteome in order to systematically evaluate DIA-LID precision
111 and dynamic range of detection. Though our experimental set-up based on normal flow
112 chromatography is not optimal from a sensitivity point of view with regard to the nano flow
113 gold standard, we tentatively applied DIA-LID tool for targeting endogenous kinases in a total
114 human cell extract.

115

116

117 **Materials and methods**

118 Details about chemicals, samples, instrumental set-up and HPLC conditions are provided as
119 Supplemental Information (SI).

120

121 **Sample preparation**

122 The 446 synthetic Cys-peptides derived from human kinases were initially mixed to form 50
123 pools of stock solutions (9 peptides per pool on average) at a theoretical individual
124 concentration of 400 $\mu\text{g}/\text{mL}$ in $\text{H}_2\text{O}/\text{ACN}$ (50/50) + 0.5 % of FA. For spectral library
125 building, 2.5 μL of each stock solution pool were diluted with acidified water to reach an
126 estimated final concentration around 1 $\mu\text{g}/\text{mL}$. These pools were then mixed in a single pool,
127 reduced with a 3-fold molar excess solution of TCEP (500 μL of a 1 mg/mL in 60mM
128 ammonium acetate), then derivatized with 5-fold molar excess of Dabcyl C2 maleimide
129 chromophore (500 μL of a 1 mg/mL methanol solution). The mixture was stirred in a sonic
130 bath, then stored in the dark for 4 h at room temperature for complete conversion.
131 Subsequently, the peptides were desalted and concentrated using an Oasis™ HLB 3 cc (60
132 mg) reversed phase cartridge (Waters, Milford, MA, USA), as described in details
133 elsewhere.²³

134 To evaluate the detection and the identification of the cysteine containing peptides in DIA by
135 photodissociation, the 354 synthetic peptides retained in the spectral library were spiked at
136 theoretical amounts ranging from 17 ng to 14.15 μg in an *Escherichia coli* proteome extract
137 (100 μg). The sample was reduced and derivatized with the DABCYL C2 maleimide
138 chromophore as described above. The samples were diluted with 3 mL of a 60 mM AA
139 solution prior to overnight digestion at 37 °C with trypsin using a 1:20 (w/w) enzyme to
140 substrate ratio. Digestion was stopped by addition of formic acid to a final concentration of
141 0.5 %. All samples were desalted as mentioned previously and concentrated in 85 μL of
142 water/MeOH (95:5, v/v) containing 0.5 % FA.

143 For a relative estimation of the detection dynamic range of DIA-LID mode, a mixture of 65
144 peptides selected among those showing the most intense signals and covering a large retention
145 time distribution were spiked in an *Escherichia coli* protein extract with different dilution
146 factors to get normalized concentrations of 1, 0.1, 0.01 and 0.001. For the precise

147 quantification of the 3 purified custom peptides, 6 different points of calibration were carried
148 out for each peptide between 0.1 ng and 1000 ng in 100 μg of *Escherichia coli*. For the
149 detection of peptides in DIA-LID directly in biological matrix, HS578T human cellular
150 extracts were used. 60 μg of proteins were derivatized and analyzed. A preliminary Bradford
151 test was performed to estimate the protein concentration and to adjust the derivatization
152 protocol. Concentrations of 1177 $\mu\text{g}/\text{mL}$ and 340 $\mu\text{g}/\text{mL}$ were obtained for the *Escherichia*
153 *coli* sample and the human cellular extract, respectively. All of these experiments were
154 achieved following the aforementioned protocol (chromophore tagging, sample digestion and
155 SPE) for the specific detection of Cys-peptides by photodissociation.

156

157 **Mass spectrometry operating condition**

158 Ionization was achieved using electrospray in positive ionization mode with an ion spray
159 voltage of 4000 V. The sheath gas and auxiliary gas (nitrogen) flow rates were set at 35 and
160 10 (arbitrary unit), respectively, with a HESI vaporizer temperature of 300°C. The ion transfer
161 capillary temperature was 300 °C with a sweep gas (nitrogen) flow rate at 5 (arbitrary unit).
162 The S-lens RF was set at 50 (arbitrary unit). For building the spectral library, we used DDA
163 (Top10) and PRM (Parallel Reaction Monitoring) modes for the pool of synthetic derivatized
164 peptides. For DDA acquisitions, in full MS scan (FMS), the resolution was set to 35 000, the
165 Automatic Gain Control (AGC) target was $5 \cdot 10^6$ and the maximum injection time was set to
166 250 ms. In MS/MS mode, the resolution was set to 17 500, the AGC target was $5 \cdot 10^6$, the
167 maximum injection time was set to 120 ms and the quadrupole isolation width was 2 Th. In
168 PRM acquisitions, the resolution was set to 17 500, the AGC target was $5 \cdot 10^6$, the maximum
169 injection time was set to 60 ms and the quadrupole isolation width was 2 Th. Synthetic Cys-
170 peptides mixtures spiked in *Escherichia coli* background and total human cellular extracts
171 were analyzed using the DIA mode. The full MS scan was performed over a m/z range of 400

172 to 848 Th with a resolution set to 35 000, an AGC target of 3.10^6 and a maximum injection
173 time of 55 ms. This m/z range was divided into 14 windows of 32 Th for LID scans with a
174 resolution set to 35 000, an AGC target of 3.10^6 and a maximum injection time of 95 ms. For
175 LID experiments (PRM, DDA and DIA), the activation time was set to 25 ms. In order to
176 avoid CID, the collision energy was set to 2 eV. For the comparison with DIA-LID,
177 experiments were also done in DIA using the classical HCD activation mode. In this case, the
178 same m/z range was divided into 14 windows of 32 Th with a resolution set to 17 500, an
179 AGC target of 3.10^6 and a maximum injection time of 75 ms. For HCD experiments, the
180 activation time was set to 3 ms and the normalized collision energy (NCE) was set to 24.

