Systematic characterization of site-specific proline hydroxylation using hydrophilic interaction chromatography and mass spectrometry

Hao Jiang[1†], Jimena Druker[1†], James W. Wilson[2], Dalila Bensaddek[3], Jason R. Swedlow[1], Sonia Rocha[2], Angus I. Lamond[1*]

[1] Division of Molecular, Cell and Developmental Biology, School of Life Sciences, University of Dundee, Dundee, United Kingdom

[2] Institute of Systems, Molecular and Integrative Biology, University of Liverpool, Liverpool, United Kingdom

[3] Bioscience core lab, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia

[†] The authors contributed equally to this work.

[*] To whom correspondence may be addressed.

Abstract

We have developed a robust workflow for the identification of proline hydroxylation sites in proteins, using a combination of hydrophilic interaction chromatography (HILIC) enrichment and high-resolution nano-Liquid Chromatography-Mass Spectrometry (LC-MS) together with refining and filtering parameters during data analysis. Using this approach, we have combined data from cell lines treated with either the prolyl hydroxylase (PHD) inhibitor, Roxadustat (FG-4592), or with the proteasome inhibitor MG-132, or with a DMSO control, to identify a total of 4,993 and 3,247 proline hydroxylation sites, respectively, in HEK293 and RCC4 cells. A subset of 1,954 (HEK293) and 1,253 (RCC4) non-collagen proline hydroxylation sites with high confidence were inhibited by FG-4592 treatment. Features characteristic of proline hydroxylated peptides were identified, which were consistent between the HEK293 and RCC4 datasets. The more hydrophilic HILIC fractions were enriched in peptides containing hydroxylated proline residues, which showed characteristic differences in charge and mass distribution, as compared with either unmodified, or oxidised peptides. Furthermore, we discovered that the intensity of the diagnostic hydroxyproline immonium ion was dependent upon parameters including the MS collision energy

34 setting, parent peptide concentration and the sequence of amino acids adjacent to the
35 modified proline. We show using synthetic peptides that a combination of retention
36 time in LC and optimised MS parameter settings allows reliable identification of proline
37 hydroxylation sites in peptides, even when multiple prolines residues are present.
38 Focussing on proteins in which newly identified proline hydroxylation sites were
39 inhibited by the pan-PHD inhibitor, FG-4592, showed enrichment for proteins involved
40 in metabolism of RNA, mRNA splicing and cell cycle regulation, including the protein
41 phosphatase 1 regulatory subunit, Repo-Man (CDCA2). We show that Repo-Man is
42 hydroxylated at P604 and in the accompanying study, [Druker et al., 2025], we present
43 a combination of cellular and biochemical evidence that hydroxylation of Repo-Man at
44 P604 is important for its function in controlling mitotic progression.
45

Introduction

Site-specific hydroxylation of proline residues is an important enzymatic protein post-translational modification (PTM), which is mediated by prolyl hydroxylase (PHD) enzymes. For example, PHDs are essential for controlling HIF-α protein levels via their proline hydroxylation activity, in the presence of iron, 2-oxoglutarate and molecular oxygen [1-3]. HIFs are key transcription factors that regulate the acute and adaptive responses to changes in oxygen availability through a variety of target genes involved in metabolism, proliferation, cell motility and cell death [4-7]. Proline hydroxylation in HIF-α leads to an increased affinity with the VHL E3-ligase complex, K48-ubiquitination and proteasomal degradation [8-10].

Until recently, the HIF-α family was regarded as the only verified target for the PHD family of enzymes. Cummins et al [11] suggested that PHDs could also regulate the kinase activity of IKK enzymes, based upon functional assays in cells, but did not provide a formal demonstration mapping sites of proline hydroxylation in the IKK complex. More recently, several studies have reported other potential protein targets for PHDs [11-19], based upon either biochemical, and/or in-cell functional analyses, including mutations, and, in some cases, also mass spectrometry-based proteomics analysis [20,21].

The use of MS-based proteomics analysis is widely established for the identification and characterisation of multiple PTMs, including phosphorylation [22-24], methylation [25], ubiquitination [26], glycosylation [27,28] and others [29-32]. The benefit of using MS is that many modification sites can be identified and quantified in a robust and unbiased way[30]. However, so far it has proven difficult to reliably identify sites of proline modification in proteins, outside the collagen family, using MS methods. This technical challenge has contributed to a degree of controversy in the field regarding whether there are any intracellular targets for PHD-mediated hydroxylation beyond the HIF family of proteins. For example, a study by Cockman et al [33] reported that purified PHD enzymes failed to recapitulate proline hydroxylation *in vitro* at sites previously identified using mass spectrometry on proteins isolated from cells. They proposed, therefore, that HIF-α may be the only physiologically relevant target for PHDs. However, an alternative interpretation for their observations could be that technical limitations with the *in vitro* assays may be responsible for the failure to detect hydroxylation of non-HIF target proteins using purified PHDs.

The controversy surrounding whether there exist in cells *bona fide* PHD target proteins outside of the HIF family highlights several technical problems that complicate the

identification of sites of enzymatic proline hydroxylation by MS. These include, (i) the lack of validated methods for enriching proline-hydroxylated peptides during sample preparation for MS and (ii) the risk of misidentification of hydroxylated proline, due to potential confusion with oxidation of Methionine (M), which can occur as a non-enzymatic, chemical modification of peptides during sample preparation. Other technical issues include possible confusion of hydroxylated proline with Leucine/Isoleucine (L/I) residues, (which have similar mass to Pro-OH), the frequent absence in spectra of fragment ions diagnostic for Pro-OH and other problems associated with low-quality MS spectra.

In previous studies using an MS-based proteomics workflow to identify proline hydroxylation sites, antibodies that recognise hydroxyproline were used to enrich either whole proteins, or peptides, prior to MS analysis [34,35], similar to antibody-based strategies commonly used to enrich other types of protein modifications for MS-based analysis. However, both the unreliability of currently available anti-hydroxyproline antibodies and the relatively low abundance of proline hydroxylated peptides at steady state, has, in practice, limited the success of this approach.

The hydroxylation of proline, which adds a hydroxyl (OH) group to the gamma carbon atom, significantly alters its chemical properties (Figure 1A). The pKa value in free proline, (Pro or P), is 10.8, while for hydroxylated Proline, (HyPro or P-OH), it is 9.68[36]. Hence, following proline hydroxylation, the resulting modified peptides will be more hydrophilic in acidic solution, which can be utilized for liquid chromatography (LC)-based separation.

In this study, we report the development of a robust approach for the enrichment and reliable identification of proline hydroxylation sites in proteins using an MS workflow that incorporates a chromatographic peptide enrichment step, using HILIC. This takes advantage of the increase in hydrophilicity that results from the addition of the hydroxyl group to proline. HILIC is known to be capable of separating polar compounds [37], including peptides [38,39] and carbohydrates [40] and has been reported previously to allow enrichment of hydroxylated peptides [41]. We show that by combining HILIC enrichment with optimised MS instrument parameters and data analysis procedures, sites of proline hydroxylation can be reliably identified and subsequently verified by the analysis of synthetic peptides. Using this approach, we report the identification of 4,993 proline hydroxylation sites in HEK293 cell extracts and 3,247 sites in RCC4 cell extracts, along with analysis of common motifs found at sites of modification and classes of target

proteins harbouring hydroxylated proteins. In the accompanying study [Druker et al., 2025], we present functional data showing that the newly identified hydroxylation of Repo-Man (CDCA2) at proline 604 is important for its role in the control of mitotic progression, supporting a wider role for PHD enzymes in regulating cell cycle and potentially other cell functions.

Results

First, we tested the performance of HILIC in separating synthetic peptides of identical length and sequence that were either unmodified, hydroxylated on histidine, or hydroxylated on proline. As shown in Figure 1B, proline hydroxylation increased peptide retention time on the HILIC column and thus proline hydroxylated peptides can be separated by HILIC from their corresponding unmodified form, while peptides with hydroxylation at different amino acids can also be separated.

The HILIC fractionation method was then applied to the analysis of tryptic peptide samples, generated from extracts of both RCC4 and HEK293 cell lines. Extracts were compared from cells treated with either the prolyl hydroxylase inhibitor, FG-4592, or with the proteasome inhibitor MG-132, to prevent degradation of proline hydroxylated proteins, with DMSO treated cell extracts used as a control. The chromatography profile, (Figure 1C, shadow green), demonstrates the reproducibility of HILIC fractionation. In total, 32 peptide fractions were collected, dried and directly used for LC-MS/MS analysis.

Using this HILIC enrichment strategy, by combining the results from the DMSO, FG-4592 and MG-132 datasets, a total of 3,588 and 5,491 distinct proline hydroxylation sites, (Table S1), were initially identified from the RCC4 and HEK293 samples, respectively. Next, by considering several unique features of HyPro, we established specific steps during data analysis, distinct from methods typically used for other PTM analysis, to enhance accurate identification of HyPro sites.

HILIC fractionation distinguished proline hydroxylated peptides from oxidised peptides

MS analysis software usually identifies post-translational peptide modifications through the detection of the mass shift of a peptide and its corresponding fragment ions as a result of the modification. However, the hydroxylation of a proline equates to a mass shift of 15.994915 Da, which is similar to the mass shift seen for the oxidation of several other amino acids. For example, the oxidation of methionine residues is especially pertinent as it represents the amino acid on which chemical oxidation occurs most frequently. Indeed, methionine oxidation can either happen endogenously in cells, as a protection mechanism from oxidative damage on other reactive residues critical to the function of the protein [42,43], or it can occur exogenously, during sample preparation for MS analysis [44]. Hence, one of the most controversial issues around the identification

153 of proline hydroxylation sites is the possibility of confusing chemical oxidation on
154 methionine, or other amino acid residues, for proline hydroxylation [33].
155 As shown above using synthetic peptides, HILIC chromatography has the potential to
156 separate the peaks of proline hydroxylated peptides from others. We therefore
157 compared the percentage/distribution of peptides identified as either unmodified,
158 methionine oxidised, or proline hydroxylated, across 32 HILIC fractions derived from
159 both HEK293 (Figure 2) and RCC4 (Figure S1) cell extracts. From the initial results of
160 HEK293 samples, most unmodified peptides were identified in fractions F1-F20
161 (Figure 2A), with oxidised peptides showing a similar pattern, but more enriched in
162 fractions F4-F14. In contrast, hydroxylated peptides were enriched in HILIC fractions
163 F13-F19, consistent with their increased hydrophilicity, with the addition of a hydroxyl
164 group causing increased retention time on HILIC. A similar HILIC fraction distribution
165 pattern was also seen with the samples from RCC4 cell extracts (Figure S1A).
166 Furthermore, the data also showed that the hydroxylated peptides have characteristic
167 features that differ from either unmodified, or oxidised peptides. For example, we
168 observed clear differences in both charge states, (Figure 2B and Figure S1B), and mass
169 distributions, (Figure 2C and Figure S1C), of peptides containing hydroxylated proline
170 residues. Thus, the most frequent charge state for both unmodified and oxidised peptide
171 ions was 2+, with fewer peptide ions identified at higher charge states. In contrast, the
172 most frequent charge state for peptide ions containing hydroxylated proline was 3+,
173 with fewer peptide ions identified with a charge of either 2+, or 4+. The mass range for
174 most of the unmodified and oxidised peptides was 800-2,220 Da, while for
175 hydroxylated peptides it was 1,900-3100 Da.
176 A possible explanation for this observed variation in peptide mass value is that proline
177 hydroxylation increases the frequency of missed cleavages when digesting proteins to
178 peptides using trypsin, resulting in longer peptide sequences, with on average a
179 correspondingly higher charge state and higher mass. Consistent with this hypothesis,
180 a higher fraction, (~65%), of the hydroxylated peptides from HEK293 samples were
181 detected with at least one missed cleavage (Figure 2D), compared with the fraction of
182 either unmodified peptides (56%), or oxidised peptides (53%). The same phenomenon
183 was observed for RCC4 samples (Figure S1D), where 52% of the proline hydroxylated
184 peptides were detected with missed cleavages, which was higher than the level of
185 peptides for either unmodified peptides (48%), or oxidised peptides (38%).

186 In summary, HILIC fractionation is shown to provide a convenient approach for
187 enriching hydroxylated peptides and separating these from both unmodified and
188 oxidised forms of the same peptide sequences.

189 **Detection of HyPro diagnostic ion**

190 The unique immonium ion (m/z of 86.0659), which is generated from the fragmentation
191 of HyPro in MS, is known to distinguish HyPro from other amino acids in peptides, or
192 PTMs (Figure S2) and thus can be diagnostic for the presence of a hydroxylated proline.
193 However, we were intrigued to see that even for synthetic peptides that are known to
194 include a HyPro residue, this immonium ion was not always detectable in the MS2
195 fragmentation spectra (Figure 3). We therefore investigated whether the instrument
196 settings used for MS analysis might affect the generation and detection of the
197 immonium ion, following fragmentation of different peptide sequences. To test this, we
198 compared MS2 spectra obtained from the hydroxylated synthetic peptides
199 LDLEMLAP(564)YIPMDDD; (derived from protein HIF1α [2]),
200 WHLSSLAP(1717)PYVK; (derived from protein CEP192 [20]) and
201 LAPITSDP(408)TEATAVGAVEASFK (derived from protein PKM2 [13]).

202 The intensity of related fragment ions was compared with each of these peptides (Figure
203 3). For all three peptides, (with 1 pmol of each analysed), the intensity of the diagnostic
204 HyPro immonium ion, (relative percentage to base peak MS2 intensity), increased
205 following an increase in the collision energy (NCE) setting in the MS instrument used
206 to generate fragment ions from peptides, (Figure 3A, 3B and 3C), while a common
207 NCE setting for HCD is ~ 30 in an Orbitrap Tribrid MS platform.

208 Importantly, these results reveal that a higher than usual collision energy must be used
209 during MS analysis to increase the probability of detecting the diagnostic HyPro
210 immonium ion. We propose that this, along with other MS analysis parameters, (see
211 below), may have contributed to some of the recent controversy in the field surrounding
212 the ability to confirm the identification of PHD-dependent HyPro residues in non-HIF
213 family protein targets.

