Enantioselective Protein Affinity Selection Mass Spectrometry (EAS-MS)

Xiaoyun Wang[1,#], Jianxian Sun[1,2,#], Shabbir Ahmad[2,#], Diwen Yang[3], Fengling Li[2], U Hang Chan[2,4], Hong Zeng[2], Conrad V. Simoben [2], Scott Houliston[5], Aiping Dong[2], Albina Bolotokova[2], Elisa Gibson[2], Maria Kutera[2,5,6], Pegah Ghiabi[2], Ivan Kondratov[7,8,9], Tetiana Matviyuk[7,10], Alexander Chuprina[7], Danai Mavridi[11,12], Christopher Lenz[11,12], Andreas C. Joerger[11,12], Benjamin D. Brown[13], Richard B. Heath[13], Wyatt W. Yue[13], Lucy K. Robbie[14], Tyler S. Beyett[14], Susanne Müller[11,12], Stefan Knapp[11,12], Rachel Harding[2,4,15], Matthieu Schapira[2,4], Peter J. Brown[16], Vijayaratnam Santhakumar[2], Suzanne Ackloo[2], Cheryl H. Arrowsmith[2,5,6], Aled M. Edwards[2], Hui Peng[1,2,3*], Levon Halabelian[2,4*]

[1]Department of Chemistry, University of Toronto, Toronto, ON, Canada

[2]Structural Genomics Consortium, University of Toronto, Toronto, ON, Canada

[3]Department of Physical & Environmental Sciences, University of Toronto Scarborough, Toronto ON, Canada

[4]Department of Pharmacology and Toxicology, University of Toronto, Toronto, ON, Canada

[5]Princess Margaret Cancer Centre, University of Toronto, Toronto, ON, Canada

[6]Department of Medical Biophysics, University of Toronto, Toronto, Ontario, Canada

[7]Enamine Ltd., Winston Churchill Street 78, 02094 Kyiv, Ukraine

[8]Enamine Germany GmbH, Industriepark Hoechst, G837, 65926 Frankfurt am Main, Germany

[9]V. P. Kukhar Institute of Bioorganic Chemistry and Petrochemistry, National Academy of Sciences of Ukraine, Akademik Kukhar Street 1, 02094 Kyiv, Ukraine

23 [10]Department of Immunology and Infectious Diseases, Harvard T. H. Chan School of Public

24 Health, Boston, MA, USA.

25 [11]Institute of Pharmaceutical Chemistry, Goethe University, Frankfurt am Main, Germany

26 [12]Structural Genomics Consortium, Goethe University, Frankfurt am Main, Germany.

27 [13]Biosciences Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon

28 Tyne, UK.

29 [14] Emory University School of Medicine, Atlanta, GA, USA

30 [15]Leslie Dan Faculty of Pharmacy, University of Toronto, Toronto, ON, Canada

31 [16]Structural Genomics Consortium, Eshelman School of Pharmacy, University of North Carolina

32 at Chapel Hill, Chapel Hill, NC 27599, USA

33 [#]These authors contribute equally to this work

34 *Corresponding authors:

35 **Hui Peng**: hui.peng@utoronto.ca; Structural Genomics Consortium, Department of Chemistry,

36 University of Toronto

37 **Levon Halabelian**: l.halabelian@utoronto.ca; Tel: 416 946 3876; Structural Genomics

38 Consortium, Department of Pharmacology and Toxicology, University of Toronto, Toronto

39   **ABSTRACT**

40   We report an enantioselective protein affinity selection mass spectrometry screening approach

41   (EAS-MS) that enables the detection of weak binders, informs about selectivity, and generates

42   orthogonal confirmation of binding. After method development with control proteins, we

43   screened 31 human proteins against a designed library of 8,210 chiral compounds. 16 binders to

44   12 targets, including many proteins predicted to be "challenging to ligand", were discovered and

45   confirmed in orthogonal assays. 7 binders to 6 targets bound in an enantioselective manner, with

46   K_Ds ranging from 3 to 20 µM. Binders for four targets (DDB1, WDR91, WDR55, and HAT1)

47   were selected for in-depth characterization using X-ray crystallography. In all four cases, the

48   mechanism for enantioselective selectivity was readily explained. EAS-MS can be used to

49   identify and characterize selective and weakly-binding ligands for novel protein targets with

50   unprecedented throughput and sensitivity.

51

52   **MAIN**

53   Bioactive small molecules are invaluable reagents in basic research and applied fields such as

54   biomedicine, agriculture and microbiology. The discovery of bioactive small molecules typically

55   begins with screening of synthetic chemical or natural product libraries. Often, libraries of

56   synthetic chemical or natural product libraries are screened in high-throughput or phenotypic

57   screening[1-3] to identify compounds that alter biochemical or cellular functions. Alternatively,

58   target-centric approaches can be employed, such as methods that measure binding to the target

59   protein directly in vitro, including fragment-based lead discovery (FBLD)[4-6], DNA-encoded

60   library (DEL) selection, or affinity selection mass spectrometry (ASMS)[7]. A significant

61   challenge across all screening strategies is to develop orthogonal assays to distinguish true hits

62   from false positives. Indeed, particularly when the initial hits bind weakly (> 10 µM K_D), this hit

3

63   verification process often demands more time and resources than the initial screen. As a result, it

64   has proven impractical for the discovery of chemical ligands for large numbers of proteins.