181

182 **Spectral library generation and data processing**

183 MS/MS data acquired in DDA and PRM modes on the pools of synthetic Cys-containing
184 peptides were searched using the Mascot database search algorithm (version 2.5.1 run on a
185 local server, Matrix Science, London, UK) against a custom-built database containing the
186 synthetic peptide sequences assembled into artificial protein sequences. Full trypsin enzyme
187 specificity was used, no missed cleavage was allowed and the dabcyI maleimide modification
188 (+391.164 Da) was set as a fixed modification on Cys residues. Mass tolerance for precursor
189 ions was set at 5 ppm, and at 0.07 Da for fragment ions. Mascot .dat results files were loaded
190 into Proline software (<http://proline.profi-proteomics.fr/>²⁵) and peptides were validated solely
191 using spectrum quality criteria, namely ion scores higher than 30 and manual inspection of
192 fragmentation spectra. The highest scored peptide spectrum match (PSM) for each peptide
193 was retained and imported into Skyline software²⁶ to build the LID reference spectral library
194 that will be further used to extract signals (user-selected specific transitions) of the selected
195 Cys-containing peptides.

196 Details and validation criteria for targeted signal extraction from DIA data using the reference
197 spectral library are provided in the results section.

198

199 **Results and discussion**

200 Despite the use of spectral libraries including high quality reference fragmentation spectra and
201 precise relative retention time coordinates, the unambiguous assignment of peptides following
202 data independent acquisition (DIA) experiments remains challenging due to the composite
203 nature of fragmentation spectra. Hence, DIA coupled to laser-induced dissociation (DIA-LID)
204 is proposed as a complementary strategy to collisional-activation in the HCD cell (DIA-HCD)
205 by leveraging the high specificity of fragmentation of Cys-peptides grafted with a Dabcyl
206 chromophore in the visible range. As a first proof of concept, a LID spectral library was built
207 from a panel of synthetic Cys-peptide used as surrogates of human kinases and ATP-binding
208 proteins. The linearity, repeatability and detection dynamic range of the LID-DIA method
209 were evaluated using spiking experiments in an *Escherichia coli* background proteome, while
210 the transferability of the method was evaluated with the detection of endogenous proteins in a
211 total human cell lysate. The library generation and data processing workflow is illustrated by
212 the synopsis of Figure 1.

213

214 **Building of LID reference spectra library**

215 Thanks to mining of public proteomics data repositories (ProteomeXchange²⁷, SRMATlas²⁸),
216 446 peptide surrogates of human-related protein kinases (one per protein) containing a single
217 cysteine residue and between 9 and 25 amino-acids were selected for parallel chemical
218 synthesis. After grafting them with a Dabcyl C2 maleimide group, pools of peptides in pure
219 solvent condition were first submitted to standard Top10 DDA experiments without any
220 constrain on the charge state of selected precursor ions. The generated MS/MS data were

221 searched against a custom-built database containing the synthetic peptide sequences
222 assembled into artificial protein sequences. Peptides identified during this first database
223 search round with high quality fragmentation spectra and reasonable sequence coverage were
224 added to the reference LID spectra library built using Skyline software. DDA experiments
225 allowed validating 209 peptides. Then, to increase the chance of detecting the 237 remaining
226 non-detected peptides, a specific monitoring using targeted PRM acquisitions was carried out
227 to complete the LID reference spectra collection. LID reference spectra and relative retention
228 time coordinates were finally recovered for 354 peptides while 92 remained undetected,
229 mostly due to synthesis failure as it is often observed with low-grade synthetic libraries. Note
230 that this partial recovery was considered not detrimental for subsequent evaluation of LID-
231 DIA figure of merit. LID spectra were recorded after a constant irradiation time of 25 ms to
232 yield, in average, to 85 % depletion of precursor ion signal. All spectra exhibit a characteristic
233 reporter ion at m/z 252.113 arising from internal fragmentation of Dabcyl chromophore as
234 well as informative backbone fragmentations²⁹ (examples are shown in Figure S1A and B in
235 Supporting Information). The relative intensity of the reporter ion at m/z at 252.113
236 corresponds in average to 20 % of the total fragment ions intensity while the sum of singly
237 and doubly charged y_n , and b_n ion signals contribute to 51 and 29 %, respectively (Figure S2)
238 (these percentages were calculated from a random selection of 40 peptides). Importantly, the
239 combination of y_n and b_n fragments overall provide consistent peptide sequence coverage of
240 60 and 30 % for the entire library, which beside the fragmentation specificity is clearly
241 another useful feature for reliable peptide assignment at a given retention time.

242

243 **DIA-LID method development and evaluation**

244 In DIA mode, the acquisition process usually combines a MS scan followed by n
245 concatenated MS/MS isolation windows, which duty cycle completion needs to ensure the

246 record of at least 8 data points per chromatographic peak.³⁰ For DIA-LID experiments, 14
247 windows of 32 Th were recorded to cover the 400 to 848 m/z range of all precursor ions
248 present in the spectral library. For comparison of the relative merit of DIA-LID towards
249 conventional DIA-HCD condition, the 400 to 848 m/z range was similarly divided in 14
250 windows of 32 Th. In order to validate our spectral library-based signal extraction strategy,
251 the 354 Cys-peptides were spiked in a whole *Escherichia coli* background proteome prior to
252 chromophore grafting and trypsin digestion. Based on previous experiments and peptide
253 signal levels measured during the spectral library generation, the 354 peptides concentrations
254 were adjusted before being spiked into 100 μg of *Escherichia coli* proteins background.
255 Dilutions of peptide stock solutions were made in order to normalize individual peptide
256 signals around 10^7 cps for which 6 fragment ions were at least detected in the pure solvent
257 condition. Figures 2 and S3 (Supporting Information) exemplify archetypes of LID and HCD
258 fragmentation patterns recorded for equivalent DIA transmission windows at two given
259 retention times. For instance, Figure 2A and B displays the extracted fragmentation spectra of
260 m/z 432-464 DIA window acquired at retention time 6.69 min in LID and HCD regime,
261 respectively. In LID, though at least four precursor ions (isotope clusters labeled with stars on
262 the LID spectrum) have been simultaneously transmitted, they remain intact since no
263 noticeable fragments are detected on both sides of them (Figure 2A), which features the sole
264 transmission of cysteine-free peptides. On the contrary, its analog recorded in HCD mode
265 exhibits a complete depletion in the precursor ion window leading to a complex pattern of
266 fragment ions both in the high and low m/z regions (Figure 2B). Such alignment comparisons
267 are straightforward provided reproducible chromatographic behavior and thus retention times
268 are achieved, which was the case thanks to the use of a conventional flow coupling enabling
269 highly stable retention time with coefficients of variation (CV) <1.5 %. Figure 2C and D
270 provides the same comparison between LID and HCD but at a higher retention time (36.40