214 Using a higher energy NCE setting increased the probability of generating the
215 diagnostic HyPro immonium ion, while the low/medium energy settings generated
216 more useful b/y ions that can used for peptide matching. However, even with equal
217 loading and identical MS fragment settings, HyPro residues located within the context
218 of different surrounding amino acid sequences showed variable intensity of the
219 diagnostic HyPro immonium ion. Thus, for both the hydroxylated CEP192 peptide and

220 PKM2 peptide, we readily observe ~15-20% intensity of the HyPro immonium ion in
221 MS2 spectra, while under identical MS analysis conditions the HIF peptide only
222 generated ~3% of the immonium ion at its maximum. We conclude that different
223 peptides present different ionisation and transmission features in MS, depending on
224 their amino acid composition and sequence, which in turn affects the ability to detect
225 sites with HyPro modification.
226 Our studies have also identified other factors during MS analysis that can contribute to
227 the efficiency of detecting the diagnostic HyPro immonium ion, including the
228 concentration of the target peptide in the samples injected into the MS. If the target
229 peptide amount is relatively low, even with optimal NCE settings, the number of
230 diagnostic immonium ions transferred to the MS detector might fall below the detection
231 limit. To test this hypothesis, different amounts of prolyl hydroxylated synthetic
232 peptides were loaded and analysed by MS, using the target acquisition mode PRM: 10
233 fmol, 50 fmol, 100 fmol, 200 fmol, 400 fmol, 800 fmol and 1,000 fmol (Figure 3). As
234 expected, both the MS2 base peak intensity and the diagnostic ion intensity, dropped as
235 the loading amount on LC-MS decreased. Even with the more sensitive PRM mode, no
236 diagnostic HyPro immonium ion was detected from the HIF1α peptide at 10 fmol
237 (Figure 3A).
238 In summary, we have demonstrated that the intensity of the diagnostic HyPro
239 immonium ion is sensitive to multiple factors involved in MS analysis, including
240 collision energy settings in the MS instrument, the amino acid sequence surrounding
241 the HyPro site in peptides and the quantity of the target modified peptide in the sample
242 under analysis.
243 **Proline hydroxylation sites profiling of HEK293 and RCC4 cell lines**
244 To further improve our workflow and reduce the chance of misidentification of HyPro,
245 the MS data were filtered by the detection of the diagnostic HyPro immonium ion. For
246 other potential HyPro sites, where the presence of the diagnostic immonium ion could
247 not be confirmed, we recommend that a score cut-off and site localisation control
248 should be used for identification. Accordingly, a score cut-off of 40 for hydroxylated
249 peptides and a localisation probability cut-off of more than 0.5 for hydroxylated
250 peptides was performed.
251 HEK293 (4 bio-replicates) and RCC4 (3 bio-replicates) cell lines were plated 24 h
252 before treatment with either FG-4592 (50 μM, 24 h), or MG-132 (20 μM, 3 h before

253 harvesting), or DMSO control (24 h). Cells were harvested, lysed, and further prepared
254 for HILIC enrichment and LC-MS/MS analysis.
255 Implementation of the above parameters resulted in the identification of 4,768 and
256 3,063 proline hydroxylation sites without the associated diagnostic ions, and 4,993 and
257 3,247 sites in total, (together with the sites with diagnostic ions), from HEK293 and
258 RCC4 samples, respectively (Table S2). These are considered as high confidence
259 identifications.
260 Within the high confidence dataset, we successfully identified known proline
261 hydroxylation sites in both cell lines, including the multiple proline sites of collagen
262 proteins COL2A1 and COL4A1 and P62 in the ribosomal protein RPS2. Also,
263 previously reported potential PHD-dependent hydroxylated proline residues were
264 identified in either the HEK293, or RCC4 datasets, such as P294 of NDRG3 [45] , P319
265 of PPP2R2A [46] and P307/P322 of ACTB [47].
266 By following the diagnostic ion filter, 225 and 184 HyPro sites were confirmed from
267 all the HEK293 and RCC4 samples, respectively (Table S2). Interestingly, more proline
268 hydroxylation sites were found from collagen proteins in the RCC4 dataset (116 of 184),
269 as compared with the HEK293 dataset (53 of 225). The hydroxylated peptide sequence
270 in the collagen proteins shared the collagen-like sequence motif GxPGxx with Gxx
271 repeats (Figure 4A and 5A). These data are not surprising, since HyPro is known to be
272 a common PTM found in collagen proteins and critical for helix stability [48,49]. No
273 consistent sequence motifs were identified in the remaining 68 (HEK293) and 172
274 (RCC4) sites from non-collagen proteins (Figure 4A and 5A), which may indicate that
275 these proteins include multiple distinct target classes, rather than all belonging to the
276 same class of targets.
277 Furthermore, to show the representation of the sequence conservation of amino acids
278 in all the proline hydroxylated peptides sequence window, sequence logo analysis was
279 performed (Figure 4B and 5B). A high frequency of the amino acids L, P, G, S and E
280 in the hydroxylated peptides was immediately apparent in the HEK293 dataset. The
281 RCC4 dataset showed a similar pattern, but without L. This indicates that different cell
282 lines show unique profiles of proteins with hydroxylated peptides and is an argument
283 against these identifications arising via non-specific chemical oxidation during sample
284 preparation, since an identical sample preparation procedure was used for both cell
285 extracts.

286 We noted that there was a high frequency of a methionine residues appearing either at
287 the first, second, or even third positions after the HyPro site (Figure S3). Bearing in
288 mind concerns that HyPro identification by MS could be confused with oxidation of
289 methionine, (see above), since both involve a mass shift of +16 Da, as a conservative
290 approach these peptides were filtered out and a set of 3,521 and 2,728 filtered HyPro
291 sites (Table S3), was generated for sequence analysis for the HEK293 and RCC4
292 datasets, respectively.
293 Next, we matched the proline hydroxylated proteins against the Pfam database
294 (https://pfam-legacy.xfam.org/[50]), to investigate protein families and the potential
295 presence of any overrepresented sequence motif. After the removal of all the collagen
296 proteins, the most common terms identified in both samples included the RNA
297 recognition motif (RRM_1), WD40 repeat, conserved site (WD40) and protein kinase
298 domain (Pkinase). Besides, the DEAD/DEAH box helicase domain (DEAD) and
299 Helicase, C-terminal (Helicase_C) were also observed in the HEK293 dataset, while
300 the Calponin homology domain (CH) and LIM domain (LIM) were more enriched in
301 RCC4 dataset. As shown in Figures 4C and 5C, there was no evidence for the over-
302 representation of a single, simple motif in any of the protein families, but we noted that
303 the amino acid frequency differed for different protein families. In the HEK293 dataset,
304 the amino acids G, S, D and E were frequently located in proximity to the hydroxylated
305 P residue for the hydroxylated peptides from RRM_1, while more L and A residues
306 were in the hydroxylated peptides from WD40. For the Helicase_C domain, E was
307 noticed frequently in the hydroxylated peptides. In the RCC4 dataset, multiple P, G and
308 S amino acids were included in the sequence for the LIM domain, akin to collagen
309 proteins, where proline hydroxylated peptides typically have multiple P and G. Also, a
310 GTP motif was present in the proline hydroxylated peptides for the Pkinase domain. A
311 previous study[21] has also reported the proline embedded in the L/xGxP consensus
312 sequences of the kinase domains, (DYRK1A and DYRK1B), as the target of
313 hydroxylation by PHD1. This is a highly conserved motif, present in most of the kinases
314 comprising the CMGC family.

315 **Reactome pathway analysis of the FG-4592 dependent dataset**

316 Next, to identify cellular functions and pathways in which proteins with PHD-catalysed
317 proline hydroxylation may be involved, we compared HyPro sites identified in extracts
318 prepared from both HEK293 and RCC4 cell lines that had been treated with either (i)
319 the prolyl hydroxylase inhibitor FG-4592, or (ii) the proteasome inhibitor MG-132,

320 with DMSO treatment as the negative control. Reactome pathway analysis was
321 performed using the HyPro sites list, (Table S4), filtered by the following conditions:
322 (i) HyPro sites only identified in extracts from either DMSO, or MG-132 treated cells,
323 but not in extracts from cells treated with FG-4592; (ii) HyPro sites from all collagen
324 proteins were removed. All the subsequent pathway analyses were performed using
325 g:Profiler [51].
326 Interestingly, the data from HEK293 cell extracts revealed a statistical enrichment for
327 proline hydroxylated proteins belonging to two main categories: (i) metabolism of RNA,
328 i.e., processing of capped intron-containing Pre-mRNA and mRNA splicing; (ii) cell
329 cycle regulatory pathways, i.e., cell cycle (mitotic), S Phase, M Phase, Mitotic
330 Anaphase, Mitotic G1 phase and G1/S transition (Table S5 and Figure 6A). The
331 equivalent data from the RCC4 cell line also revealed proteins enriched in the same
332 metabolism of RNA and cell cycle (mitotic) pathways, (Table S5 and Figure 6B), but
333 additionally included enrichment of proteins involved in signalling by Rho GTPases,
334 Axon guidance and L1 family of cell adhesion molecules (L1CAMs) interactions [52]. By
335 comparing the HEK293 and RCC4 datasets, the most highly enriched pathways in
336 common to both cell lines are metabolism of RNA, cell cycle (mitotic) and cellular
337 response to heat stress/stress/stimuli.
338 In total, 1,412 identical proline hydroxylation sites were identified in both the HEK293
339 and RCC4 datasets, (Figure 6C, generated on http://barc.wi.mit.edu/tools/venn/). After
340 filtering to remove sites at which hydroxylation was not inhibited by FG-4592 treatment,
341 167 proline hydroxylation sites, (corresponding to 136 proteins), were identified in both
342 cell lines as strong candidates for being PHD targets (Table S4). By analysing
343 Reactome pathway for the proteins in which these sites were located, (Table S5 and
344 Figure 6D), metabolism of RNA and cell cycle regulatory pathways were enriched
345 again. This is consistent with, a previous study [20] that has also shown that the proline
346 hydroxylation of the centrosomal protein Cep192, mediated by PHD1, could be directly
347 linked to the control of cell cycle progression.
348 RCC4 cells have a mutant VHL gene, where HIF1α and HIF2α cannot be targeted for
349 proteasomal degradation, thus providing a model for analysis of VHL-independent
350 mechanisms related to cell cycle progression. To determine whether PHD enzymes
351 influence cell cycle progression, independent of VHL and HIF-α stabilisation, each of
352 the three PHD isoforms were independently depleted by siRNA and analysed by
353 fluorescence-activated cell sorting (FACS). Interestingly, following depletion of PHD1

354 in RCC4 cells, the G1-phase and S-phase populations decreased, while the G2/M
355 populations significantly increased (Figure 6E). In contrast, the knockdown of PHD2
356 and PHD3 resulted in no significant change in the respective levels of the different cell
357 cycle populations. Similarly, 24-hour treatment of RCC4 cells with the PHD inhibitor,
358 FG-4592, resulted in a significant reduction in the G1-phase population and an increase
359 in cells in the G2/M-phase, (Figure 6F), consistent with the PHD1 knockdown data.

360 **Validation of proline hydroxylation sites using synthetic peptides**

361 While the above data provide robust identification of HyPro sites in proteins, even using
362 HILIC fractionation enrichment and score/probability cut-off, it cannot be completely
363 excluded that some false proline hydroxylation identifications. Therefore, we have
364 included an extra validation step for identification of hydroxylated peptides using
365 synthetic peptides to characterise cognate MS spectra for HyPro-containing peptide
366 sequences. This focussed on the following criteria:

367 1. Comparison of the LC retention time of the experimentally measured peptides
368 and corresponding synthetic peptides.
369 2. Investigation of MS1 and MS2 ion features, including manual analysis of the
370 spectra to confirm the results obtained from software, determining the fragment
371 b/y ions with high quality, and noting the ions neighbouring the hydroxylation
372 sites.
373 3. Identification of the diagnostic HyPro immonium ion.

374 This approach is illustrated for the peptide sequence:
375 KPLLSPIPELPEVP(600)EM(602)TP(604)SIP(607)SIRR, derived from the protein
376 Repo-Man (CDCA2), which was first discovered as a chromatin-associated regulator
377 of protein phosphatase one (PP1) [53] and identified here as a potential PHD target from
378 the HEK293 dataset (Table S1 and Table S4). In this peptide, multiple prolines
379 (P600/P604/P607) could potentially be hydroxylated, but none had a localisation
380 probability high enough (>90%) to confirm unambiguously the actual HyPro site.
381 Manual inspection of the MS/MS spectra indicated that amino acid residues M602,
382 P604 or P607 could be either oxidised, or hydroxylated, each causing a +16 Da mass
383 shift. Some of the fragment ions in the MS/MS spectra were missing that would have
384 allowed us to distinguish between hydroxylation of either P604, or P607 residues, or
385 oxidation at M602. This can be a common situation when analysing peptides isolated
386 from cell or tissue extracts, especially when investigating low-abundance
387 peptides/proteins.

388 To distinguish between these possibilities, we first investigated how either the
389 hydroxylation of P, or the oxidation of M, would alter peptide chromatography in
390 RPLC-MS analysis. This was achieved by comparing a set of synthetic peptides that
391 were either hydroxylated, oxidised, or unmodified and determining their respective
392 retention times (Figure 7A). This revealed that the peptides hydroxylated at either P604,
393 or P607, had retention times that peaked at 76.54 minutes and 77.37 minutes,
394 respectively. By contrast, the retention time of the unmodified peptide was 80.26
395 minutes, while that of the methionine oxidised peptide was 74.30 minutes. We conclude
396 that the retention time measured by RPLC thus provides a discerning feature for HyPro
397 modified peptides that can be used to validate protein hydroxylation sites and reduce
398 the chance of misidentification of methionine oxidation as a potential HyPro site.
399 To validate the MS1 and MS2 ion features of the prolyl-hydroxylated peptide, we next
400 performed an IP-MS experiment in HeLa cells transiently transfected with a Repo-Man
401 expression construct. For this analysis, the cells were arrested in Prometaphase by
402 treatment with nocodazole. Cell extracts were prepared and the corresponding synthetic
403 peptides, with heavy isotope labelled lysine, (K, +8Da), were spiked into each sample
404 before LC-MS/MS analysis, to facilitate identification and quantification of the proline
405 hydroxylated Repo-Man peptide. By measuring the RPLC retention time of synthetic
406 peptides, the oxidation at M602 (75.55 min) and hydroxylation at either P604 (77.89
407 min), or P607 (79.29 min), could be distinguished, despite each having a similar mass
408 shift of +16 Da (Figure 7B).
409 We note that the Repo-Man peptides hydroxylated at either P604, or P607, could be
410 further modified by oxidation also at M602, causing a mass shift +32 Da (Figure 7C).
411 By manually checking the experimental LC profile and MS/MS spectra (Figures 6C
412 and 7D, generated via Universal Spectrum Explorer[54]), it was clear that a Repo-Man
413 peptide could be detected with a mass shift of 32 Da, indicating that both hydroxylation
414 at P604 and oxidation at M602 could occur at the same time. The site of modification
415 was confirmed from the MS/MS spectra via comparison with the spectra from the
416 synthetic peptides, based upon similarity in the respective MS2 fragmentation patterns
417 (Figure 7D).
418 The diagnostic immonium ion was not detected from either of the proline hydroxylation
419 site in these MS2 spectra, neither for the endogenous Repo-Man peptides from HeLa
420 cells, nor for synthetic hydroxylated Repo-Man peptides. As described above, the
421 detection of the HyPro immonium ion in MS spectra depends on the collision energy

422 and the quantity of target peptides. Using the typical NCE setting of 27% on a Q
423 Exactive MS platform, the diagnostic immonium ion of the synthetic Repo-Man proline
424 hydroxylated peptide was barely detectable (Figure 7E). However, as the NCE setting
425 was increased, there was a corresponding increase in the intensity of the HyPro
426 diagnostic immonium ion detected. In contrast with the HyPro immonium ion, varying
427 energy settings showed little to no effect on the efficiency of detection of the immonium
428 ions from L/I (Figure 7F).
429 Finally, by comparing the RPLC retention times of the respective spiked-in heavy
430 isotope-labelled synthetic peptides with hydroxylation at either P604 (72.54 min, with
431 M602 oxidation), or hydroxylation at P607 (73.57 min, with M602 oxidation), these
432 could be clearly distinguished. Therefore, we conclude that the Repo-Man protein in
433 HeLa cell extracts is hydroxylated specifically at P604.
434 **Hydroxylation of Repo-Man at P604 is PHD-dependent**
435 Next, we investigated further whether the hydroxylation of P604 on Repo-Man is PHD-
436 dependent in cells. To this end we performed SILAC (Stable Isotope Labelling of amino
437 acids in cell culture)-based quantitative proteomics in both HeLa and HEK293 cells,
438 comparing cells grown +/- the PHD inhibitor FG-4592. Asynchronous cells were grown
439 either in 'heavy' K4+R6 media, i.e., containing a combination of heavy isotope labelled
440 lysine (K) and Arginine (R), or in 'light' K0+R0 (normal) media. Although HIF1α
441 stabilisation in response to FG-4592 is efficient (Figure S4A), FG-4592 was added to
442 cells growing in heavy media for 24h to ensure that the PHD activity was inhibited
443 throughout the cell cycle, while DMSO was added in the light media as a control
444 (Figure 8A). Total cell extracts were prepared, subjected to HILIC enrichment and the
445 resulting samples were analysed by LC-MS/MS.
446 The spectrum of the Repo-Man tryptic peptide (KPLLSPIPELPEVPEMTPSIPSIRR)
447 containing hydroxylated P604, oxidised M602 and the diagnostic ion is shown in Figure
448 8B. The abundance of the hydroxylated (OH-P604) Repo-Man peptide in each
449 condition i.e., DMSO and FG-4592, was quantified with its intensity shown in Figure
450 8C. The Repo-Man peptide hydroxylated at P604 was only detected in extracts of cells
451 grown in the DMSO control media and not in extracts from cells grown in the presence
452 of the PHD inhibitor FG-4592 (Figure 8C).
453 As a control for the FG-4592 treatment, we quantified levels of HIF1α, which, as
454 expected, was only detected in the FG-4592 treated samples (Figure 8D). Furthermore,
455 the FG-4592-dependent differences observed in Repo-Man hydroxylation in HeLa and

456 HEK293 cells were not due to changes in total Repo-Man protein levels, as the ratio
457 H/L is ~ 1 (Figure 8E), confirming no significant protein level alterations in cells treated
458 with FG-4592. As an additional control for the possible effect of SILAC media on
459 protein expression, we measured Histone H4 levels, which demonstrated no significant
460 changes in response to FG-4592 treatment (Figure 8E). In summary, these data show
461 that Repo-Man hydroxylation at P604 in cells is PHD-dependent.