65   Small molecules with a stereocenter may bind to their targets in an enantioselective way[10], which

66   can be leveraged to identify potential steroselective protein binders when screening libreraied

67   containing chiral compounds, as well as to develop negative controls for chemoproteomics

68   studies and for experiments with chemical probes[11-13]. To exploit this phenomenon, we

69   developed a scalable orthogonal screening strategy, termed "Enantioselective Protein Affinity

70   Selection Mass Spectrometry (EAS-MS)", to both identify and characterize chemical ligands for

71   previously unliganded proteins.

72

73   **RESULTS**

74   **Design and benchmarking an Enantioselective Protein Affinity-Selection Mass**

75   **Spectrometry (EAS-MS) platform**

76   *Method concept*: The EAS-MS screening concept workflow (Fig. 1a) is a variation of previous

77   methods used for bioactive natural product discovery[14-16], by including the orthogonal

78   enantioselectivity evidence. First, pools of ~600 compounds from an 8,210 member screening

79   library (Supplementary Table 1), comprising racemic mixtures of drug-like compounds at a

80   concentration of 0.1 µM, are incubated with purified and quality-controlled polyhistidine-tagged

81   proteins immobilized on magnetic nickel beads. At this concentration, all compounds are far

82   below their solubility limit in aqueous solution. To embed a measure of selectivity in the screen,

83   8-12 different proteins are screened in parallel at a concentration of 1 µM in solution. This

84   translates to an effective protein concentration of 100 µM on the beads. After washing the beads,

85   compounds that remain bound to the proteins and the beads, which comprise a combination of

86   real binders and nonspecific binders, are eluted in methanol. A small aliquot of each methanol

87   eluate is subject to liquid chromatography (LC) followed by high-resolution mass spectrometry

88   in order to identify any compounds in the eluates, as well as to estimate their abundance relative

89   to the eluates from other targets screened in parallel. Our concept is that some measure of

90   binding specificity can be inferred by observing strong binding of a compound to one protein but

91   not to the other proteins screened in parallel (8-12 proteins in each batch). For the compounds

92   that appear to bind one target preferentially over the others, another aliquot from the same

93   methanol eluate is analyzed by chiral chromatography, and the ratio of the two enantiomers in

94   the protein eluate is assessed. If one of the two enantiomers is enriched in the protein eluate

95   compared to their ratio in the screening library, then this provides orthogonal evidence for

96   compound binding.

97   *Proof of concept for EAS-MS with known enantioselective protein-ligand pairs.* To pilot the

98   EAS-MS concept, we selected four chiral compounds known to bind to their respective protein

99   partners in an enantioselective manner. The first test case was lenalidomide, a chiral analogue of

100   thalidomide that binds enantioselectively to the cereblon (CRBN) protein[17-19]. As shown in Fig.

101   1b, when screened against four proteins (CRBN, DCAF1, PRMT5, and USP21), lenalidomide

102   was selectively enriched in the methanol eluate from CRBN beads; no significant enrichment

103   was observed on beads containing any of the three other human proteins, which served as

104   negative controls. The CRBN methanol eluate was then subjected to chiral chromatography. We

105   recovered (*S*)-lenalidomide in the eluate from CRBN-containing beads at far greater quantities

106   compared to its (*R*-) enantiomer, despite both being present in equal amounts in the input.

107   Enantioselective binding was also observed for (*S*)-OICR0036766, (*S*)-LLY283, and (*R*)-

108   BAY805, to their cognate protein targets, DCAF1[20], PRMT5[21], and USP21[22] (Fig. 1c, 1d, and

109   1e).

110   *Racemate library design and characterization*: With proof of concept in hand, we set out to

111   apply the method to novel proteins, in screening mode. To this end, we assembled a diverse,

112   drug-like screening library comprising 8,210 commercially available racemates, and developed

113   chromatographic method(s) to resolve the enantiomers for as many of the racemates as was

114   practical. These chemicals were selected by considering chemical diversity, drug-like properties

115   and MS sensitivity (Fig. 1f and 1 g). To accomplish this, each of the 8,210 racemic mixtures was

116   subject to chiral liquid chromatography under 8 different conditions (4 columns $\times$ 2 LC methods).

117   In total, ~7,000 enantiomer pairs could be resolved under one of these conditions. The most

118   effective chromatographic steps involved using IG and IA columns (Fig. 1h). Thus, for any of

119   these 7,000 compounds, it is possible to compare the ratio of enantiomers in the screening library

120   and the protein eluant, as described in the workflow outlined above (Fig 1a).