271 min). Most of the Cys-peptides are eluted later in the gradient (between 25 and 40 % of ACN)
272 due to the increased retention factor caused by the Dabcyl group²³. This chromatographic
273 behavior also contributes to extract them from the bulk of their cysteine-free congeners. At
274 this retention time, the characteristic signal of reporter ion at m/z 252.113 is detected both in
275 LID (Figure 2C) and HCD (Figure 2D) modes, indicating that at least one derivatized Cys-
276 peptide ion fell within the m/z 592-624 DIA window. The DIA-LID data processing using our
277 spectral library extraction approach identified peptide (NILWSAENECFK). It is noteworthy
278 that upon photoirradiation (Figure 2C), the reporter ion and specific backbone fragments
279 (y_{10}^{++} , y_9^{++} , b_2^+ , b_3^+ , b_4^+ and a_2^+ ions labeled with black triangles) related to the targeted peptide
280 sequence are the major species, while the concurrently transmitted cysteine-free peptides
281 remain intact (* labeled ions). In HCD mode (Figure 2D), the spectrum is more complex due
282 to co-fragmentation. Indeed, in contrast to LID for which only derivatized Cys-peptides are
283 specifically photo-fragmented, in HCD all cysteine-free peptides present in the DIA window
284 are also fragmented (yielding ions labeled with empty squares). The spectrum is dominated by
285 the intense reporter ion at m/z 252.113. This could be explained by preferential protonation on
286 the chromophore moiety which drives the fragmentation mechanism in HCD (charge-driven
287 fragmentation), while the energetic redistribution seems more efficient in LID.²⁹ As for
288 backbone fragments, due to intense contributions from other cysteine-free ions in the DIA
289 window, only a_2^+ , b_2^+ and b_3^+ fragment ions can be clearly assigned to the aforementioned
290 kinase related peptide. Together, all DIA-LID spectra validate the concept of streamlined
291 spectral library-based signal extraction using Skyline by leveraging the specific fragmentation
292 of the subset of Cys-peptides upon visible LID.

293 Overall, the spectral library-based extraction of DIA-LID data lead to the successful
294 identification of 272 peptides out of 354 (76 %) by applying the following validation filters: a
295 minimum of 3 detected transitions, a Dotp value (attesting for the correlation of relative

296 fragment intensities with those measured on the reference LID spectrum) equal or above 0.8
297 and a retention time tolerance of +/- 2.5 min (centered on the reference retention time
298 recorded in the spectral library). In fact, 96 % of the 272 peptides were detected with at least 4
299 of the most intense fragments contained in the spectral library. Applying these stringent
300 validation criteria for a control experiment consisting in the DIA-LID analysis of a non-
301 spiked *Escherichia coli* lysate derivatized with Dabcyl but free of human kinase-related
302 peptides led to no identification. This tight specificity of LID (Figures 3A, B and C)
303 compared to indiscriminant HCD mode (Figures 3D, E and F) is similarly highlighted by the
304 extracted ion chromatograms of fragment ions corresponding to three identified peptides. LID
305 allows significant depletion of interfering signals in the chromatograms over different DIA
306 windows. In HCD mode, the detection of noise signal resulting from co-fragmentation of
307 cysteine-free peptides leads to poor Dopt scores and difficult identification of peptides. After
308 thorough manual inspection, 14 peptides out of the 82 missing candidates were putatively
309 retrieved after enlarging the retention time tolerance to +/- 6 min. This result is likely the
310 consequence of a retention time shift caused by a matrix effect, which strengthens the
311 proposal of DIA workflows based on sample type-specific spectral libraries¹⁰. No signal could
312 be assigned for any of the remaining lacking 68 peptides that may be attributed to lower
313 concentration than estimated for some peptides in the synthetic library or signal suppression
314 caused by the matrix.

315

316 **Response curve, linearity and sensitivity of Cys-peptides detection with DIA-LID**

317 In order to assess the dynamic range of detection and precision of our DIA-LID
318 method, we selected a subset of 65 peptides based on their signal intensity and retention time
319 distribution and set up a new concentration-balanced mixture that was spiked into 100 μg of
320 *Escherichia coli* total lysate. Since the true peptide concentrations were not accessible for

321 those crude peptides, the dilution curves were normalized to 1, 0.1, 0.01 and 0.001 where the
322 value of 1 corresponds to adjusted concentrations of the stock solutions as described above.
323 By analyzing the results from injection triplicates, both slopes and coefficients of
324 determination R^2 values of the response curves demonstrated a linearity over 3 decades with
325 all R^2 better than 0.95 (Table S1). Imprecision values of peak areas expressed by the
326 coefficient of variation (CV) were better than 20 % for 94 % of the concentration points
327 (Figure 4 and Table S1). In order to better evaluate the intrinsic sensitivity of our DIA-LID
328 experimental set-up, here based on a first generation of Q-Exactive Orbitrap mass
329 spectrometer, response curves were acquired in triplicate for 3 custom peptides (purity of 70
330 %) with concentrations spanning 0.1 to 1000 ng spiked in 100 μ g of *Escherichia coli*
331 proteome. Linearity was observed between 1 ng and 1000 ng ($R^2 = 0.99$) both for
332 VMCDSNDLPK and DICNDVLSLLEK peptides, and between 5 ng and 1000 ng ($R^2 =$
333 0.9824) for YMPQNPCIIATK peptide (Figure S4). Again, the CVs were all better than 20 %
334 (Table S2). Assuming that the limit of quantification (LOQ) corresponds to the lowest
335 concentration of the linearity range for which a CV better than 20 % is achieved, we could
336 thus express these LOQ at 210, 170 and 840 femtomoles injected on column. By
337 extrapolation, the LOQ that could be then reasonably expected by shifting from the present
338 normal flow (300 μ L/min) to a more sensitive nano flow format (200 nL/min) should attain
339 the low femtomole on column range, in line with the recently published multisite comparison
340 of SWATH-MS performances⁹.