462 **Recombinant PHD1 hydroxylates Repo-Man at P604 *in vitro***

463 To determine whether Repo-Man could be a substrate for PHD hydroxylation *in vitro*,
464 we tested if recombinant PHD1 would hydroxylate a synthetic Repo-Man peptide at the
465 P604 site that was hydroxylated in cell extracts (Figure 9, A-B). To avoid any
466 misinterpretation between oxidation and hydroxylation, we used a synthetic Repo-Man
467 peptide as a substrate that was already oxidised at M602
468 (KPLLSPIPELPEVPEM(OX)TPSIPSIRR). Following incubation *in vitro* with
469 recombinant PHD1, samples were then analysed by LC-MS/MS. Figure 9A compares
470 the MS2 spectra obtained from incubating the Repo-Man peptide with recombinant
471 PHD1, together with a reference synthetic peptide that is hydroxylated on P604. By
472 comparing the fragment ions from MS2 spectra obtained from incubating the Repo-
473 Man peptide +/- recombinant PHD1, especially the Y5 and Y8 fragment ions derived
474 from the parent ion, (KPLLSPIPELPEVPEM(OX)TPSIPSIRR +16Da), we confirm
475 that the PHD1 hydroxylation site is P604 and not P607 (Figure 9, A-B).

476 As a control, we performed in parallel the *in vitro* hydroxylation using the HIF1α
477 peptide, (QDTDLEMLAPYIPMDDDFQLR) (Figure 9, C-D). Taking advantage of the
478 availability of an antibody that recognises OH-P564 in HIF1α, a dot blot was performed
479 with part of the *in vitro* reaction. As a control for the OH-564-HIF1α antibody, varying
480 concentrations of the HIF1α synthetic peptide were tested on the dot blot (Figure 9C).
481 The rest of the sample was analysed by MS. Figure 9D shows the MS spectra of HIF1α,
482 with a +16Da shift only seen when the peptide was incubated with PHD1 and with the
483 resulting hydroxylated peptide having a similar retention time to that observed for a
484 positive control OH-564-HIF1α synthetic peptide. We note that the efficiency of the *in*
485 *vitro* hydroxylation reaction, using only recombinant PHD1, was significantly higher
486 with the HIF1α peptide than with the Repo-Man peptide. It is possible that the
487 hydroxylation efficiency for Repo-Man *in cellulo* is increased by other factors, and/or
488 by the sequence context of the endogenous protein, which are not reflected in the
489 simplified *in vitro* assay.

490 In summary, the data demonstrate that a Repo-Man peptide can be *in vitro* hydroxylated
491 at P604 by recombinant PHD1, consistent with data above showing that Repo-Man is
492 a target for PHD hydroxylation in Hela and HEK293 cells.
493

494 **Discussion**

495 In this study, we have successfully established a robust and reproducible workflow for
496 the reliable identification of proline hydroxylation sites in proteins, using mass
497 spectrometry. This workflow includes a sample preparation step to enrich proline
498 hydroxylated peptides, prior to LC-MS/MS analysis. Enrichment is achieved using
499 HILIC fractionation, which exploits the change in hydrophilicity of peptides resulting
500 from addition of a hydroxyl group on proline.

501 In addition to hydroxylated peptide enrichment using HILIC fractionation, we have
502 identified several parameters involved in the MS protocol and subsequent data analysis
503 that affect the ability to detect specific proline hydroxylation sites and distinguish these
504 from alternative forms of amino acid oxidation. For example, tryptic peptides modified
505 with HyPro were found to show differences in peptide mass and charge state, as
506 compared with either unmodified peptides, or peptides with oxidised methionine.

507 We note that it was previously reported to be difficult to distinguish between peptides
508 with either HyPro ($C_6H_{13}NO_2$), or Leu/Ile ($C_5H_9NO_3$), because their corresponding
509 residue mass difference is only 0.03639 Da (Figure S2). However, using modern MS
510 instruments for proline hydroxylation identification, this concern is mitigated, as
511 modern instruments are now capable of high mass resolution at both MS1 and MS2
512 levels. For example, using the Orbitrap Fusion MS, the MS1 and MS2 resolution values
513 were 70,000 and 15,000, respectively. This corresponds to 10 ppm and 20 ppm as the
514 mass tolerance parameter in data searching software and more recent MS instruments
515 could provide even higher resolution under similar instrument scanning rate. This
516 provides sufficient mass resolution to distinguish whether a peptide contains either
517 HyPro, or Leu/Ile.

518 Of particular interest, we have found that the generation of the diagnostic HyPro
519 immonium ion in MS2 spectra is very sensitive to the collision energy setting used for
520 peptide fragmentation and is also sensitive to the amino acid sequence of the peptide
521 and its concentration in the analysis sample. In contrast, other immonium ions, e.g., for
522 Leu/Ile, do not show this sensitivity to collision energy and are detected with similar
523 efficiency over a broad range of energy settings. We propose that these observations
524 can explain, at least in part, why the diagnostic HyPro immonium ion was only detected
525 for 4.5% of the HyPro sites we mapped here from HEK293 cell extracts and 5.6% for
526 RCC4 in the hydroxylation profiling experiment. This discovery, regarding the
527 sensitivity of MS analysis parameters for generation and detection of diagnostic HyPro

528 immonium ions, may be a significant contributory factor in some of the recent
529 controversies surrounding the identification of proline hydroxylated PHD target
530 proteins. As shown in this study, the use of synthetic peptides, combined with optimised
531 analysis procedures, provides a reliable strategy for identifying cellular proteins
532 modified by proline hydroxylation. Thus, we recommend that for HyPro site mapping
533 and for the purpose of validating specific target sites, a workflow is used that combines
534 higher NCE energy settings in MS, (combined with optimisation using synthetic
535 peptides), using a highly sensitive MS scan mode, (such as SIM, PRM for target
536 analysis), to increase the detection limit and using HILIC enrichment to increase the
537 concentration of HyPro modified peptides in samples being analysed.
538 Having validated the analysis strategy for detecting peptides modified with proline
539 hydroxylation, we proceeded to characterise HyPro sites in extracts of both the HEK293
540 and RCC4 human cell lines. Several data analysis filters were applied to compare the
541 sites, e.g., removing peptides mapped to collagen proteins, removing peptides
542 containing methionine residues close to the HyPro sites and removing peptides where
543 the proline hydroxylation sites were still observed after treating cells with PHD
544 inhibitors. Using these filtered peptides, sequence motif analysis showed that proline
545 hydroxylation sites were enriched in the RNA recognition motif (RRM_1), protein
546 kinase domain (Pkinase) and WD40 repeat, conserved site (WD40).
547 Given the potential wider biological importance of PHD-mediated regulation of key
548 cellular processes, it is worth considering how the identification of proline hydroxylated
549 target proteins can be further improved in future. For example, although we show here
550 that proline hydroxylated peptides can be usefully enriched by HILIC fractionation,
551 prior to LC-MS/MS analysis, we note that the enrichment efficiency is not as high as
552 that possible with some methods for enrichment of peptides carrying other forms of
553 PTM, particularly in the case of enrichment for phosphorylated peptides using TiO_2 [55]
554 or IMAC chromatography [56]. HILIC fractionation will also enrich peptides with
555 increased hydrophilicity due to modifications such as phosphorylation and
556 glycosylation [57], which might complicate the detection of hydroxylated peptides. It will
557 be interesting, therefore, to test whether the enzymatic removal of other modifications
558 during sample preparation, such as using PNGase F to remove the N-glycans from
559 proteins [58], might further affect the efficiency of proline hydroxylation identification
560 using HILIC. Another issue worth noting is the high abundance of proline hydroxylated
561 peptides derived from collagen proteins. Since MS is a concentration-sensitive analysis

562 tool, high abundance ions from the collagen proteins may inhibit the ionisation of others,
563 lower abundant proline hydroxylated ions derived from other, medium/low abundant
564 proteins, thereby reducing the chance of their identification. To mitigate this problem,
565 it may be helpful to include extra steps during sample preparation to selectively remove
566 collagen proteins/peptides, prior to MS analysis. Furthermore, as previously mentioned,
567 proline hydroxylated proteins tend to show increased levels of missed cleavages during
568 trypsin digestion. A strategy of using multiple proteases, other than trypsin, has
569 previously been used to increase the identification depth in proteomics studies [59,60] and
570 this may also be interesting to test in future for identification of proline hydroxylation
571 sites.
572 We show here that a Repo-Man peptide containing P604 and the surrounding sequence
573 is recognised and hydroxylated *in vitro* by purified, recombinant PHD1, but note that
574 the efficiency of this *in vitro* hydroxylation is lower than that seen for hydroxylation of
575 a HIF1α peptide. We suggest that *in cellulo* the activity of PHD enzymes towards many
576 target proteins may depend upon additional targeting subunits, analogous to the role of
577 targeting subunits for protein phosphatase PP1 and PP2 catalytic subunits [61-64], such as
578 Repo-Man itself and B56 [65,66]. The efficiency of hydroxylation may also be increased
579 by other features not replicated in the basic *in vitro* system, e.g. post-translational
580 modifications of residues on the target sequence surrounding the hydroxylation site. In
581 this regard it is interesting that even for the canonical PHD target HIF1α, the LIMD1
582 protein was identified as a bridging factor between PHDs and HIF1α [67]. It will be
583 interesting in future to investigate this further with experiments to evaluate factors that
584 can stimulate the efficiency of site-specific hydroxylation *in vitro* on Repo-Man and
585 other PHD targets.
586 Reactome pathway analysis showed that the proteins identified as targets of proline
587 hydroxylation in both the HEK293 and RCC4 cell lines were enriched preferentially
588 for factors involved in cell cycle regulation and RNA processing, suggesting possible
589 roles for PHD-mediated regulation in both these processes. This is consistent with our
590 previous studies showing that the regulation of cell cycle progression can be controlled
591 by mechanisms involving PHD-mediated proline hydroxylation, including HyPro
592 modification of the centrosomal protein CEP192 [20]. With regard to cell cycle regulation,
593 it is interesting that we also detected in this study hydroxylation of P604 in the mitotic
594 regulator Repo-Man, which is a chromatin-associated subunit of a protein phosphatase
595 one complex. We show further that hydroxylation of Repo-Man at P604 is PHD-

596 dependent in cells and that this site is a substrate for hydroxylation by recombinant
597 PHD1 *in vitro*.
598 In the accompanying study by Druker et al., we provide detailed experimental evidence
599 showing that hydroxylation of Repo-Man at P604 is of functional importance for
600 controlling progression through mitosis, acting, at least in part, via hydroxylation of
601 Repo-Man at P604 regulating the interaction of Repo-Man with the B56-PP2A
602 phosphatase complex during chromosome alignment and thereby controlling the levels
603 of Histone H3T3 phosphorylation. Based upon the many novel proline hydroxylation
604 sites we have identified, we propose that future experimental studies may show further
605 important roles for PHD-mediated cellular regulation linked with stress responses and
606 oxygen sensing, particularly relating to the control of RNA processing and cell cycle
607 progression.
608

609 **Materials and Methods**

610 **Cell culture**

611 HEK293, RCC4 and HeLa cells were cultured in Dulbecco's Modified Eagle Medium
612 (Gibco, # 41966-029), supplemented with 10% Fetal Bovine Serum (FBS, Gibco #
613 A3169801), 100 U/ml penicillin and streptomycin and 2 mM L-Glutamine. Cell lines
614 were maintained at 37°C with 5% CO2 in humidified incubator.
615 HEK293 or RCC4 cells were plated in 10-cm plates and cultured for 24 h. DMSO or
616 FG-4592 (Selleck # S1007, 50 µM) was then added into the media for the following 24
617 h incubation. MG-132 (Calbiochem, 20 µM) was added for 3 h.
618 For SILAC [68], HEK293 and HeLa cells were cultured in DMEM for SILAC media
619 (Thermo scientific # 88364), supplemented with 10% Dialyzed FBS (Gibco #A33820),
620 100 U/ml penicillin, streptomycin, 2 mM L-Glutamine and either '"Light" amino acids
621 (K0R0, Lysine and Arginine), or "Heavy" amino acids (K4R6, $^{13}C_4$-Lysine and $^{13}C_6$-
622 Arginine). The media was changed every 1–2 days. Cells were kept in SILAC media
623 for 6 cell passages to achieve >95% labelling. Before the last passage, cells were plated
624 in 10-cm plates and 24 h later. Either DMSO or 50 µM FG-4592 (Selleck # S1007),
625 were added for the last 24 h in light, or heavy media, respectively. After incubation,
626 cells were washed with PBS and trypsinised. Then cells from each condition were
627 resuspended in PBS. Cells from both conditions were mixed at 1:1 ratio (cell number)
628 and centrifuged at 1,000 g to collect the pellets.

629 **Lysates of HEK293, RCC4 and HeLa cells**

630 HEK293, RCC4 and HeLa cells were harvested in PBS and the pellets were lysed in
631 500 µL of lysis buffer (50 mM Tris pH 7.5, 150 mM NaCl, 1% NP40, 0.5% Sodium
632 deoxycholate and 0.1% SDS) containing protease inhibitors (Roche, Complete Mini
633 EDTA-Free) and phosphatase inhibitors (PhosSTOP, Roche). The lysates were
634 incubated on ice for 20 min and cleared by centrifugation at 4°C for 15 min at 13,000g.
635 The supernatant was transferred to a new tube for further analysis.