121   *EAS-MS identifies low affinity binders*. Rapid size-exclusion chromatography (SEC) is the most

122   common approach used to resolve protein binders and non-binders in AS-MS library screening[7].

123   However, this method is known to be limited by the inability to detect weakly bound compounds,

124   perhaps due to the challenges in retaining compounds with fast off-rates during chromatography.

125   We explored whether the EAS-MS platform in which the protein is concentrated on beads and

126   employs rapid wash steps, had the potential to enhance the sensitivity of AS-MS screens. Using

127   102 previously-characterized ligands across 8 proteins, we tested the sensitivity of EAS-MS. The

128   EAS-MS platform successfully captured nearly all high-affinity ($K_D < 1$ μM) ligands, half of the

129   moderate-affinity (1 μM $< K_D <$ 10 μM) ligands, and 20% of the low-affinity ($K_D >$ 10 μM)

130   ligands (Fig. 1i) without substantially increasing the false positive rate. In contrast, all moderate-

131   and low-affinity ligands were lost by the SEC-coupled AS-MS approach. The apparent

132   sensitivity of EAS-MS is a significant advantage for chemical ligand discovery, or to assess the

6

133   ligandability of a new protein, as binders can be discovered within a relatively small chemical

134   library and by using lower amounts of protein.

135

136   **Application of the EAS-MS platform to novel targets**

137   With evidence of increased sensitivity and an ability to potentially provide orthogonal evidence

138   for specificity using enantioselectivity, we set out to explore the application of the EAS-MS

139   platform to new proteins. We selected 31 human proteins with diverse biological functions and

140   with varying levels of known or predicted 'ligandability' (Fig. 2a) as assessed by their drug-like

141   density (DLID) scores[23]. We purposely included several targets (*e.g.,* WDR91 and DCAF1)

142   known to be 'ligandable' [20,24,25]. Among the 31 proteins, 3 were predicted to have high

143   ligandability (DLID > 1), 11 of medium ligandability (0.5 < DLID < 1), and 16 low ligandability

144   (DLID < 0.5) (Supplementary Fig. 1).

145   The 31 proteins were grouped in 4 batches of 8 proteins and screened against 14 pools of ~600

146   compounds. In total, 118 compounds (candidate binders) exhibited > 5-fold enrichment for one

147   protein compared to the others screened in parallel, corresponding to an average hit rate of

148   ~0.014% (Fig. 2b, and Supplementary Table 2). Candidate binders were detected for all three

149   high ligandable targets, seven of the eleven targets with medium ligandability scores, and nine of

150   the sixteen targets predicted to be the most challenging. Importantly, these hits were distributed

151   across the entire chemical space of the EAS-MS library (Supplementary Fig. 2).

152   To confirm binding, we tested the binding of 101 candidate binders to each of the 18 targets

153   using orthogonal methods. To this end, we used surface plasmon resonance spectroscopy (SPR),

154   a technique that requires relatively small amounts of protein and when successful can provide

155   quantitative orthogonal confirmation of binding. Ideally, SPR assay methods are developed using

7

156   a positive control binder, which we lacked for nearly all of the proteins. Accordingly, positive

157   results by SPR can be interpreted, but negative results are inconclusive.

158   The binding of 16 hits to 12 targets was confirmed by SPR, ranging from 1-4 hits per target and

159   with K_D values ranging from 2 to 87 µM (Fig. 2c and 2d; full list in Supplementary Table 2).

160   Notably, 13 of the 16 SPR-confirmed hits exhibited low-affinity ($K_D > 10$ µM) (Fig. 1i).

161   Interestingly, the degree of enrichment in the primary screen of the SPR-validated hits showed a

162   strong correlation with the K_D values (Fig. 2e), demonstrating the potential application of the

163   EAS-MS method for semi-quantitative read-outs.

164   As argued above, the 85 candidate binders that were unable to be confirmed using SPR are not

165   necessarily false positives for three reasons. First, they might not be able to be detected because

166   the assay for the protein target could not be optimized using a positive control compound.

167   Second, they might also be *bone fide* binders with affinities beyond the sensitivity of SPR under

168   the non-optimized solution conditions used. Third, the compounds might be insoluble at the

169   concentrations required and conditions used for SPR (up to 0.2 mM of the compound) and might

170   confound the read-outs[26]. The third possibility is a common issue when attempting to

171   characterize weakly binding compounds by any biophysical method.

172

173   **Enantioselective binding provides orthogonal evidence of compound binding**

174   Our EAS-MS method, which creates an effective protein concentration of 100 µM on the beads,

175   enables the screening of compounds at a concentration of 100 nM, far below their solubility

176   limits. In support of solubility being a confounding issue for SPR orthogonal validation, we

177   observed that SPR confirmation rates improved significantly for EAS-MS hits with higher fold

178   enrichment values. For EAS-MS hits without enantioselectivity that were enriched > 5 fold, the

179   SPR-confirmation rate was 12.2%; this increased to 22.2% for more potent hits with fold

180   enrichment values > 20 (Fig. 2f). For the more potent binders that were enantioselective

181   enrichment, the SPR confirmation rate increased further to 50%. Even subject to the caveats of

182   SPR, this trend clearly demonstrates the potential value of enantioselective enrichment in hit

183   confirmation.