341

342 **DIA-LID applicability**

343 Overall, the DIA-LID metrics assessed by response curves reveals the validity of DIA-
344 LID approach for selective, precise and sensitive detection of Cys-peptides in a complex
345 proteome. DIA-LID reliability was ultimately investigated by tracking endogenous kinases in

346 the HS578T human cell line. From triplicate analyses, 5 kinase-related peptides were
347 confidently identified (Table S3 and Figure 5) with acceptable precision as coefficients of
348 variation of peak area calculated on the most intense fragment were all better than 25 %. In
349 the first instance, this identification rate may appear relatively disappointing but could have
350 been in part anticipated due to the adverse normal flow chromatographic configuration used
351 in the present study. Indeed, the increased sensitivity of the nano flow standard, suitable for
352 extensive detection of weakly abundant proteins, is the unique format compatible with all
353 miniaturized kinase enrichment tools required for human kinome assay³¹⁻³³. Nonetheless, a
354 focus on the extracted ion chromatogram of the reporter ion at m/z 252.113 between RT 27
355 and 35 min. (Figure S5) and randomly selected DIA MS-MS spectra of different m/z
356 transmission windows give us a glimpse on the potential of DIA-LID to massively and
357 specifically detect Cys-peptides. Interestingly, DIA-LID could also insert into a proteomic
358 workflow integrating prior chemoselective capture of Cys-peptides, which was already
359 proved highly efficient in enhancing the detection of low abundance proteins, hence in
360 expanding proteome coverage^{34,35}. To reach that goal, the next step will be focused on
361 spectral library expansion towards two peptides for all human proteins bearing at least two
362 cysteine residues.

363

364 **Conclusion**

365 We have herein documented a streamlined, peptide-centric detection of Cys-peptides
366 by combining visible laser induced dissociation with Data Independent Acquisition (DIA-
367 LID). Collectively, the response curve experiments advance the robustness of the entire DIA-
368 LID workflow, as highlighted by good CVs metrics over a dynamic range of 3 orders of
369 magnitude. The present DIA-LID configuration foreshadows a range of detection sensitivity
370 in line with more specifically dedicated instrumental set-ups mostly running in nanoLC

371 configuration, deployed today for DIA studies in proteomics. In this regard, further work
372 needs to be done to expand the spectral library to a second Cys-peptide per kinase-related
373 protein and ultimately to all human protein sequence harboring cysteine residues and to
374 endow the present mass spectrometer with a chromatographic format compatible with kinase
375 enrichment tools. Such a configuration is indeed mandatory to attain the detection sensitivity
376 level required for proteomics, especially when the sample amounts are limited. Overall, DIA-
377 LID is at this stage not intended to replace conventional DIA-CID but could provide a
378 complementary tool to streamline the detection of a subset of proteins or peptides whose
379 detection faces specificity issues.

380

381

382 **Supporting Information**

383 The Supporting Information is available free of charge on the ACS Publications website:

384 Additional example of MS/MS spectra, analytical performance levels (repeatability and
385 linearity) and endogenous kinases detected in a tryptic digest of HS578T cell whole proteome.

386

387

388 **Acknowledgment**

389 The authors would like to thank Diane Cala and Ruth Rimokh for providing the *Escherichia*
390 *coli* and human HS578T cells, as well as Joelle Saulnier for her assistance in protein
391 concentration measurements. This work was supported by the French Proteomic
392 Infrastructure ProFI (ANR-10-INBS-08-03).