636 **Immunoprecipitation of Repo-Man in HeLa cells**

637 Prometaphase arrested HeLa cells were resuspended in 200 µL of lysis buffer (50 mM
638 Tris pH 7.5, 150 mM NaCl, 1% NP40, 0.5% Sodium deoxycholate and 0.1% SDS)
639 containing protease inhibitors (Roche, Complete Mini EDTA-Free) and phosphatase
640 inhibitors (PhosSTOP, Roche). The lysates were incubated on ice for 20 min and
641 cleared by centrifugation at 4°C for 15 min at 13,000 g. The supernatant was then
642 diluted by adding 300 µL of dilution buffer (50 mM Tris pH 7.5, 150 mM NaCl). A
643 total of 500 µL diluted lysates was transferred to a new tube containing pre-washed
644 GFP-Trap magnetic agarose beads (ChromoTek gtma) and incubated at 4°C for 2 h.
645 After incubation, the beads were washed twice with washing buffer (50 mM Tris pH
646 7.5, 150 mM NaCl and 0.05% NP40) and once with PBS, then transferred to a clean
647 tube in the last wash. The beads were resuspended in LDS sample buffer (Invitrogen).
648 After heating the beads at 70°C for 10 min, the eluted proteins were transferred to a
649 new tube for MS analysis.

650 **Sample preparation for proteomics analysis**

651 All protein samples were sonicated for 10 cycles (30s on/off) using the Bioruptor® Pico,
652 and then centrifugated at 20,000g for 15 min. After adding TCEP (10 mM final
653 concentration, Sigma-Alrich), the supernatant proteins were denatured and reduced at
654 95°C for 10 min, and then alkylated with 40 mM IAA (final concentration, Sigma-
655 Alrich) in dark at room temperature for 30 min. The protein concentration was
656 measured using EZQ™ Protein Quantitation Kit (Thermo Fisher Scientific) by
657 following the manual. The proteins from each sample were further processed using SP3
658 protocol as described [69]. In brief, protein samples were mixed with SP3 beads (1: 10,
659 protein: beads). For the HILIC fractionation samples from cell lysates, the starting
660 material of proteins was 500 ug. For IP-MS samples, 100 ug of beads were added into
661 the sample regardless of the protein amount. The proteins samples were then incubated
662 and cleaned with SP3 beads, then digested with lysC/trypsin mixture (1: 50, enzyme:

663 protein, Promega) at 37°C overnight. For IP-MS samples, 50 µL of 2% DMSO (in H_2O)
664 was used to elute the peptides. The peptide samples were directly used for LC-MS
665 analysis. For the HILIC fractionation samples, after the overnight digestion using
666 lysC/trypsin mixture, PNGase F (1000 U for 1 mg proteins/peptides) was added into
667 the samples for another incubation at 37°C for 4 h. 200 µL of 0.1% TFA (in H_2O) was
668 used to elute the peptides from SP3 beads. The peptide concentration was measured
669 using Pierce™ Quantitative Fluorometric Peptide Assay (Thermo Fisher Scientific) by
670 following the manual.

671 **HILIC fractionation**

672 The 200 µL of peptide samples were added with 200 µL of 0.1% Trifluoroacetic acid
673 (TFA) in Acetonitrile (ACN) and mixed well to have a clear solution. 3 more times of
674 0.1% TFA (in ACN) were added to reach a final concentration of 80% ACN with 0.1%
675 TFA. It's crucial to have a clear solution before HILIC fractionation. Centrifugation
676 might be needed to remove possible insoluble content. HILIC was performed using a
677 Dionex RSLCnano HPLC (Thermo Fisher Scientific). Peptides were injected onto a 4.6
678 mm ID × 15 cm TSKgel Amide-80 column (3 um pore size), using a 90-min multistep
679 gradient with a constant flow rate of 0.4 mL/min: 0-20 min, 80%B; 20-30 min, 80%-
680 70%B, 30-60 min, 70%-60%B, 60-65 min, 60%-0%B, 65-70 min, 0%B (flow rate at
681 0.2 mL/min). The mobile phases were 0.1% TFA in H_2O for solvent A and 0.1% TFA
682 in ACN for solvent B. Fractionations were collected by every 126 s, starting from 3
683 min to 70 min. All the peptide fractions were then dried using SpeedVac concentrator
684 (45°C or lower temperature, Thermo Fisher Scientific). The samples could be stored in
685 -20°C or resuspended in 0.1% FA and directly used for LC-MS analysis.

686 **LC-MS/MS analysis**

687 All the HILIC fractionation and IP samples were analysed by using a Q Exactive Plus
688 mass spectrometer platform (Thermo Fisher Scientific), equipped with a Dionex Ultra-
689 high-pressure liquid-chromatography system (RSLCnano). RPLC was performed using
690 a Dionex RSLCnano HPLC (Thermo Fisher Scientific). Peptides were injected onto a
691 75 µm × 2 cm PepMap-C18 pre-column and resolved on a 75 µm × 50 cm RP- C18
692 EASY-Spray temperature-controlled integrated column-emitter (Thermo Fisher
693 Scientific), using a 2-hour multistep gradient with a constant flow rate of 300 nl/min.
694 The mobile phases were H_2O incorporating 0.1% Formic acid FA (solvent A) and 80%
695 ACN incorporating 0.1% FA (solvent B). The gradient for HILIC enriched samples was:
696 0-6 min, 1%B; 6-10 min, 1%-8%B; 10-80 min, 8%-28%B; 80-95 min, 28%-38%B. The

697 gradient for IP-MS samples was: 0-6 min, 5%B; 6-91 min, 5%-38%B. The MS data
698 were acquired under the control of Xcalibur software in a data-dependent acquisition
699 mode (DDA) using top N mode. The survey scan was acquired in the orbitrap covering
700 the m/z range from 375 to 1,600 with a mass resolution of 70,000 and an automatic gain
701 control (AGC) target of $3e^6$ ions with 20 ms maximum injection time. For HILIC
702 peptide fractions, the top 20 most intense ions were selected for fragmentation using
703 HCD with 27% NCE collision energy and an isolation window of 1.6 Da with a
704 dynamic exclusion of 60 s. The MS2 scan was acquired in the orbitrap with a mass
705 resolution of 17,500 with the m/z range starts from either 80 or 100. The AGC target
706 was set to 1e5 with a maximum injection time of 60 ms. For IP-MS samples, the target
707 PRM mode was applied for fragmentation, with a AGC target of $2e^5$ and a maximum
708 injection time of 120 ms.
709 All the synthetic peptides were analysed using an Orbitrap Fusion Tribrid mass
710 spectrometer platform and RSLCnano system. The gradient was: 0-6 min, 5%B; 6-31
711 min, 5%-38%B; 31-35 min, 38-95%B; 35-40 min, 95%. The MS data were acquired
712 under the control of Xcalibur software in a target PRM. The survey scan was acquired
713 in the orbitrap covering the m/z range from 375 to 1,575 with a mass resolution of
714 60,000 and an automatic gain control (AGC) target of $4e^5$ ions with 50 ms maximum
715 injection time. The target ions were selected for fragmentation using HCD with 30%
716 NCE collision energy and an isolation window of 1.6 Da with a dynamic exclusion of
717 60 s. The MS2 scan was acquired in the orbitrap with a mass resolution of 30,000 with
718 the m/z range starts from 80. The AGC target was set to 5e4 with a maximum injection
719 time of 54 ms. The PRM inclusion list was shown in Table S6.
720 **MS Data analysis**
721 The MS data were analysed together using MaxQuant [70,71] (v. 2.3.1.0). The FDR
722 threshold was set to 1% for each of the respective Peptide Spectrum Match (PSM) and
723 Protein levels. The data was searched with the following parameters: quantification
724 type was set to LFQ, stable modification of carbamidomethyl (C), variable
725 modifications, oxidation (M), acetylation (protein N terminus), and hydroxylation (P),
726 with maximum of 2 missed tryptic cleavages threshold. First, proteins and peptides
727 were searched against the homo sapiens database from UniProt (SwissProt April 2023)
728 without adding hydroxylation (P) as a variable modification. Then a fasta file was
729 generated based on the identified protein list and used to identify the hydroxylation (P)
730 peptides and sites. The result tables generated from MaxQuant were processed and

filtered using Perseus [72] (v. 1.6.15.0). For SILAC data analysis, multiplicity for quantification was set to 2 with maximum 3 labelled AAs. “Arg6” and “Lys4” were selected in the heavy label channel.

In vitro hydroxylation assay and dot blot

For *in vitro* hydroxylation, either 3 µM of Repo-Man synthetic peptide (M602 oxidised) (KPLLSPIPELPEVPEM(OX)TPSIPSIRR), or 3 µM of HIF1α peptide (LDLEMLAPYIPMDDD), was incubated +/- 300 nM of recombinant PHD1/EGLN2 (Active motif #81064). The hydroxylation reaction was performed in a final volume of 30 µL of 20 mM Tris-HCl pH 7.5, 5 mM KCl, 1.5 mM $MgCl_2$, 1 mM DTT, 100 µM α-Ketoglutarate (Sigma #75890), 100 µM L-ascorbic acid (Sigma #A7506) and 50 µM Ammonium iron II sulphate hexahydrate (Fluka #09719), for 2.5 h at 30°C. After incubation, reactions were stopped by addition of 1 mM EDTA and samples subjected to LC-MS/MS analysis. In the case of HIF1α, 5 µL of the reaction was dot blotted onto a nitrocellulose membrane. Serial dilutions of P564 hydroxylated HIF1α peptide, (LDLEMLAP(Hydro)YIPMDDD), were dot blotted in the same membrane as a positive control. Membrane was air dried and blocked with Intercept Blocking buffer TBS (#92760001 0LiCOR) and incubated overnight with anti-HIF1α OH-P564 antibody (1/1000 dilution). After three washes with TBS_T of 5 min each with the last wash being done with TBS, the membrane was incubated with the secondary antibody donkey anti-rabbit IRDye 680RD (#926-68073 and imaged using a Typhoon Fuji Imager.

siRNA transfections of PHD isoforms

RCC4 cells were transfected with 20 µM siRNA of each PHD isoform using Interferin (Peqlab). A random scrambled sequence was used as the control. The siRNA sequences used were previously presented in [73].

Flow cytometric analysis of cell cycle distributions

Adherent RCC4 cells were harvested, pooled, and washed once in PBS prior to being fixed in ice-cold 70% (v/v) ethanol/distilled water. Afterward, the cells were twice washed in PBS and resuspended in Guava Cell Cycle Reagent (Luminex Corp.; #4500-0220) for 30 minutes at room temperature. The cell cycle distribution with a Guava easyCyte HT machine and software. Cells with DNA content between 2N and 4N were allocated into phases G1-, S-, or G2/M of the cell cycle. The cell profiles were gated according to the control in each independent experiment and the number of cells in each phase are expressed as a percentage of the total number of cells counted.

Acknowledgements

We thank all our collaborators who worked on the project as well as Alejandro Brenes and Michael Batie for all the insightful discussions. We would also like to thank all members of the Lamond Laboratory. This work was funded by the Wellcome Trust grant (206293/Z/17/Z) with additional support provided by grants from BBSRC (BB/V010948/1; APP3732) & UKRI (EP/Y010655/1).

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE [74] partner repository with the dataset identifier PXD044783 (RCC4) and PXD044663 (HEK293).

References

1. Epstein, A.C., Gleadle, J.M., McNeill, L.A., Hewitson, K.S., O'Rourke, J., Mole, D.R., Mukherji, M., Metzen, E., Wilson, M.I., Dhanda, A., et al. (2001). C. elegans EGL-9 and mammalian homologs define a family of dioxygenases that regulate HIF by prolyl hydroxylation. Cell *107*, 43-54. 10.1016/s0092-8674(01)00507-4.
2. Jaakkola, P., Mole, D.R., Tian, Y.M., Wilson, M.I., Gielbert, J., Gaskell, S.J., von Kriegsheim, A., Hebestreit, H.F., Mukherji, M., Schofield, C.J., et al. (2001). Targeting of HIF-alpha to the von Hippel-Lindau ubiquitylation complex by O2-regulated prolyl hydroxylation. Science *292*, 468-472. 10.1126/science.1059796.
3. Kaelin, W.G., Jr., and Ratcliffe, P.J. (2008). Oxygen sensing by metazoans: the central role of the HIF hydroxylase pathway. Mol Cell *30*, 393-402. 10.1016/j.molcel.2008.04.009.
4. Rocha, S. (2007). Gene regulation under low oxygen: holding your breath for transcription. Trends Biochem Sci *32*, 389-397. 10.1016/j.tibs.2007.06.005.
5. Duan, C. (2016). Hypoxia-inducible factor 3 biology: complexities and emerging themes. Am J Physiol Cell Physiol *310*, C260-269. 10.1152/ajpcell.00315.2015.
6. D'Ignazio, L., Batie, M., and Rocha, S. (2017). Hypoxia and Inflammation in Cancer, Focus on HIF and NF-kappaB. Biomedicines *5*. 10.3390/biomedicines5020021.
7. Biddlestone, J., Bandarra, D., and Rocha, S. (2015). The role of hypoxia in inflammatory disease (review). Int J Mol Med *35*, 859-869. 10.3892/ijmm.2015.2079.
8. Maxwell, P.H., Wiesener, M.S., Chang, G.W., Clifford, S.C., Vaux, E.C., Cockman, M.E., Wykoff, C.C., Pugh, C.W., Maher, E.R., and Ratcliffe, P.J. (1999). The tumour suppressor protein VHL targets hypoxia-inducible factors for oxygen-dependent proteolysis. Nature *399*, 271-275. 10.1038/20459.
9. Ivan, M., Kondo, K., Yang, H., Kim, W., Valiando, J., Ohh, M., Salic, A., Asara, J.M., Lane, W.S., and Kaelin, W.G., Jr. (2001). HIFalpha targeted for VHL-mediated destruction by proline hydroxylation: implications for O2 sensing. Science *292*, 464-468. 10.1126/science.1059817.
10. Ohh, M., Park, C.W., Ivan, M., Hoffman, M.A., Kim, T.Y., Huang, L.E., Pavletich, N., Chau, V., and Kaelin, W.G. (2000). Ubiquitination of hypoxia-inducible factor requires direct binding to the beta-domain of the von Hippel-Lindau protein. Nat Cell Biol *2*, 423-427. 10.1038/35017054.
11. Cummins, E.P., Berra, E., Comerford, K.M., Ginouves, A., Fitzgerald, K.T., Seeballuck, F., Godson, C., Nielsen, J.E., Moynagh, P., Pouyssegur, J., and Taylor, C.T. (2006). Prolyl hydroxylase-1 negatively regulates IkappaB kinase-beta, giving insight into hypoxia-induced NFkappaB activity. Proc Natl Acad Sci U S A *103*, 18154-18159. 10.1073/pnas.0602235103.
12. Zheng, X., Zhai, B., Koivunen, P., Shin, S.J., Lu, G., Liu, J., Geisen, C., Chakraborty, A.A., Moslehi, J.J., Smalley, D.M., et al. (2014). Prolyl hydroxylation by EglN2 destabilizes FOXO3a

by blocking its interaction with the USP9x deubiquitinase. Genes Dev *28*, 1429-1444. 10.1101/gad.242131.114.