184   Among the 118 candidate binders identified in the screen of 31 proteins, clear enantioselective

185   binding was detected for 32 binders to 14 targets (Fig. 2c). Of these targets, we were successful

186   in developing an SPR assay for 6, and could calculate K_D values for 7 binders, with affinities

187   ranging from 3-20 µM (Supplementary Figs. 4-8). There were 8 targets and 25 candidate binders

188   for which we were unable to generate convincing binding data using either SPR under standard

189   conditions or with other biophysical methods, such as differential scanning fluorimetry or [19]F-

190   NMR. At this point, we cannot be certain if these 25 enantioselective candidate binders are false

191   positives or are true positives that are challenging to assay.

192

193   **In-depth characterization of enantioselective binding to 4 targets**

194   Four targets – DDB1, WDR91, WDR55, and HAT1 - predicted to have 'medium' or 'low'

195   ligandbility scores, as measured by their DLID scores (Fig. 2a, Supplementary Fig. 1), were

196   selected for further analysis.

197

198   *Damage-specific DNA-binding protein 1 (DDB1)* is a multidomain protein involved in protein

199   homeostasis[27]. We screened histidine-tagged DDB1 and 7 other proteins in parallel against the

200   racemate library and identified a compound (XS381952) specifically enriched in the methanol

201   eluate from DDB1 beads (Fig. 3a). When XS381952 in the DDB1-bead eluate was analyzed by

202   chiral chromatography, one of the two enantiomers was clearly enriched compared with the ratio

203   of the enantiomers in the starting library, providing compelling evidence of specific binding (Fig.

204   3b). To rigorously characterize the binding, we resolved and characterized the two enantiomers

205   of XS381952 using preparative chiral chromatography and electronic circular dichroism (ECD)

206   spectroscopy (Supplementary Fig. 3), and measured their binding to DDB1 using SPR. We found

207   that (S)-XS381952 bound DDB1 with a K_D of 2 µM (Supplementary Fig. 4), significantly more

208   potent than its (R)-counterpart (estimated $K_D > 73$ µM). To elucidate the mechanism of binding,

209   we determined the crystal structure of (S)-XS381952 with DDB1 (Supplementary Table 3;

210   Supplementary Fig. 9a), and found that (S)-XS381952 is sandwiched between two aromatic

211   residues, W1047 and Y1114, via π-stacking interactions (Fig. 3c). The enantioselectivity was

212   explained by the positioning of the ethoxyphenyl moiety linked to the chiral carbon, which

213   adopted a nearly 90° angle from the compound backbone, orienting toward V1132 and

214   stabilizing against the 1129-1140 α-helix of DDB1. The (R)-enantiomer would not be able to fit

215   into the binding site due to steric hindrance.

216

217   *WD40 repeat containing protein 91 (WDR91)* is a 747-residue protein with a C-terminal WDR

218   domain that plays a critical role in endosomal maturation[29]. After screening the WDR domain of

219   WDR91 with racemate library, four compounds were enriched in WDR91-bead eluates

220   compared with the other proteins screened in parallel. Two of these hits, XS838489 (Fig. 3d, 3e)

221   and XS837729 (Supplementary Fig. 5), bound in an enantioselective manner. The remaining two

222   hits, XS381295 and XS381186, did not show enantioselectivity in its binding. The binding of all

223   four hits was characterized by SPR, and three hits displayed single digit micromolar K_D values

224   (Supplementary Fig. 5), while the fourth hit, XS381186, displayed 29 µM K_D (Supplementary

225   Fig. 5). To elucidate the binding mode of racemic XS838489, we co-crystallized it with the

226   WDR domain of WDR91 (residues 392-747), revealing the (R)-enantiomer bound to WDR91

227   (Fig. 3f) (Supplementary Table 3; Supplementary Fig. 9b). This binding pocket lies between two

228 β-propellers and is surrounded by hydrophobic residues, including L465, L467, L477, and A459.

229 The amide nitrogen linked to the chiral carbon forms a hydrogen bond with the backbone oxygen

230 of T547, favoring the (R)-enantiomer for binding. The co-crystal structure of WDR91 with

231 XS381295 showed it binding to the same side pocket also in an enantioselective manner

232 (Supplementary Fig. 10), even though this was not apparent by EAS-MS. In the case of

233 XS381295, its chlorophenyl ring fits into a hydrophobic pocket formed by the aliphatic side

234 chains of L477, L465, L467, A459, as well as T547 and M550 (Supplementary Fig. 10). We

235 suspect that the inability to detect enantioselective binding for XS381295 by chiral

236 chromatography may be due to its rapid racemization.

237

238 *WD40 repeat containing protein (WDR55)* is a 383-residue WDR protein that adopts a seven-

239 bladed β-propeller fold and functions as a nucleolar protein involved in ribosomal biogenesis[28].