393

394 **Conflict of interests**

395 The authors declare no conflict of interest

396

397 **Data access**

398 DIA-LID data are available via PeptideAtlas (<http://www.peptideatlas.org/PASS/PASS01120>)

399

400 **Author ORCID number**

401 Jérôme Lemoine: 0000-0002-5497-6880

402 Marion Girod: 0000-0003-0728-2111

403 **References**

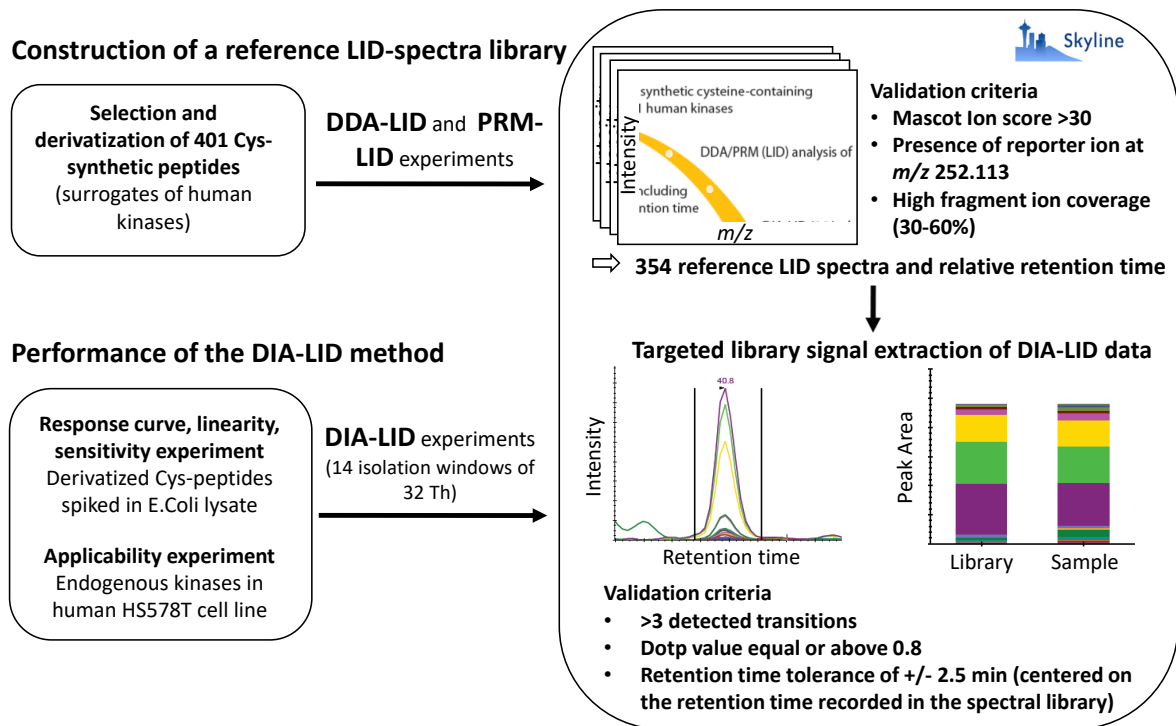
- 404 (1) Venable, J. D.; Dong, M.-Q.; Wohlschlegel, J.; Dillin, A.; Yates, J. R. *Nat. Methods*
405 **2004**, *1* (1), 39–45.
- 406 (2) Gillet, L. C.; Navarro, P.; Tate, S.; Rost, H.; Selevsek, N.; Reiter, L.; Bonner, R.;
407 Aebersold, R. *Mol. Cell. Proteomics* **2012**, *11* (6), O111.016717.
- 408 (3) Liu, Y.; Huttenhain, R.; Collins, B.; Aebersold, R. *Expert Rev. Mol. Diagn.* **2013**, *13*
409 (8), 811–825.
- 410 (4) Sajic, T.; Liu, Y.; Aebersold, R. *Proteomics. Clin. Appl.* **2015**, *9* (3–4), 307–321.
- 411 (5) Rosenberger, G.; Koh, C. C.; Guo, T.; Rost, H. L.; Kouvonon, P.; Collins, B. C.;
412 Heusel, M.; Liu, Y.; Caron, E.; Vichalkovski, A.; Faini, M.; Schubert, O. T.; Faridi, P.;
413 Ebhardt, H. A.; Matondo, M.; Lam, H.; Bader, S. L.; Campbell, D. S.; Deutsch, E. W.;
414 Moritz, R. L.; Tate, S.; Aebersold, R. *Sci. data* **2014**, *1*, 140031.
- 415 (6) Lambert, J.-P.; Ivosev, G.; Couzens, A. L.; Larsen, B.; Taipale, M.; Lin, Z.-Y.; Zhong,
416 Q.; Lindquist, S.; Vidal, M.; Aebersold, R.; Pawson, T.; Bonner, R.; Tate, S.; Gingras,
417 A.-C. *Nat. Methods* **2013**, *10* (12), 1239–1245.
- 418 (7) Guo, T.; Kouvonon, P.; Koh, C. C.; Gillet, L. C.; Wolski, W. E.; Rost, H. L.;
419 Rosenberger, G.; Collins, B. C.; Blum, L. C.; Gillessen, S.; Joerger, M.; Jochum, W.;
420 Aebersold, R. *Nat Med* **2015**, *21* (4), 407–413.
- 421 (8) Liu, Y.; Buil, A.; Collins, B. C.; Gillet, L. C. J.; Blum, L. C.; Cheng, L.-Y.; Vitek, O.;
422 Mouritsen, J.; Lachance, G.; Spector, T. D.; Dermitzakis, E. T.; Aebersold, R. *Mol.*
423 *Syst. Biol.* **2015**, *11* (1), 786.
- 424 (9) Collins, B. C.; Hunter, C. L.; Liu, Y.; Schilling, B.; Rosenberger, G.; Bader, S. L.;
425 Chan, D. W.; Gibson, B. W.; Gingras, A.-C.; Held, J. M.; Hirayama-Kurogi, M.; Hou,
426 G.; Krisp, C.; Larsen, B.; Lin, L.; Liu, S.; Molloy, M. P.; Moritz, R. L.; Ohtsuki, S.;
427 Schlapbach, R.; Selevsek, N.; Thomas, S. N.; Tzeng, S.-C.; Zhang, H.; Aebersold, R.
428 *Nat. Commun.* **2017**, *8* (1), 291.
- 429 (10) Muntel, J.; Xuan, Y.; Berger, S. T.; Reiter, L.; Bachur, R.; Kentsis, A.; Steen, H. J.
430 *Proteome Res.* **2015**, *14* (11), 4752–4762.
- 431 (11) Crowe, M. C.; Brodbelt, J. S. *Anal. Chem.* **2005**, *77* (17), 5726–5734.
- 432 (12) Borotto, N. B.; McClory, P. J.; Martin, B. R.; Hakansson, K. *Anal. Chem.* **2017**, *89*
433 (16), 8304–8310.
- 434 (13) Wilson, J. J.; Brodbelt, J. S. *Anal. Chem.* **2006**, *78* (19), 6855–6862.
- 435 (14) Agarwal, A.; Diedrich, J. K.; Julian, R. R. *Anal. Chem.* **2011**, *83* (17), 6455–6458.