13. Luo, W., Hu, H., Chang, R., Zhong, J., Knabel, M., O'Meally, R., Cole, R.N., Pandey, A., and Semenza, G.L. (2011). Pyruvate kinase M2 is a PHD3-stimulated coactivator for hypoxia-inducible factor 1. Cell *145*, 732-744. 10.1016/j.cell.2011.03.054.
14. Ullah, K., Rosendahl, A.H., Izzi, V., Bergmann, U., Pihlajaniemi, T., Maki, J.M., and Myllyharju, J. (2017). Hypoxia-inducible factor prolyl-4-hydroxylase-1 is a convergent point in the reciprocal negative regulation of NF-kappaB and p53 signaling pathways. Sci Rep *7*, 17220. 10.1038/s41598-017-17376-0.
15. Guo, J., Chakraborty, A.A., Liu, P., Gan, W., Zheng, X., Inuzuka, H., Wang, B., Zhang, J., Zhang, L., Yuan, M., et al. (2016). pVHL suppresses kinase activity of Akt in a proline-hydroxylation-dependent manner. Science *353*, 929-932. 10.1126/science.aad5755.
16. Qian, X., Zhou, Q., Ouyang, Y., Wu, X., Sun, X., Wang, S., Duan, Y., Hu, Z., Hou, Y., Wang, Z., et al. (2025). Transferrin promotes fatty acid oxidation and liver tumor growth through PHD2-mediated PPARalpha hydroxylation in an iron-dependent manner. Proc Natl Acad Sci U S A *122*, e2412473122. 10.1073/pnas.2412473122.
17. Liu, X., Tang, J., Wang, Z., Zhu, C., Deng, H., Sun, X., Yu, G., Rong, F., Chen, X., Liao, Q., et al. (2024). Oxygen enhances antiviral innate immunity through maintenance of EGLN1-catalyzed proline hydroxylation of IRF3. Nat Commun *15*, 3533. 10.1038/s41467-024-47814-3.
18. Wang, Z., Yan, M., Ye, L., Zhou, Q., Duan, Y., Jiang, H., Wang, L., Ouyang, Y., Zhang, H., Shen, Y., et al. (2024). VHL suppresses autophagy and tumor growth through PHD1-dependent Beclin1 hydroxylation. EMBO J *43*, 931-955. 10.1038/s44318-024-00051-2.
19. Erber, L., Luo, A., and Chen, Y. (2019). Targeted and Interactome Proteomics Revealed the Role of PHD2 in Regulating BRD4 Proline Hydroxylation. Mol Cell Proteomics *18*, 1772-1781. 10.1074/mcp.RA119.001535.
20. Moser, S.C., Bensaddek, D., Ortmann, B., Maure, J.F., Mudie, S., Blow, J.J., Lamond, A.I., Swedlow, J.R., and Rocha, S. (2013). PHD1 links cell-cycle progression to oxygen sensing through hydroxylation of the centrosomal protein Cep192. Dev Cell *26*, 381-392. 10.1016/j.devcel.2013.06.014.
21. Strowitzki, M.J., Cummins, E.P., and Taylor, C.T. (2019). Protein Hydroxylation by Hypoxia-Inducible Factor (HIF) Hydroxylases: Unique or Ubiquitous? Cells *8*. 10.3390/cells8050384.
22. Mann, M., Ong, S.E., Gronborg, M., Steen, H., Jensen, O.N., and Pandey, A. (2002). Analysis of protein phosphorylation using mass spectrometry: deciphering the phosphoproteome. Trends Biotechnol *20*, 261-268. 10.1016/s0167-7799(02)01944-3.
23. Steen, H., Jebanathirajah, J.A., Rush, J., Morrice, N., and Kirschner, M.W. (2006). Phosphorylation analysis by mass spectrometry: myths, facts, and the consequences for qualitative and quantitative measurements. Mol Cell Proteomics *5*, 172-181. 10.1074/mcp.M500135-MCP200.
24. Bekker-Jensen, D.B., Bernhardt, O.M., Hogrebe, A., Martinez-Val, A., Verbeke, L., Gandhi, T., Kelstrup, C.D., Reiter, L., and Olsen, J.V. (2020). Rapid and site-specific deep phosphoproteome profiling by data-independent acquisition without the need for spectral libraries. Nat Commun *11*, 787. 10.1038/s41467-020-14609-1.
25. Bremang, M., Cuomo, A., Agresta, A.M., Stugiewicz, M., Spadotto, V., and Bonaldi, T. (2013). Mass spectrometry-based identification and characterisation of lysine and arginine methylation in the human proteome. Mol Biosyst *9*, 2231-2247. 10.1039/c3mb00009e.
26. Udeshi, N.D., Mertins, P., Svinkina, T., and Carr, S.A. (2013). Large-scale identification of ubiquitination sites by mass spectrometry. Nat Protoc *8*, 1950-1960. 10.1038/nprot.2013.120.
27. Morelle, W., and Michalski, J.C. (2007). Analysis of protein glycosylation by mass spectrometry. Nat Protoc *2*, 1585-1602. 10.1038/nprot.2007.227.
28. Illiano, A., Pinto, G., Melchiorre, C., Carpentieri, A., Faraco, V., and Amoresano, A. (2020). Protein Glycosylation Investigated by Mass Spectrometry: An Overview. Cells *9*. 10.3390/cells9091986.
29. Larsen, M.R., Trelle, M.B., Thingholm, T.E., and Jensen, O.N. (2006). Analysis of posttranslational modifications of proteins by tandem mass spectrometry. Biotechniques *40*, 790-798. 10.2144/000112201.
30. Witze, E.S., Old, W.M., Resing, K.A., and Ahn, N.G. (2007). Mapping protein post-translational modifications with mass spectrometry. Nat Methods *4*, 798-806. 10.1038/nmeth1100.

31. Doll, S., and Burlingame, A.L. (2015). Mass spectrometry-based detection and assignment of protein posttranslational modifications. ACS Chem Biol *10*, 63-71. 10.1021/cb500904b.
32. Aslebagh, R., Wormwood, K.L., Channaveerappa, D., Wetie, A.G.N., Woods, A.G., and Darie, C.C. (2019). Identification of Posttranslational Modifications (PTMs) of Proteins by Mass Spectrometry. Adv Exp Med Biol *1140*, 199-224. 10.1007/978-3-030-15950-4_11.
33. Cockman, M.E., Lippl, K., Tian, Y.M., Pegg, H.B., Figg, W.D.J., Abboud, M.I., Heilig, R., Fischer, R., Myllyharju, J., Schofield, C.J., and Ratcliffe, P.J. (2019). Lack of activity of recombinant HIF prolyl hydroxylases (PHDs) on reported non-HIF substrates. Elife *8*. 10.7554/eLife.46490.
34. Arsenault, P.R., Heaton-Johnson, K.J., Li, L.S., Song, D., Ferreira, V.S., Patel, N., Master, S.R., and Lee, F.S. (2015). Identification of prolyl hydroxylation modifications in mammalian cell proteins. Proteomics *15*, 1259-1267. 10.1002/pmic.201400398.
35. Zhou, T., Erber, L., Liu, B., Gao, Y., Ruan, H.B., and Chen, Y. (2016). Proteomic analysis reveals diverse proline hydroxylation-mediated oxygen-sensing cellular pathways in cancer cells. Oncotarget *7*, 79154-79169. 10.18632/oncotarget.12632.
36. Eberhardt, E.S., Panisik, N., Jr., and Raines, R.T. (1996). Inductive Effects on the Energetics of Prolyl Peptide Bond Isomerization: Implications for Collagen Folding and Stability. J Am Chem Soc *118*, 12261-12266. 10.1021/ja9623119.
37. Buszewski, B., and Noga, S. (2012). Hydrophilic interaction liquid chromatography (HILIC)--a powerful separation technique. Anal Bioanal Chem *402*, 231-247. 10.1007/s00216-011-5308-5.
38. Boersema, P.J., Mohammed, S., and Heck, A.J. (2008). Hydrophilic interaction liquid chromatography (HILIC) in proteomics. Anal Bioanal Chem *391*, 151-159. 10.1007/s00216-008-1865-7.
39. Yoshida, T. (2004). Peptide separation by Hydrophilic-Interaction Chromatography: a review. J Biochem Biophys Methods *60*, 265-280. 10.1016/j.jbbm.2004.01.006.
40. Fu, Q., Liang, T., Li, Z., Xu, X., Ke, Y., Jin, Y., and Liang, X. (2013). Separation of carbohydrates using hydrophilic interaction liquid chromatography. Carbohydr Res *379*, 13-17. 10.1016/j.carres.2013.06.006.
41. Bensaddek, D., Nicolas, A., and Lamond, A.I. (2015). Evaluating the use of HILIC in large-scale, multi dimensional proteomics: Horses for courses? Int J Mass Spectrom *391*, 105-114. 10.1016/j.ijms.2015.07.029.
42. Kim, G., Weiss, S.J., and Levine, R.L. (2014). Methionine oxidation and reduction in proteins. Biochim Biophys Acta *1840*, 901-905. 10.1016/j.bbagen.2013.04.038.
43. Luo, S., and Levine, R.L. (2009). Methionine in proteins defends against oxidative stress. FASEB J *23*, 464-472. 10.1096/fj.08-118414.
44. Klont, F., Bras, L., Wolters, J.C., Ongay, S., Bischoff, R., Halmos, G.B., and Horvatovich, P. (2018). Assessment of Sample Preparation Bias in Mass Spectrometry-Based Proteomics. Anal Chem *90*, 5405-5413. 10.1021/acs.analchem.8b00600.
45. Lee, D.C., Sohn, H.A., Park, Z.Y., Oh, S., Kang, Y.K., Lee, K.M., Kang, M., Jang, Y.J., Yang, S.J., Hong, Y.K., et al. (2015). A lactate-induced response to hypoxia. Cell *161*, 595-609. 10.1016/j.cell.2015.03.011.
46. Di Conza, G., Trusso Cafarello, S., Loroch, S., Mennerich, D., Deschoemaeker, S., Di Matteo, M., Ehling, M., Gevaert, K., Prenen, H., Zahedi, R.P., et al. (2017). The mTOR and PP2A Pathways Regulate PHD2 Phosphorylation to Fine-Tune HIF1alpha Levels and Colorectal Cancer Cell Survival under Hypoxia. Cell Rep *18*, 1699-1712. 10.1016/j.celrep.2017.01.051.
47. Luo, W., Lin, B., Wang, Y., Zhong, J., O'Meally, R., Cole, R.N., Pandey, A., Levchenko, A., and Semenza, G.L. (2014). PHD3-mediated prolyl hydroxylation of nonmuscle actin impairs polymerization and cell motility. Mol Biol Cell *25*, 2788-2796. 10.1091/mbc.E14-02-0775.
48. Sipila, K.H., Drushinin, K., Rappu, P., Jokinen, J., Salminen, T.A., Salo, A.M., Kapyla, J., Myllyharju, J., and Heino, J. (2018). Proline hydroxylation in collagen supports integrin binding by two distinct mechanisms. J Biol Chem *293*, 7645-7658. 10.1074/jbc.RA118.002200.
49. Rappu, P., Salo, A.M., Myllyharju, J., and Heino, J. (2019). Role of prolyl hydroxylation in the molecular interactions of collagens. Essays Biochem *63*, 325-335. 10.1042/EBC20180053.
50. Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G.A., Sonnhammer, E.L.L., Tosatto, S.C.E., Paladin, L., Raj, S., Richardson, L.J., et al. (2021). Pfam: The protein families database in 2021. Nucleic Acids Res *49*, D412-D419. 10.1093/nar/gkaa913.
51. Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., and Vilo, J. (2019). g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). Nucleic Acids Res *47*, W191-W198. 10.1093/nar/gkz369.

52. Kiefel, H., Bondong, S., Hazin, J., Ridinger, J., Schirmer, U., Riedle, S., and Altevogt, P. (2012). L1CAM: a major driver for tumor cell invasion and motility. Cell Adh Migr *6*, 374-384. 10.4161/cam.20832.
53. Trinkle-Mulcahy, L., Andersen, J., Lam, Y.W., Moorhead, G., Mann, M., and Lamond, A.I. (2006). Repo-Man recruits PP1 gamma to chromatin and is essential for cell viability. J Cell Biol *172*, 679-692. 10.1083/jcb.200508154.
54. Schmidt, T., Samaras, P., Dorfer, V., Panse, C., Kockmann, T., Bichmann, L., van Puyvelde, B., Perez-Riverol, Y., Deutsch, E.W., Kuster, B., and Wilhelm, M. (2021). Universal Spectrum Explorer: A Standalone (Web-)Application for Cross-Resource Spectrum Comparison. J Proteome Res *20*, 3388-3394. 10.1021/acs.jproteome.1c00096.
55. Thingholm, T.E., Jorgensen, T.J., Jensen, O.N., and Larsen, M.R. (2006). Highly selective enrichment of phosphorylated peptides using titanium dioxide. Nat Protoc *1*, 1929-1935. 10.1038/nprot.2006.185.
56. Villen, J., and Gygi, S.P. (2008). The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry. Nat Protoc *3*, 1630-1638. 10.1038/nprot.2008.150.
57. Qing, G.Y., Yan, J.Y., He, X.N., Li, X.L., and Liang, X.M. (2020). Recent advances in hydrophilic interaction liquid interaction chromatography materials for glycopeptide enrichment and glycan separation. Trac-Trend Anal Chem *124*. ARTN 115570 10.1016/j.trac.2019.06.020.
58. Wu, F., Tabang, D.N., Wang, D., Odorico, J.S., and Li, L. (2025). Proteome-wide Investigation of Proline Hydroxylation in Pancreatic Ductal Adenocarcinoma Using DiLeu Isobaric Labeling Strategy. Mol Cell Proteomics, 100969. 10.1016/j.mcpro.2025.100969.
59. Giansanti, P., Tsiatsiani, L., Low, T.Y., and Heck, A.J. (2016). Six alternative proteases for mass spectrometry-based proteomics beyond trypsin. Nat Protoc *11*, 993-1006. 10.1038/nprot.2016.057.
60. Swaney, D.L., Wenger, C.D., and Coon, J.J. (2010). Value of using multiple proteases for large-scale mass spectrometry-based proteomics. J Proteome Res *9*, 1323-1329. 10.1021/pr900863u.
61. Heroes, E., Lesage, B., Gornemann, J., Beullens, M., Van Meervelt, L., and Bollen, M. (2013). The PP1 binding code: a molecular-lego strategy that governs specificity. FEBS J *280*, 584-595. 10.1111/j.1742-4658.2012.08547.x.
62. Bollen, M., Peti, W., Ragusa, M.J., and Beullens, M. (2010). The extended PP1 toolkit: designed to create specificity. Trends Biochem Sci *35*, 450-458. 10.1016/j.tibs.2010.03.002.
63. Hubbard, M.J., and Cohen, P. (1991). Targeting subunits for protein phosphatases. Methods Enzymol *201*, 414-427. 10.1016/0076-6879(91)01038-4.
64. Cohen, P. (1989). The structure and regulation of protein phosphatases. Annu Rev Biochem *58*, 453-508. 10.1146/annurev.bi.58.070189.002321.
65. Smith, R.J., Cordeiro, M.H., Davey, N.E., Vallardi, G., Ciliberto, A., Gross, F., and Saurin, A.T. (2019). PP1 and PP2A Use Opposite Phospho-dependencies to Control Distinct Processes at the Kinetochore. Cell Rep *28*, 2206-2219 e2208. 10.1016/j.celrep.2019.07.067.
66. de Castro, I.J., Budzak, J., Di Giacinto, M.L., Ligammari, L., Gokhan, E., Spanos, C., Moralli, D., Richardson, C., de Las Heras, J.I., Salatino, S., et al. (2017). Repo-Man/PP1 regulates heterochromatin formation in interphase. Nat Commun *8*, 14048. 10.1038/ncomms14048.
67. Foxler, D.E., Bridge, K.S., James, V., Webb, T.M., Mee, M., Wong, S.C., Feng, Y., Constantin-Teodosiu, D., Petursdottir, T.E., Bjornsson, J., et al. (2012). The LIMD1 protein bridges an association between the prolyl hydroxylases and VHL to repress HIF-1 activity. Nat Cell Biol *14*, 201-208. 10.1038/ncb2424.
68. Ong, S.E., and Mann, M. (2006). A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC). Nat Protoc *1*, 2650-2660. 10.1038/nprot.2006.427.
69. Hughes, C.S., Moggridge, S., Muller, T., Sorensen, P.H., Morin, G.B., and Krijgsveld, J. (2019). Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. Nat Protoc *14*, 68-85. 10.1038/s41596-018-0082-x.
70. Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol *26*, 1367-1372. 10.1038/nbt.1511.
71. Tyanova, S., Temu, T., and Cox, J. (2016). The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. Nat Protoc *11*, 2301-2319. 10.1038/nprot.2016.136.
72. Tyanova, S., and Cox, J. (2018). Perseus: A Bioinformatics Platform for Integrative Analysis of Proteomics Data in Cancer Research. Methods Mol Biol *1711*, 133-148. 10.1007/978-1-4939-7493-1_7.