240 Although WDR55 is predicted to be a challenging target according to its DLID score, racemic

241 'compound XS381774' was enriched (Fig. 3g, 3h), with a subsequently confirmed apparent K_D

242 of 11 μM, as measured by SPR, and displayed enantioselectivity (Supplementary Fig. 6). The

243 enantioselective binding of XS381774 was of some interest, because the enantiomers differed by

244 the orientation of a single methyl group. The mechanism of enantioselective binding was

245 revealed by co-crystallization of racemic XS381774 with WDR55 (Supplementary Table 3;

246 Supplementary Fig. 9). The (S)-enantiomer of XS381774 preferentially co-crystallized in a

247 hydrophobic side pocket of WDR55, and binding was mediated through hydrophobic

248 interactions with the aliphatic side chains of residues I100, I108, V110, V122, L141, L135, and

249 W153 (Fig. 3i). A methyl group which forms the chiral center points toward W153 and I108,

250 enhancing hydrophobic interactions and that explained the enantioselective enrichment. To

251 quantify the degree of selectivity, we separated and purified the individual enantiomers and

11

252   evaluated their binding affinities to WDR55 using SPR. The (*S*)-enantiomer of XS381774

253   exhibited at least 5-fold higher binding affinity compared to the (*R*)-enantiomer, with K_D values

254   of 5 μM and > 25 μM, respectively (Supplementary Fig. 6).

255

256   *Histone Acetyltransferase 1 (HAT1)* is an enzyme that acetylates histone lysine residues, a

257   modification associated with a transcriptionally active chromatin state[30]. Despite two decades of

258   efforts in both academia and industry, there are few, if any, small molecule ligands for HAT1. In

259   our EAS-MS screen, XS380871 was selectively enriched by HAT1 (Fig. 3j). Interestingly,

260   XS380871 contains two chiral centers, and two of its four stereoisomers, were selectively

261   enriched by HAT1 (Fig. 3k). The binding of XS380871 to HAT1, was confirmed by SPR (K_D =

262   12 μM) as well as ^{19}F NMR (Supplementary Fig. 7). To determine the molecular basis of binding,

263   we incubated and co-crystallized the racemic mixture of XS380871 with HAT1 (Supplementary

264   Table 3; Supplementary Fig. 9). The co-structure revealed that the compound binds at the acetyl-

265   Coenzyme A (acetyl-CoA) binding site (Fig. 3l). The chiral carbon (*S*) linked to the nitrophenyl

266   moiety exhibited an enantioselective binding mode, with the nitrophenol substituent forming

267   hydrogen bonds with the backbone nitrogen of G253 (via its nitro group) and the backbone

268   nitrogen of G249 (via its hydroxyl group). There would however be enough space elsewhere in

269   the structure to accommodate the (*R*) enantiomer, albeit with potentially reduced affinity. In

270   contrast, the fluorophenyl moiety at the second chiral carbon, bound to a hydrophobic subpocket

271   lined by residues A275, V238, M241 and Y282; here the other enantiomer would be

272   incompatible with binding due to severe steric clashes with the region around M241. This

273   observation is consistent with the enantioselectivity data obtained from the EAS-MS analysis.

274

275   **DISCUSSION**

276 Despite the extensive efforts over the past decades[6], approximately 80% of all human proteins

277 still lack chemical ligands[5,7,8]. The slow pace of chemical ligand discovery is largely due to the

278 intrinsic limtations of the current chemical ligand discovery strategies. Here, we introduce the

279 EAS-MS approach for high-throughput chemical ligand discovery: 1) The high sensitivity of the

280 EAS-MS method in capturing weak-affinity hits ensures a high success rate, even with a

281 relatively small chemical library (~8,000 compounds). This will reduce the time and financial

282 cost for scalable screening. 2) Hit compounds may be able to be rapidly confirmed using

283 orthogonal enantioselectivity information, bypassing time-consuming biophysical validations

284 required for scalable and early-stage screening. The concept is that EAS-MS provides increased

285 evidence of the specificity of compound binding not through use of an orthogonal assay with an

286 orthogonal compound. Arguably, if a compound binds in an enantioselective manner in the

287 screen at ~100 nM concentration, it may provide sufficient evidence of binding to obviate the

288 need to test the compound at higher concentrations in an orthogonal biophysical assay. We

289 envision that the EAS-MS approach will provide a convenient solution for scalable, proteome-

290 wide chemical ligand discovery.

291 EAS-MS method may also be able to provide semi-quantitative information for identified hits.

292 As shown in Supplementary Fig. 11, strong correlations were observed between the recovered

293 MS signals of confirmed hits and their K_D values. This raises the possibility that the MS signals

294 of EAS-MS hits can be used to generate large high-quality datasets that support artificial

295 intelligence (AI)-based drug discovery.