- 436 (15) Antoine, R.; Dugourd, P. *Phys. Chem. Chem. Phys.* **2011**, *13* (37), 16494–16509.
- 437 (16) Joly, L.; Antoine, R.; Broyer, M.; Dugourd, P.; Lemoine, J. *J. Mass Spectrom.* **2007**, *42*
438 (6), 818–824.
- 439 (17) Aponte, J. R.; Vasicek, L.; Swaminathan, J.; Xu, H.; Koag, M. C.; Lee, S.; Brodbelt, J.
440 S. *Anal. Chem.* **2014**, *86* (13), 6237–6244.
- 441 (18) O’Brien, J. P.; Pruet, J. M.; Brodbelt, J. S. *Anal. Chem.* **2013**, *85* (15), 7391–7397.
- 442 (19) Lin, D.; Li, J.; Slebos, R. J. C.; Liebler, D. C. *J. Proteome Res.* **2010**, *9* (10), 5461–
443 5472.
- 444 (20) Diedrich, J. K.; Julian, R. R. *Anal. Chem.* **2010**, *82* (10), 4006–4014.
- 445 (21) Parker, W. R.; Holden, D. D.; Cotham, V. C.; Xu, H.; Brodbelt, J. S. *Anal. Chem.*
446 **2016**, *88* (14), 7222–7229.
- 447 (22) Cotham, V. C.; Wine, Y.; Brodbelt, J. S. *Anal. Chem.* **2013**, *85* (11), 5577–5585.
- 448 (23) Girod, M.; Biarc, J.; Enjalbert, Q.; Salvador, A.; Antoine, R.; Dugourd, P.; Lemoine, J.
449 *Analyst* **2014**, *139* (21), 5523–5530.
- 450 (24) Enjalbert, Q.; Girod, M.; Simon, R.; Jeudy, J.; Chirot, F.; Salvador, A.; Antoine, R.;
451 Dugourd, P.; Lemoine, J. *Anal. Bioanal. Chem.* **2013**, *405* (7), 2321–2331.
- 452 (25) Carapito, C.; Lane, L.; Benama, M.; Opsomer, A.; Mouton-Barbosa, E.; Garrigues, L.;
453 Gonzalez de Peredo, A.; Burel, A.; Bruley, C.; Gateau, A.; Bouyssie, D.; Jaquinod, M.;
454 Cianferani, S.; Burret-Schiltz, O.; Van Dorsselaer, A.; Garin, J.; Vandenbrouck, Y. *J.*
455 *Proteome Res.* **2015**, *14* (9), 3621–3634.
- 456 (26) MacLean, B.; Tomazela, D. M.; Shulman, N.; Chambers, M.; Finney, G. L.; Frewen,
457 B.; Kern, R.; Tabb, D. L.; Liebler, D. C.; MacCoss, M. J. *Bioinformatics* **2010**, *26* (7),
458 966–968.
- 459 (27) Deutsch, E. W.; Csordas, A.; Sun, Z.; Jarnuczak, A.; Perez-Riverol, Y.; Ternent, T.;
460 Campbell, D. S.; Bernal-Llinares, M.; Okuda, S.; Kawano, S.; Moritz, R. L.; Carver, J.
461 J.; Wang, M.; Ishihama, Y.; Bandeira, N.; Hermjakob, H.; Vizcaino, J. A. *Nucleic*
462 *Acids Res.* **2017**, *45* (D1), D1100–D1106.
- 463 (28) Kusebauch, U.; Campbell, D. S.; Deutsch, E. W.; Chu, C. S.; Spicer, D. A.; Brusniak,
464 M.-Y.; Slagel, J.; Sun, Z.; Stevens, J.; Grimes, B.; Shteynberg, D.; Hoopmann, M. R.;
465 Blattmann, P.; Ratushny, A. V.; Rinner, O.; Picotti, P.; Carapito, C.; Huang, C.-Y.;
466 Kapousouz, M.; Lam, H.; Tran, T.; Demir, E.; Aitchison, J. D.; Sander, C.; Hood, L.;
467 Aebersold, R.; Moritz, R. L. *Cell* **2016**, *166* (3), 766–778.
- 468 (29) Garcia, L.; Lemoine, J.; Dugourd, P.; Girod, M. *Rapid Commun. Mass Spectrom.* **2017**,

- 469 31 (23), 1985-1992.
- 470 (30) Bruderer, R.; Bernhardt, O. M.; Gandhi, T.; Xuan, Y.; Sonderrmann, J.; Schmidt, M.;
471 Gomez-varela, D.; Reiter, L. *Mol. Cell. Proteomics* **2017**, No. 34, 2296–2309.
- 472 (31) Worboys, J. D.; Sinclair, J.; Yuan, Y.; Jorgensen, C. *Nat Meth* **2014**, *11* (10), 1041–
473 1044.
- 474 (32) Xiao, Y.; Wang, Y. *Mass Spectrom. Rev.* **2016**, *35* (5), 601–619.
- 475 (33) McAllister, F. E.; Niepel, M.; Haas, W.; Huttlin, E.; Sorger, P. K.; Gygi, S. P. *Anal.*
476 *Chem.* **2013**, *85* (9), 4666–4674.
- 477 (34) Wang, S.; Zhang, X.; Regnier, F. E. *J. Chromatogr. A* **2002**, *949* (1–2), 153–162.
- 478 (35) Guo, J.; Gaffrey, M. J.; Su, D.; Liu, T.; Camp, D. G.; Smith, R. D.; Qian, W.-J. *Nat.*
479 *Protoc.* **2014**, *9* (1), 64–75.
- 480
- 481