73. Culver, C., Sundqvist, A., Mudie, S., Melvin, A., Xirodimas, D., and Rocha, S. (2010). Mechanism of hypoxia-induced NF-kappaB. Mol Cell Biol *30*, 4901-4921. 10.1128/MCB.00409-10.

74. Perez-Riverol, Y., Bai, J., Bandla, C., Garcia-Seisdedos, D., Hewapathirana, S., Kamatchinathan, S., Kundu, D.J., Prakash, A., Frericks-Zipper, A., Eisenacher, M., et al. (2022). The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. Nucleic Acids Res *50*, D543-D552. 10.1093/nar/gkab1038.

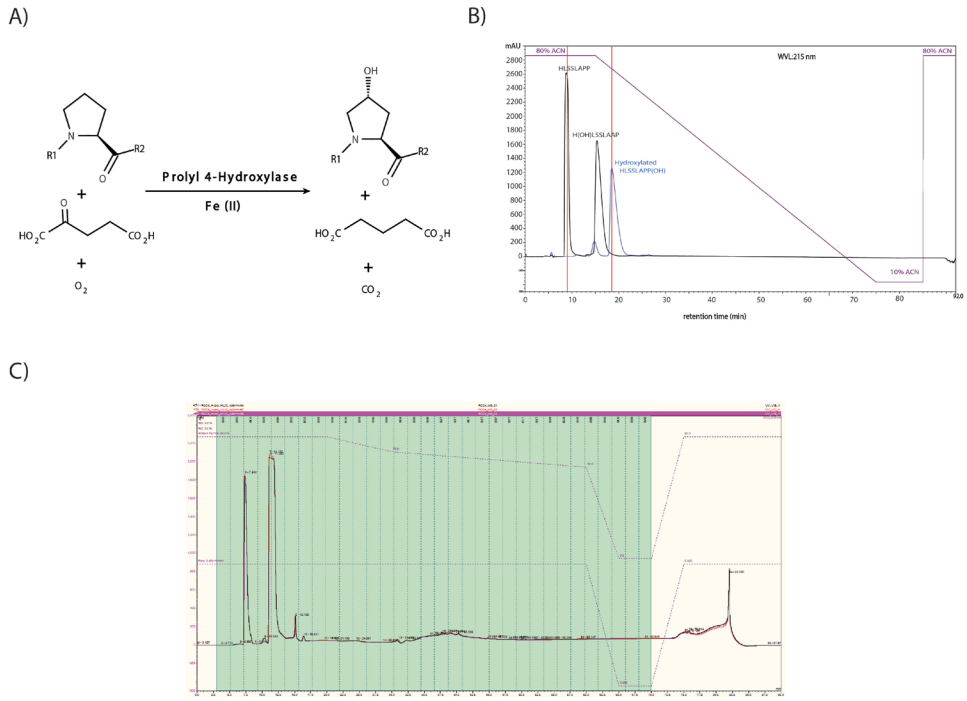


Figure 1: A) The schematic diagram of proline hydroxylation. B) HILIC chromatography showing the retention time difference of the synthetic peptides (HLSSLAPP, with no modification, or hydroxylation at the histidine/proline). C) HILIC chromatography profile of peptides fractionation from RCC4 samples (3 replicates, overlay).

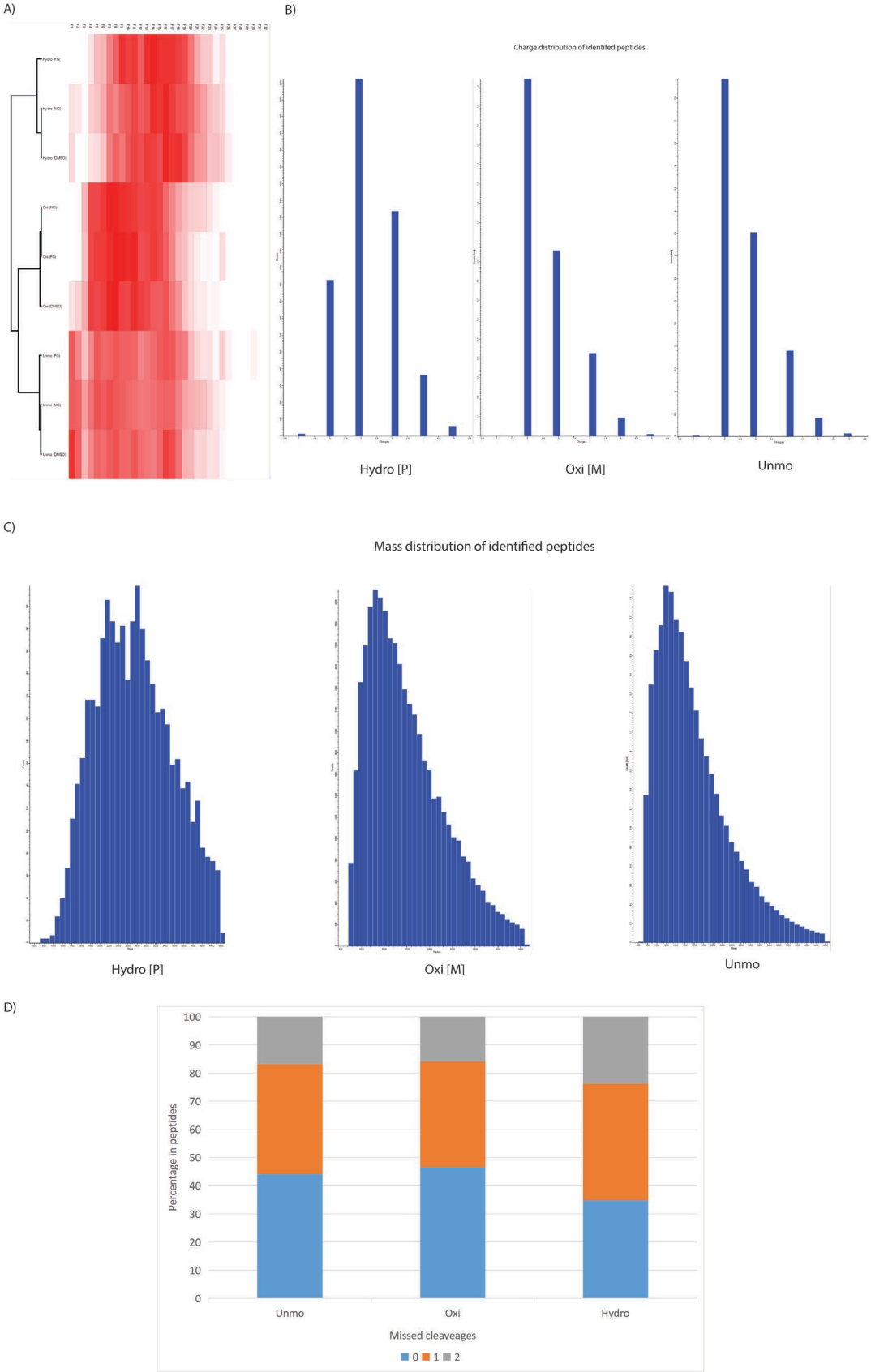


Figure 2: HEK293 dataset, A) The hydrophilicity difference of peptides with hydroxylation (P), oxidation(M) and unmodified peptides across HILIC fractions from

all experiments with different treatment. The hydrophilicity increases from the Fraction 1 to Fraction 32. B) The charge distribution of peptides with hydroxylation (P), oxidation (M) and unmodified peptides. C) The mass distribution of peptides with hydroxylation (P), oxidation (M) and unmodified peptides. D) The missed cleavage difference of peptides with hydroxylation (P), oxidation (M) and unmodified peptides.

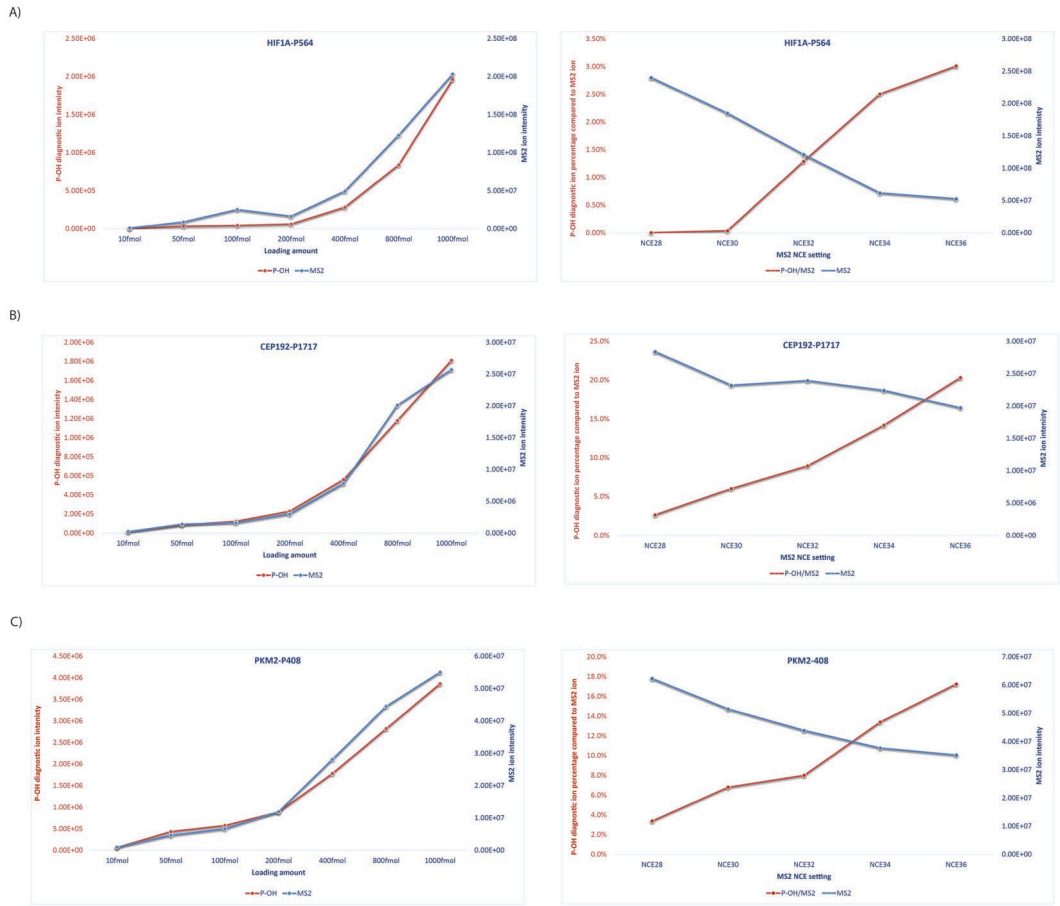


Figure 3: A) The intensity behaviour of HyPro diagnostic ion (m/z at 86.060) under different loading amount (left) and NCE settings (right), using the synthetic peptide HIF1α-P564, LDLEMLAPYIPMDDD (m/z at 833.9018). B) The intensity behaviour of HyPro diagnostic ion (m/z at 86.060) under different loading amount (left) and NCE settings (right), using the synthetic peptide CEP192-1717, WHLSSLAPPYVK (m/z at 707.38). C) The intensity behaviour of HyPro diagnostic ion (m/z at 86.060) under different loading amount (left) and NCE settings (right), using the synthetic peptide LAPITSDP (8) TEATAVGAVEASFK (m/z at 707.38). All ions were extracted under ±10ppm mass tolerance.

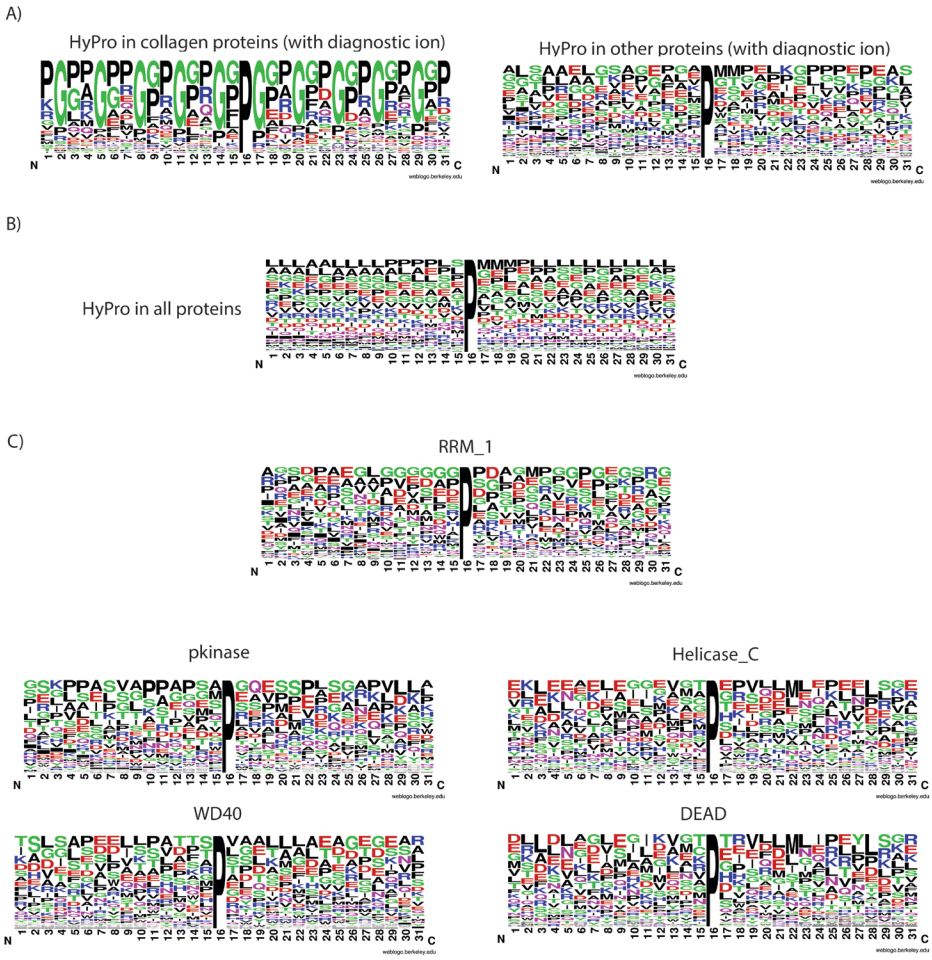


Figure 4: A) Sequence Logo for peptide window with hydroxylated proline site from collagen proteins (with diagnostic ion) and from other proteins (with diagnostic ion). B) Sequence Logo for peptide window with hydroxylated proline site from all the proteins. C) Sequence Logo for proline hydroxylated peptides from proteins included in different protein families (RRM_1 domain, pkinase domain, Helicase_C domain, WD40 domain and DEAD domain)