296 Not all EAS-MS candidate binders could be confirmed using biophysical methods, such as SPR.

297 Our limited data indicates this may be due to differences in their underlying biophysical

298 principles: the SPR method can only detect hit compounds at concentrations above their K_D,

299 which requires compounds to be soluble at high concentrations whereas EAS-MS is insensitive

13

300 to chemical concentrations. This discrepancy most apparent for weak EAS-MS hits, which may

301 fail validation by SPR due to their low water solubility. Moving forward, we propose a general

302 strategy of obtaining analogues of EAS-MS hits with greater predicted solubility as a strategy to

303 facilitate biophysical validation.

304 **REFERENCES**

305 1   Moffat, J. G., Rudolph, J. & Bailey, D. Phenotypic screening in cancer drug discovery -
306   past, present and future. *Nature reviews. Drug discovery* **13**, 588-602 (2014).
307   https://doi.org:10.1038/nrd4366
308 2   Eder, J., Sedrani, R. & Wiesmann, C. The discovery of first-in-class drugs: origins and
309   evolution. *Nature reviews. Drug discovery* **13**, 577-587 (2014).
310   https://doi.org:10.1038/nrd4336
311 3   Macarron, R. *et al.* Impact of high-throughput screening in biomedical research. *Nature*
312   *reviews. Drug discovery* **10**, 188-195 (2011). https://doi.org:10.1038/nrd3368
313 4   Parker, C. G. *et al.* Ligand and target discovery by fragment-based screening in human
314   cells. *Cell* **168**, 527-541 e529 (2017). https://doi.org:10.1016/j.cell.2016.12.029
315 5   Abbasov, M. E. *et al.* A proteome-wide atlas of lysine-reactive chemistry. *Nature*
316   *chemistry* **13**, 1081-1092 (2021). https://doi.org:10.1038/s41557-021-00765-4
317 6   Offensperger, F. *et al.* Large-scale chemoproteomics expedites ligand discovery and
318   predicts ligand behavior in cells. *Science* **384**, eadk5864 (2024).
319   https://doi.org:10.1126/science.adk5864
320 7   Prudent, R., Annis, D. A., Dandliker, P. J., Ortholand, J. Y. & Roche, D. Exploring new
321   targets and chemical space with affinity selection-mass spectrometry. *Nature reviews.*
322   *Chemistry* **5**, 62-71 (2021). https://doi.org:10.1038/s41570-020-00229-2
323 8   Durant, G., Boyles, F., Birchall, K. & Deane, C. M. The future of machine learning for
324   small-molecule drug discovery will be driven by data. *Nature computational science* **4**,
325   735-743 (2024). https://doi.org:10.1038/s43588-024-00699-0
326 9   Wognum, C. *et al.* A call for an industry-led initiative to critically assess machine
327   learning for real-world drug discovery. *Nature Machine Intelligence* **6**, 1120-1121
328   (2024).
329 10   Wang, Y. *et al.* Expedited mapping of the ligandable proteome using fully functionalized
330   enantiomeric probe pairs. *Nature chemistry* **11**, 1113-1123 (2019).
331   https://doi.org:10.1038/s41557-019-0351-5
332 11   Favalli, N. *et al.* Stereo- and regiodefined DNA-encoded chemical libraries enable
333   efficient tumour-targeting applications. *Nature chemistry* **13**, 540-548 (2021).
334   https://doi.org:10.1038/s41557-021-00660-y
335 12   Tao, Y. *et al.* Targeted protein degradation by electrophilic PROTACs that
336   stereoselectively and site-specifically engage DCAF1. *Journal of the American Chemical*
337   *Society* **144**, 18688-18699 (2022). https://doi.org:10.1021/jacs.2c08964
338 13   Ogasawara, D. *et al.* Chemical tools to expand the ligandable proteome: diversity-
339   oriented synthesis-based photoreactive stereoprobes. *bioRxiv : the preprint server for*
340   *biology* (2024). https://doi.org:10.1101/2024.02.27.582206

14   Yang, D. *et al.* Widespread formation of toxic nitrated bisphenols indoors by heterogeneous reactions with HONO. *Science advances* **8**, eabq7023 (2022). https://doi.org/10.1126/sciadv.abq7023

15   Liu, J. *et al.* Diindoles produced from commensal microbiota metabolites function as endogenous CAR/Nr1i3 ligands. *Nat Commun* **15**, 2563 (2024). https://doi.org/10.1038/s41467-024-46559-3

16   Liu, J. *et al.* The omega-3 hydroxy fatty acid 7(S)-HDHA is a high-affinity PPARα ligand that regulates brain neuronal morphology. *Science signaling* **15**, eabo1857 (2022). https://doi.org/10.1126/scisignal.abo1857

17   Ito, T. *et al.* Identification of a primary target of thalidomide teratogenicity. *Science* **327**, 1345-1350 (2010). https://doi.org/10.1126/science.1177319

18   Mori, T. *et al.* Structural basis of thalidomide enantiomer binding to cereblon. *Sci Rep* **8**, 1294 (2018). https://doi.org/10.1038/s41598-018-19202-7