482 **Figures**

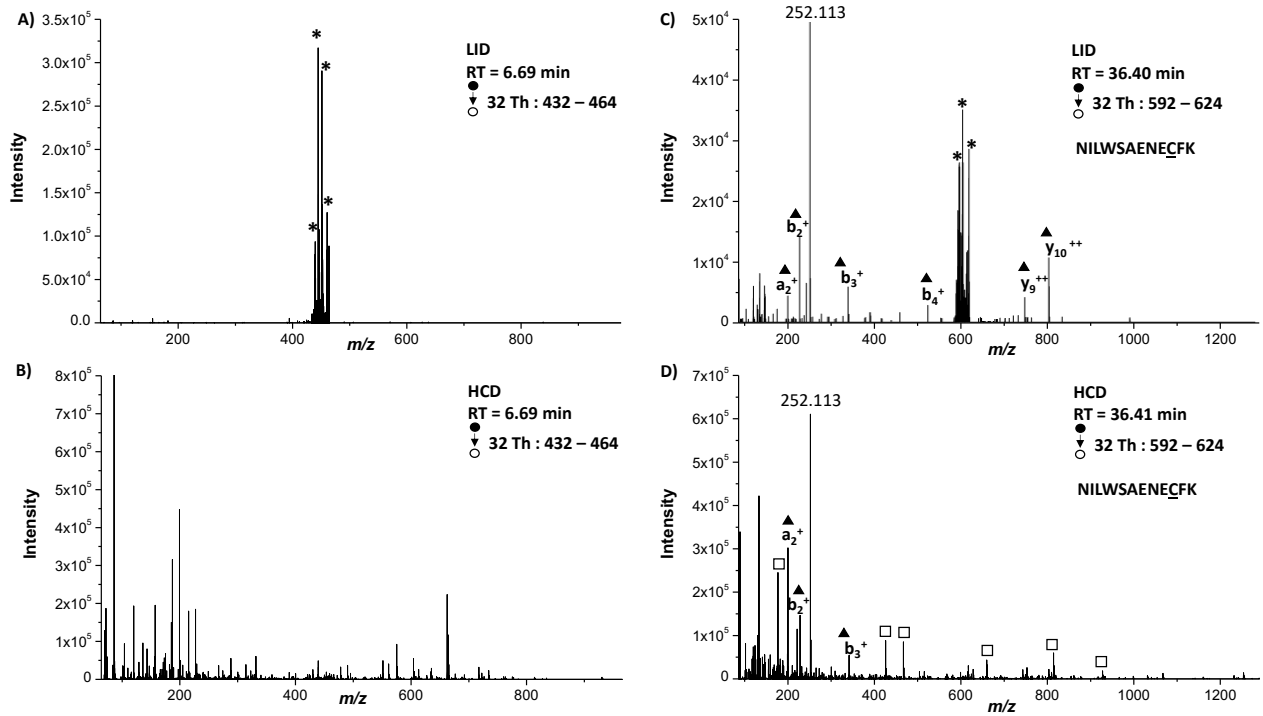
483



484

485 **Figure 1.** Workflow for LID reference spectra library generation and evaluation of the DIA-
486 LID method based on targeted signal extraction of Cys-peptides grafted with a Dabcyl
487 maleimide group.

488



489

490 **Figure 2.** Example of DIA-MS/MS spectra at retention time (RT) 6.69 and 36.40 min

491 recorded either in LID (A and C) or HCD mode (B and D) during the analysis of 354 Cys-

492 peptides derived from human kinase-related proteins spiked in an *Escherichia coli* proteome

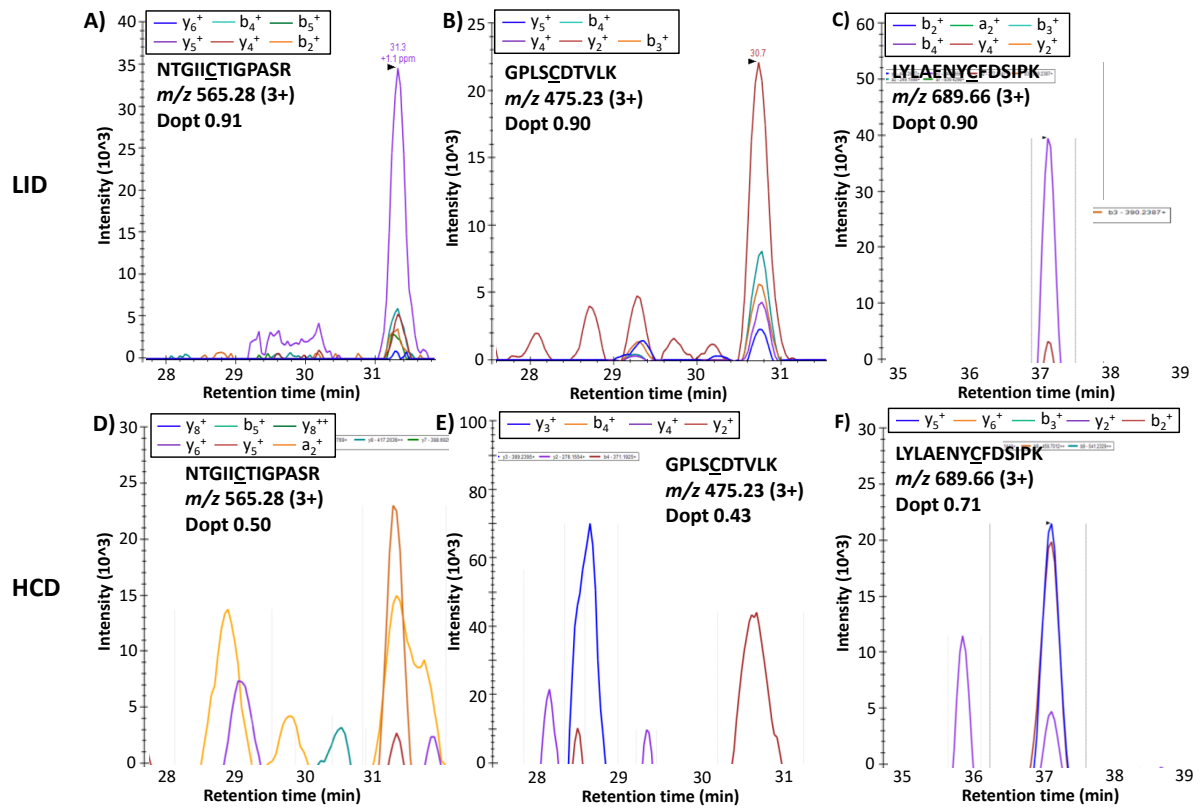
493 prior to trypsin digestion and dabcyI chromophore derivatization. (*) indicates peptide ions

494 free of cysteine residue, (▲) highlights specific Cys-peptide fragment ions and (□)

495 highlights fragment ions from cysteine-free peptides.