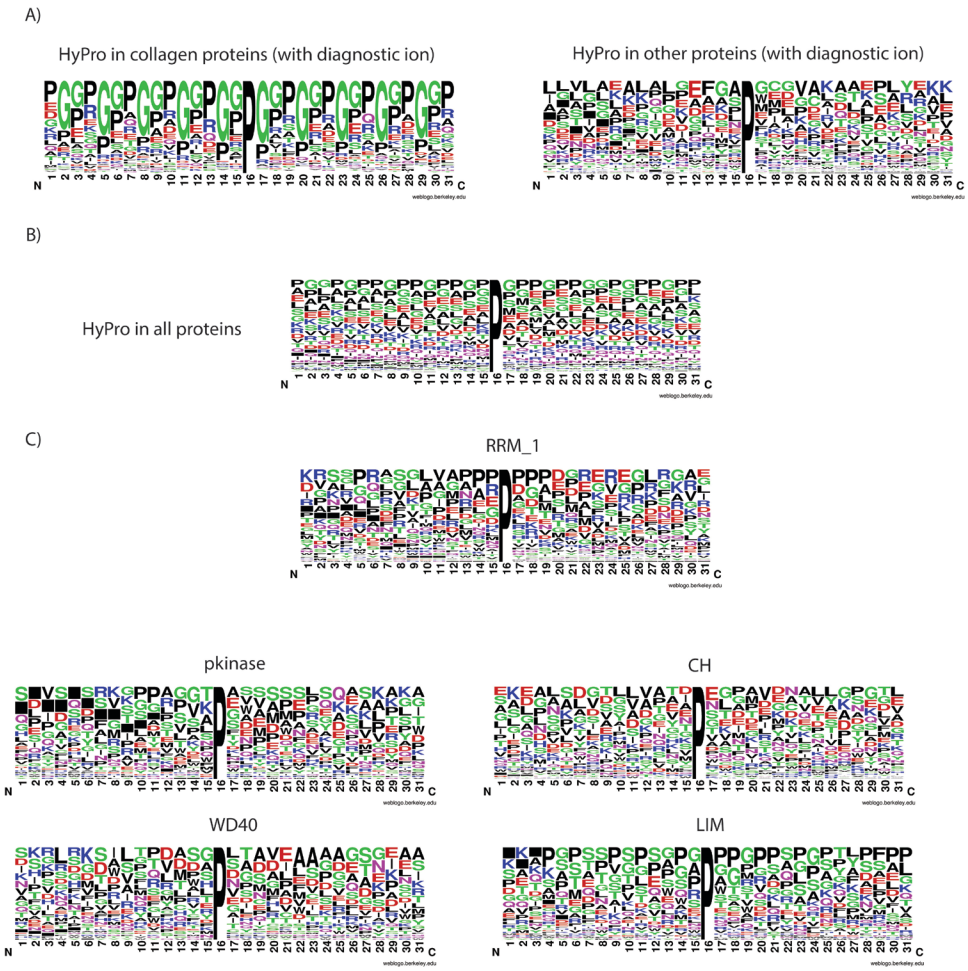


Figure 5: A) Sequence Logo for peptide window with hydroxylated proline site from collagen proteins (with diagnostic ion) and from other proteins (with diagnostic ion). B) Sequence Logo for peptide window with hydroxylated proline site from all the proteins. C) Sequence Logo for proline hydroxylated peptides from proteins included in different protein families (RRM_1 domain, pkinase domain, CH domain, WD40 domain and LIM domain)

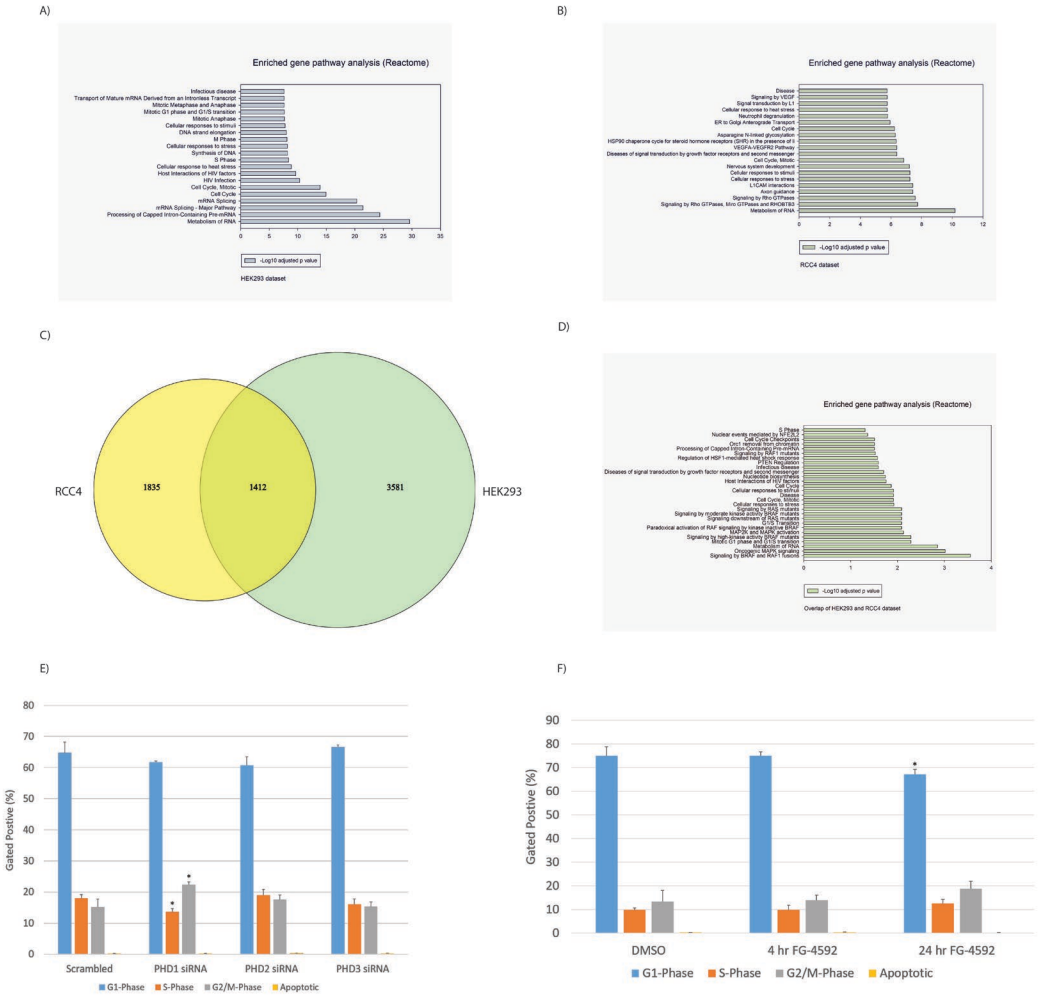


Figure 6: A) Statistical gene enrichment pathway analysis (Reactome) of the proteins only hydroxylated in DMSO/MG-132 treated HEK293 cells but not in FG-4592 treatment. B) Statistical gene enrichment pathway analysis (Reactome) of the proteins only hydroxylated in DMSO/MG-132 treated RCC4 cells but not in FG-4592 treatment. C) The Venn diagram of the identified proline hydroxylation sites between HEK293 and RCC4 dataset. D) Statistical gene enrichment pathway analysis (Reactome) of the proteins, which not only have the same hydroxylation proline sites from HEK293 and RCC4 dataset (DMSO/MG-132 treated), but also not hydroxylated in any FG-4592 treatment. E) RCC4 cells were treated with the indicated siRNAs and the cell cycle profile was determined by flow cytometry. The values presented represent the averages from three independent experiments. The error bars indicate the standard deviations. F) RCC4 cells were treated with the PHD inhibitor FG-4592 for the indicated times points

prior to fixing. After which the cell cycle profile was determined by flow cytometry. The averages from three independent experiments are presented, whilst the error bars indicate the standard deviations. The cell profiles were gated according to the control in each independent experiment and the number of cells in each phase are expressed as a percentage of the total number of cells present. Statistical analysis was performed according to the Student's T test vs the control; *p < 0.05.

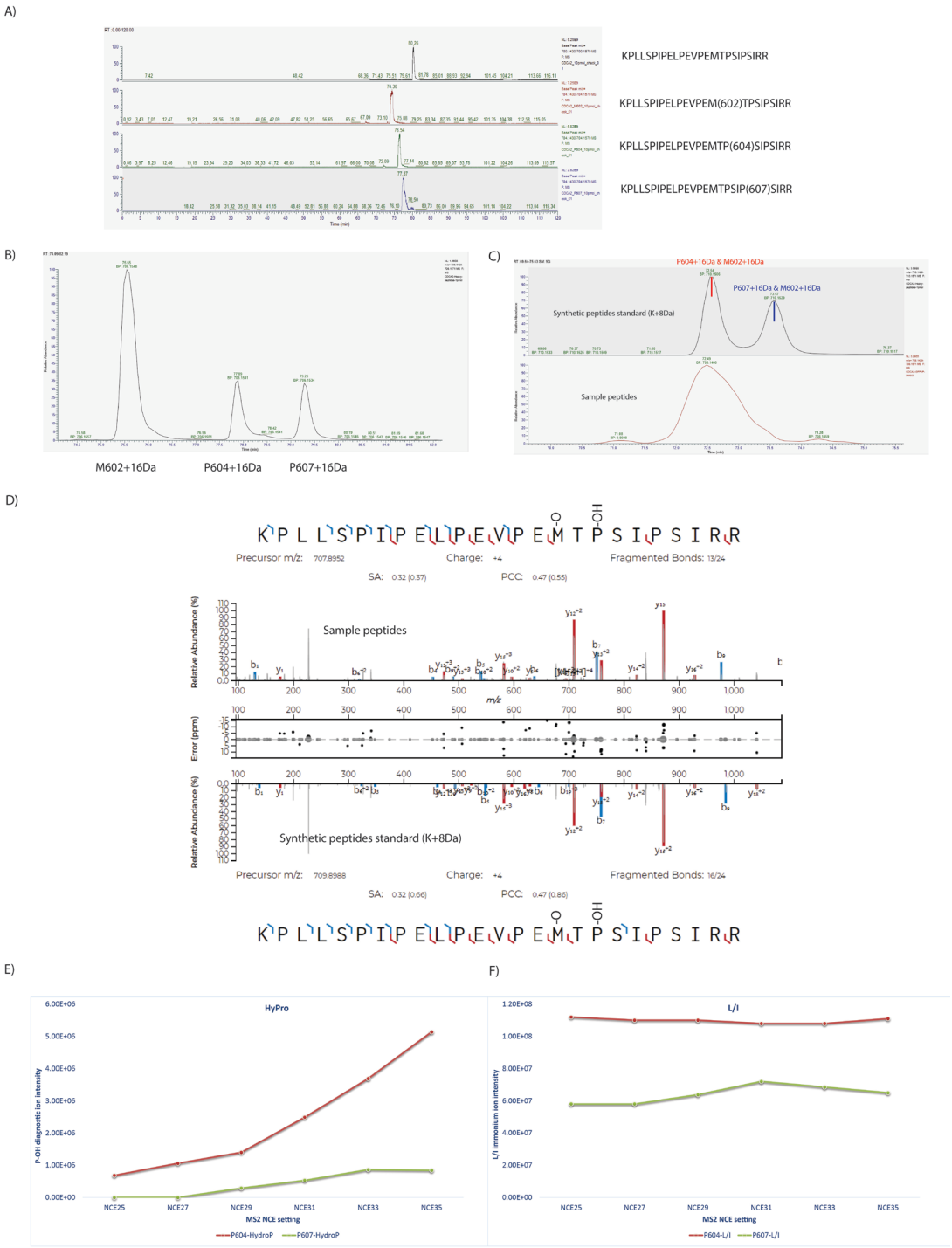


Figure 7: A) The retention time difference of the synthetic peptides (KPLLSPIPELPEVPEMTPSIPSIRR) with or without oxidation/hydroxylation (+16Da mass shift) at different amino acids using RPLC-MS analysis, individually. B) The retention time difference of the synthetic peptides (KPLLSPIPELPEVPEMTPSIPSIRR, K+8Da) with or without oxidation/hydroxylation (+16Da mass shift) at different amino acids using RPLC-MS analysis, mixed. C) The retention time comparison of the M602 oxidised peptides (KPLLSPIPELPEVPEM(O)TPSIPSIRR) with hydroxylation (+32Da mass shift) at different proline residues, between the spiked-in heavy isotope labelled synthetic peptide and the peptide in the IP sample. D) The MS/MS spectrum of the sequence KPLLSPIPELPEVPEM(O)TP(OH)SIPSIRR from the spiked-in heavy isotope labelled synthetic peptide (scan no. 21243) and the peptide from the IP sample (scan no. 25891). The intensity behaviour of E) HyPro diagnostic ion (m/z at 86.060) and F) L/I immonium ion (m/z at 86.097) under different NCE settings, using the synthetic hydroxylated peptide KPLLSPIPELPEVPEMTP(604)SIPSIRR and KPLLSPIPELPEVPEMTPSIP(607)SIRR (m/z at 704.15). All ions were extracted under ±10ppm mass tolerance.

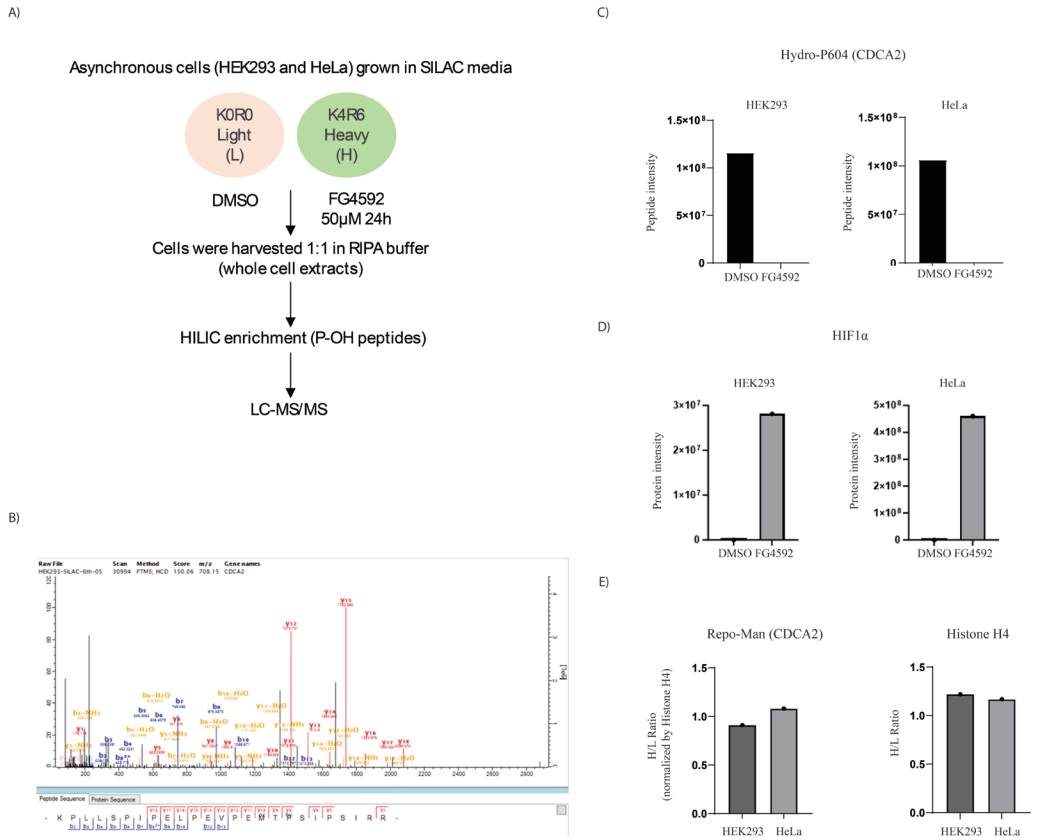


Figure 8: MS analysis of Repo-Man hydroxylation at P604. A) Diagram of proline hydroxylation sites analysis in asynchronous HeLa and HEK293 cells using SILAC, Cells were grown in SILAC media for 6 passages and 24 h before harvesting cells were treated with FG-4592 50 μM or DMSO. DMSO treated cells (control) were light labelled, and FG-4592 treated cells are heavy labelled. B) MS/MS spectra of endogenous Repo-Man peptide (KPLLSPIPELPEVPEMTPSIPSIRR), where matched fragment ions (b/y) were marked. Both the peptide precursor ion (m/z 708.15) and its corresponding fragment ions (y8 and y10) confirmed the identification of KPLLSPIPELPEVPEM(602-Ox)TP(604-OH)SIPSIRR. The diagnostic ion (m/z 86.06) of hydroxylated proline was also shown as "P-OH". C) Quantification analysis of P604 hydroxylation in Repo-Man in both HeLa and HEK293 cells, under the treatment of either DMSO or FG-4592. The bar chart was based on the intensity of the peptide containing P604 hydroxylation. D) Quantification analysis of HIF1α in both HeLa and HEK293 cells, where the protein raw intensity was used for the bar chart since there's no HIF1α detected in DMSO treated cells. E) Quantification analysis of Repo-Man (CDCA2) and Histone H4 in both HeLa and HEK293 cells, where H/L represents the heavy to light ratio using SILAC quantification.