19   Chamberlain, P. P. *et al.* Structure of the human Cereblon-DDB1-lenalidomide complex reveals basis for responsiveness to thalidomide analogs. *Nature structural & molecular biology* **21**, 803-809 (2014). https://doi.org/10.1038/nsmb.2874

20   Li, A. S. M. *et al.* Discovery of nanomolar DCAF1 small molecule ligands. *J Med Chem* **66**, 5041-5060 (2023). https://doi.org/10.1021/acs.jmedchem.2c02132

21   Bonday, Z. Q. *et al.* LLY-283, a potent and selective inhibitor of arginine methyltransferase 5, PRMT5, with antitumor activity. *ACS medicinal chemistry letters* **9**, 612-617 (2018). https://doi.org/10.1021/acsmedchemlett.8b00014

22   Göricke, F. *et al.* Discovery and characterization of BAY-805, a potent and selective inhibitor of ubiquitin-specific protease USP21. *J Med Chem* **66**, 3431-3447 (2023). https://doi.org/10.1021/acs.jmedchem.2c01933

23   Sheridan, R. P., Maiorov, V. N., Holloway, M. K., Cornell, W. D. & Gao, Y. D. Drug-like density: a method of quantifying the "bindability" of a protein target based on a very large set of pockets and drug-like ligands from the Protein Data Bank. *J Chem Inf Model* **50**, 2029-2040 (2010). https://doi.org/10.1021/ci100312t

24   Ahmad, S. *et al.* Discovery of a first-in-class small-molecule ligand for WDR91 using DNA-encoded chemical library selection followed by machine learning. *J Med Chem* **66**, 16051-16061 (2023). https://doi.org/10.1021/acs.jmedchem.3c01471

25   Ackloo, S. *et al.* A target class ligandability evaluation of WD40 repeat-containing proteins. *J Med Chem* (2024). https://doi.org/10.1021/acs.jmedchem.4c02010

26   Helmerhorst, E., Chandler, D. J., Nussio, M. & Mamotte, C. D. Real-time and label-free bio-sensing of molecular interactions by surface plasmon resonance: A laboratory medicine perspective. *The Clinical biochemist. Reviews* **33**, 161-173 (2012).

27   Jackson, S. & Xiong, Y. CRL4s: the CUL4-RING E3 ubiquitin ligases. *Trends in biochemical sciences* **34**, 562-570 (2009). https://doi.org/10.1016/j.tibs.2009.07.002

28   Iwanami, N. *et al.* WDR55 is a nucleolar modulator of ribosomal RNA synthesis, cell cycle progression, and teleost organ development. *PLoS genetics* **4**, e1000171 (2008). https://doi.org/10.1371/journal.pgen.1000171

29   Casanova, J. E. & Winckler, B. A new Rab7 effector controls phosphoinositide conversion in endosome maturation. *The Journal of cell biology* **216**, 2995-2997 (2017). https://doi.org/10.1083/jcb.201709034

30   Nagarajan, P. *et al.* Histone acetyl transferase 1 is essential for mammalian development, genome stability, and the processing of newly synthesized histones H3 and H4. *PLoS genetics* **9**, e1003518 (2013). https://doi.org/10.1371/journal.pgen.1003518

388   31   Hong, Y., Welch, C. J., Piras, P. & Tang, H. Enhanced structure-based prediction of
389   chiral stationary phases for chromatographic enantioseparation from 3D molecular
390   conformations. *Anal Chem* **96**, 2351-2359 (2024).
391   https://doi.org:10.1021/acs.analchem.3c04028
392   32   Xu, H., Lin, J., Zhang, D. & Mo, F. Retention time prediction for chromatographic
393   enantioseparation by quantile geometry-enhanced graph neural network. *Nat Commun* **14**,
394   3095 (2023). https://doi.org:10.1038/s41467-023-38853-3
395

396 **METHODS**

397 **Reagents**

398 Methanol (HPLC grade), ultrapure water (HPLC grade), acetonitrile (HPLC grade), DMSO,

399 glycerol, and formic acid were purchased from Fisher Scientific (Ottawa, ON, CA). Ni-NTA

400 magnetic agarose beads and Triton X-100 were obtained from Sigma-Aldrich (St. Louis, MO,

401 USA). Tris base, sodium chloride (NaCl), Tris(2-Carboxyethyl)phosphine (TCEP) were sourced

402 from BioShop Canada Inc. (Burlington, ON, CA). Imidazole was provided by Bio Basic Canada

403 Inc. (Markham, ON, CA). The chemical library was purchased from Enamine US Inc.

404 (Monmouth Jct., NJ, USA) and ChemDiv (San Diego, CA, USA). Stock solutions of the

405 chemicals was prepared in DMSO and stored at -20 $^{\circ}$C in the dark until use.