496

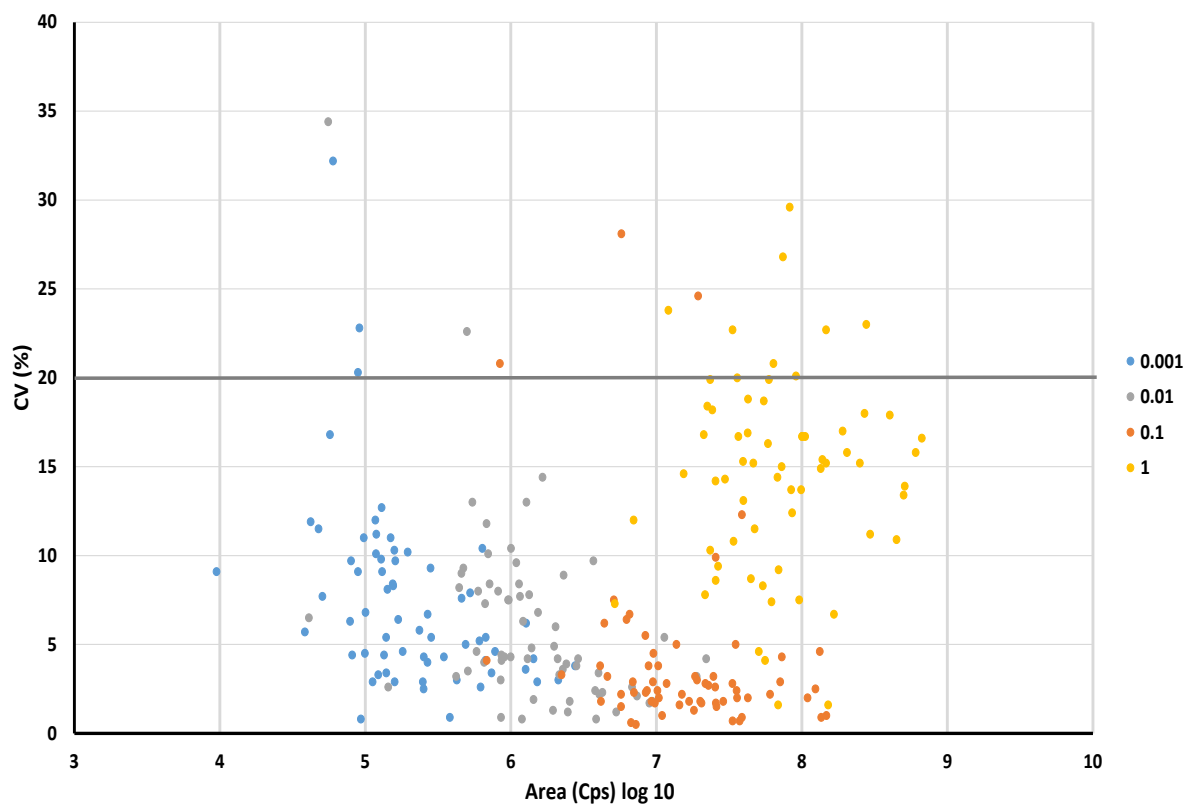
497



499

500 **Figure 3.** Extracted ion chromatograms of fragment ions from 3 derivatized Cys-peptides
 501 spiked in an *Escherichia coli* lysate monitored in DIA-LID (A, B, C) and DIA-HCD (D, E,
 502 F).

503



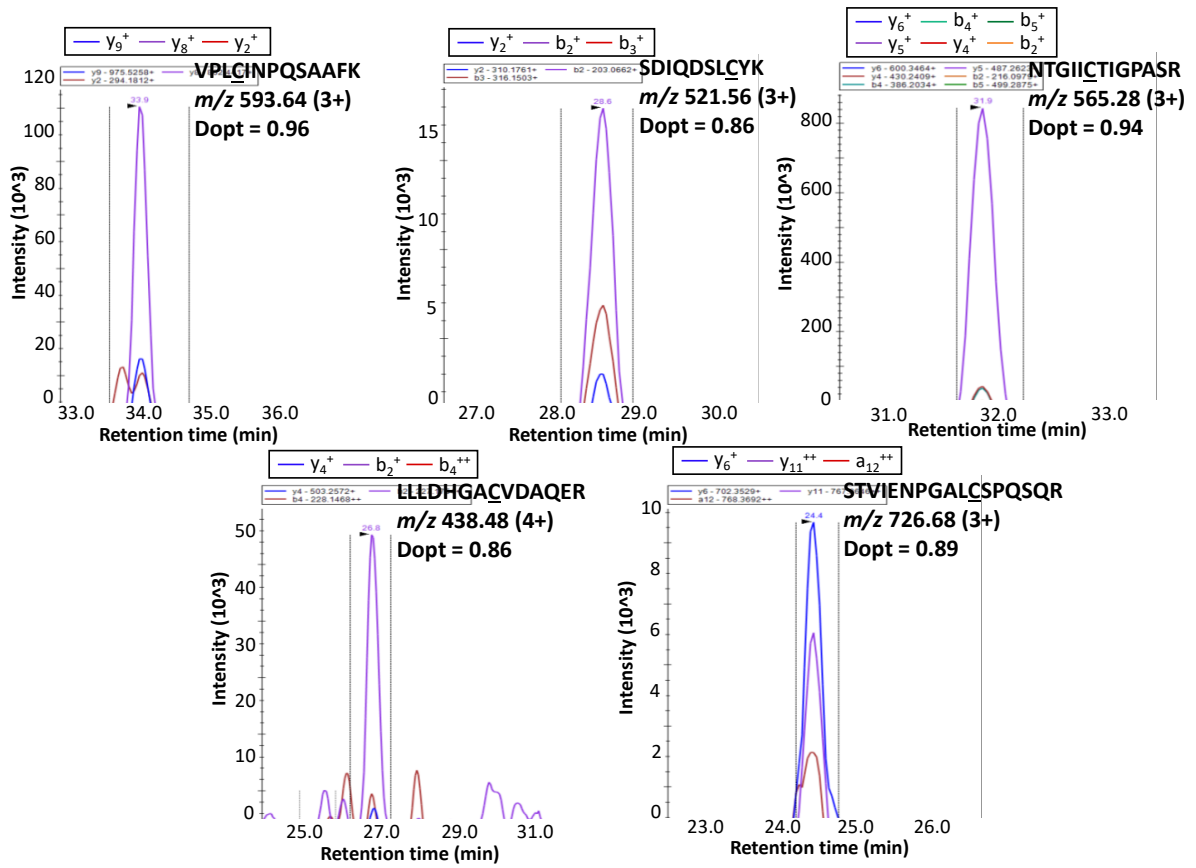
504

505 **Figure 4.** Coefficient of variation (CV) of peak areas (most intense transition) calculated for
 506 triplicate dilution points (1, 0.1, 0.01 and 0.001) for 65 derivatized Cys-peptides spiked into
 507 an *Escherichia coli* lysate.

508

509

510



511

512 **Figure 5.** Extracted ion chromatograms of all fragment ions detected for 5 Cys-peptide
 513 surrogates of endogenous kinases detected in a tryptic digest of HS578T cell whole proteome
 514 in DIA-LID.

515