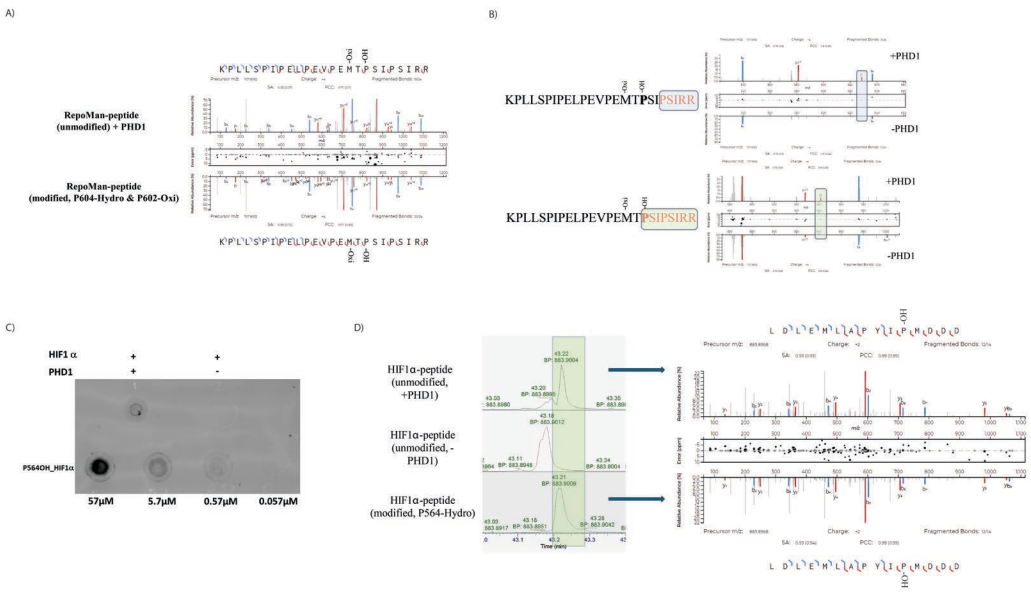


Figure 9: *In vitro* hydroxylation of Repo-Man synthetic peptide by recombinant PHD1. The reactions were performed in parallel using Repo-Man (KPLLSPIPELPEVPEM(602-OX)TPSIPSIRR) and HIF1α (LDLEMLAPYIPMDDD) synthetic peptides as substrates in combination with (+) or without (-) recombinant PHD1. A) MS/MS spectra of the hydroxylated peptide KPLLSPIPELPEVPEM(602-

1099 OX)TP(604-OH)SIPSIRR from Repo-Man. The mirror image shows the fragment ions
1100 matching results between the experimental hydroxylated peptide (top) and its synthetic
1101 standard peptide (bottom), where matched b/y ions were highlighted. B) Detailed
1102 MS/MS spectra of the hydroxylated synthetic peptide KPLLSPIPELPEVPEM(602-
1103 OX)TP(604-OH)SIPSIRR from Repo-Man. The mirror image shows the fragment ions
1104 matching results of the synthetic peptide incubated with or without PHD1. The y8 ion
1105 represents the fragment of [P(604-OH)SIP(607)SIRR], while the y5 ion represents the
1106 fragment of [P(607)SIRR], which confirmed the hydroxylation takes place at P604 only
1107 when incubation with PHD1. C) Dot blot analysis of the in vitro hydroxylation of
1108 HIF1α peptide, with and without recombinant PHD1 (upper dots). Titration of the OH-
1109 564 HIF1α peptide as a control of the primary antibody. Membrane was incubated with
1110 anti-OH-564 HIF1α and developed using far-red secondary antibody in typhoon. D)
1111 LC-MS analysis of *in vitro* hydroxylation of HIF1α peptide (LDLEMLAPYIPMDDD).
1112 The left panel shows the retention time difference in LC separation profile of HIF1α
1113 synthetic peptide with or without PHD1, compared to the HIF1α hydroxylated (P546-
1114 OH) synthetic peptide. The mirror image in the right panel shows the MS/MS fragment
1115 ions matching results between the experimental hydroxylated HIF1α peptide (top) and
1116 its synthetic standard peptide (bottom), where matched b/y ions were highlighted.

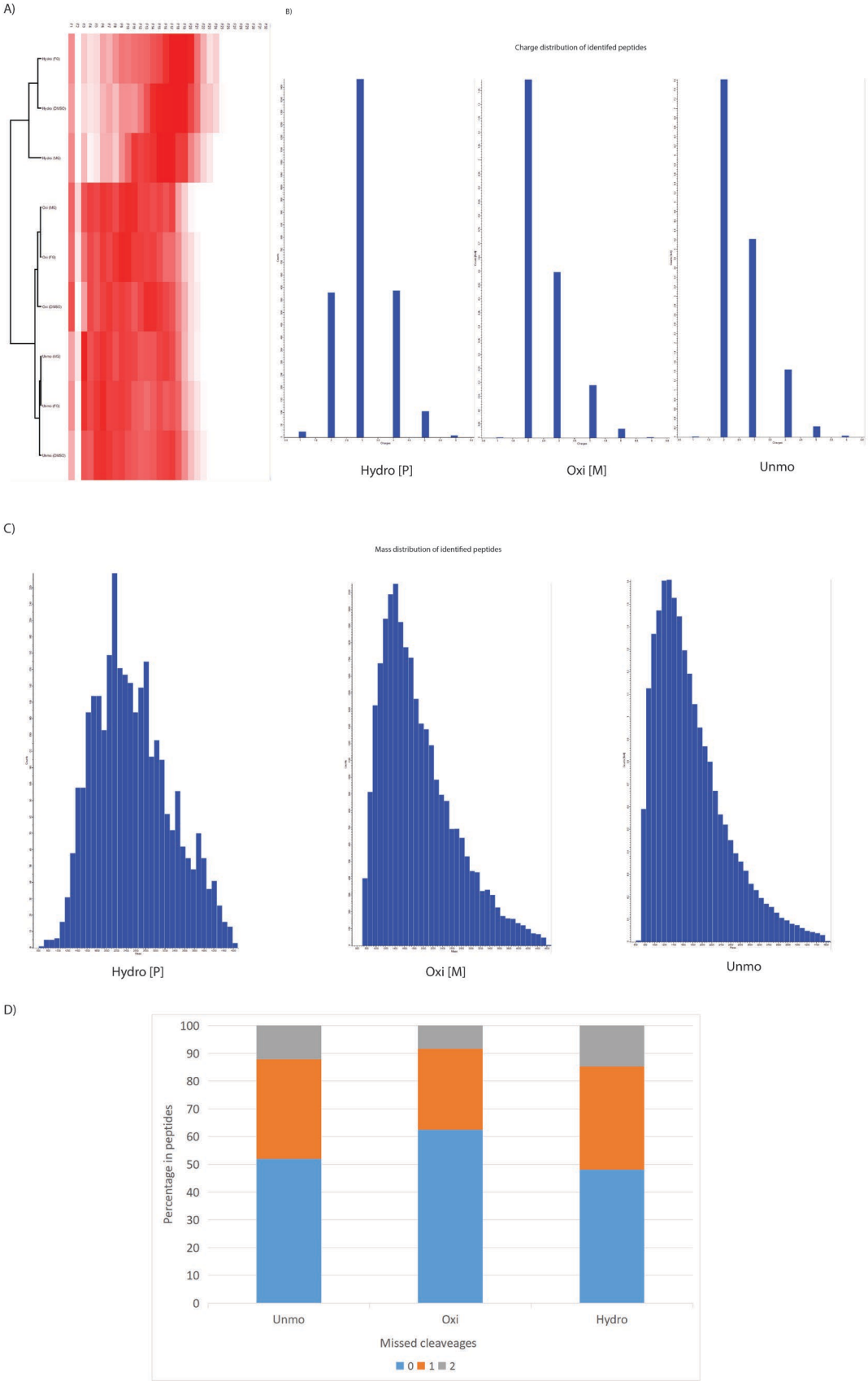


Figure S1: RCC4 dataset, A) The hydrophilicity difference of peptides with hydroxylation (P), oxidation(M) and unmodified peptides across HILIC fractions from

all experiments with different treatment. The hydrophilicity increases from the Fraction 1 to Fraction 32. B) The charge distribution of peptides with hydroxylation (P), oxidation (M) and unmodified peptides. C) The mass distribution of peptides with hydroxylation (P), oxidation (M) and unmodified peptides. D) The missed cleavage difference of peptides with hydroxylation (P), oxidation (M) and unmodified peptides.

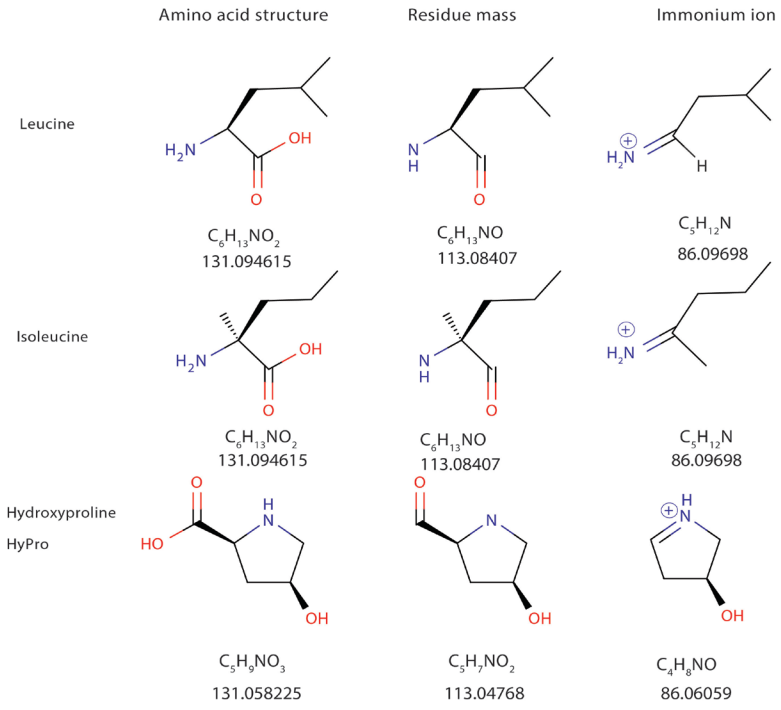


Figure S2: The structure, mass and immonium ion difference of Leucine, isoleucine and hydroxyproline.

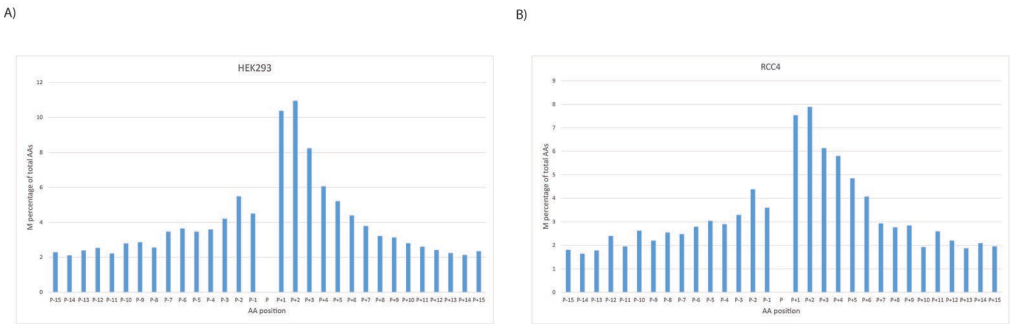
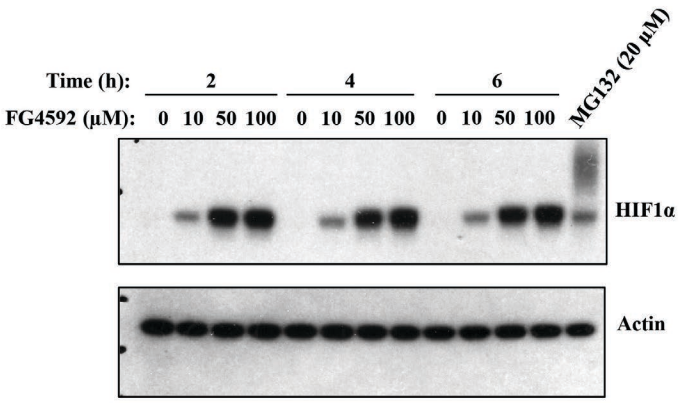


Figure S3: The percentage of M in total amino acids at each position from A) HEK293 and B) RCC4 dataset, centred at the hydroxyproline residue.

A)



B)

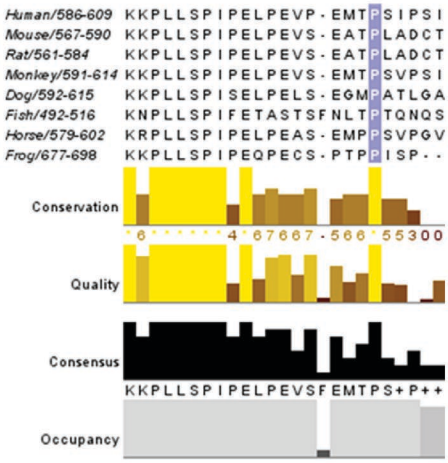


Figure S4: A) Western blot analysis of HIF1α from HEK293 cell lysates treated with different concentrations and times of FG-4592 as indicated in the figure. Actin was used as a loading control. MG-132 was added for 3 h before cells were harvested. B) Sequence alignment of Repo-Man peptide region (covering P604) across different species is displayed.