406

407 **EAS-MS protein affinity selection experiments**

408 The EAS-MS library of 8,210 chemicals was divided into 14 chemical pools by using a robotic

409 system, with an average of ~600 chemicals in each pool. Each of the 14 chemical pools was

410 incubated with purified recombinant His-tagged human proteins on 96-well plates, in the binding

411 buffer (150 mM NaCl, 50 mM Tris-HCl, 0.1 mM TCEP, 0.5% glycerol, 0.01% Triton X-100,

412 and pH 7.5). 5 µL of Ni-NTA magnetic beads was added to the binding buffer, adjusted to a final

413 volume of 250 µL per sample. The final concentrations were set to 1 µM for the proteins and 100

414 nM for each chemical. All EAS-MS screenings were performed in triplicate. Incubation was

415 carried out on a rotary mixer at 4 $^{\circ}$C for 30 min. After incubation, the 96-well plate was placed

416 on a magnetic plate to separate the beads, and the incubation buffer was removed. The beads

417 were then washed twice with 100 µL of washing buffer 1 (150 mM NaCl, 50 mM Tris-HCl, 0.1

418 mM TCEP, 0.5% glycerol, 0.01% Triton X-100, 5 mM imidazole, and pH 7.5), followed by one

419 wash with 100 µL of washing buffer 2 (150 mM NaCl, 50 mM Tris-HCl, 5 mM imidazole, and

17

420   pH 7.5). The beads and washing buffer 2 were transferred to a new plate. After removing the

421   washing buffer 2, 120 µL of methanol was added to each well to denature proteins and release

422   chemical ligands. The methanol extracts separated from magnetic beads were directly subjected

423   to LC-MS analysis. In parallel, the extracts were also subjected to SDS-PAGE to confirm the

424   presence of target proteins on the beads.

425

426   **Identification of putative hits using LC-MS**

427   The EAS-MS extracts were analyzed using an Orbitrap Exploris 240 mass spectrometer

428   equipped with a Vanquish UHPLC system (Thermo Fisher Scientific, CA, USA).

429   Chromatographic separation was conducted on an Accucore Vanquish C_{18} reverse-phase column

430   (50 mm × 2.1 mm × 3 µm) at a flow rate of 0.3 mL/min. The injection volume was 1 µL.

431   Ultrapure water with 0.1% formic acid (A) and methanol with 0.1% formic acid (B) were used as

432   the two mobile phases. Gradient elution started with 5% B, increased to 80% at 1.5 min, then

433   reached to 100% at 3 min, held for 1.1 min, and finally returned to 5% B over 1.5 min. The

434   column temperature was maintained at 40 $^{\circ}$C, and the sample compartment at 7 $^{\circ}$C. Data

435   acquisition was performed in full MS^1 scan mode (150-520 *m/z*) with a resolution of R = 60,000,

436   in both positive and negative ionization modes. The detailed instrumental parameters of LC-MS

437   system are provided in Table S4.

438

439   **Chiral analysis of putative hits**

440   *1) Building the chiral separation database:* the accurate prediction of suitable chiral separation

441   conditions for resolving the enantiomers of a given compound remains a challenge[31,32]. To

442   address this, we decided to establish the chiral separation database for the EAS-MS chemical

18

443 library, under eight conditions (4 columns × 2 LC gradient methods, see the details below). Each

444 of the 14 chemical pools were individually injected under eight conditions to build the database.

445 *2) Eight chiral separation conditions:* four columns from DAICEL Chemical Industries, LTD.

446 were employed: CHIRALPAK IA (250 × 4.6 mm), CHIRALPAK IBN (250 × 2.1 mm),

447 CHIRALPAK IC (250 × 2.1 mm), and CHIRALPAK IG (250 × 2.1 mm). These columns were

448 selected because they have been demonstrated to achieve sufficient separations for the biggest

449 number of chiral compounds[31].

450 Two LC gradient methods were employed for each column. In **method 1**, acetonitrile with 0.1%

451 formic acid (A) and methanol with 0.1% formic acid (B) were used as mobile phases. Gradient

452 elution started at 40% B, held for 1 min, then increased to 80% at 3 min, reached 100% at 8 min,

453 held for 1 min, followed by a return to 40% over 2 min, and held for another 1 min. The flow

454 rate was set at 0.3 mL/min for columns IBN, IC, and IG, and at 1.0 mL/min for column IA.

455 In **method 2**, H_2O with 0.1% formic acid (A) and methanol with 0.1% formic acid (B) were used

456 as mobile phases. Gradient elution began with 40% B, increased to 80% at 6 min, reached 100%

457 at 18 min, held for 6 min, and then return to 40% over 6 min. The flow rate was 0.2 mL/min for

458 columns IBN, IC, and IG, and 1.0 mL/min for column IA. The column temperature was

459 maintained at 40 ^{o}C, and the sample compartment was kept at 7 ^{o}C. The injection volume was 1

460 μL.

461 *3) Chiral detection of putative hits under optimal conditions:* Once putative hits were detected

462 from initial EAS-MS screening, the optimal separation condition for the hit compounds was

463 selected by searching against the chiral separation database established in step 2) above. The

464 same EAS-MS extracts were then analyzed under the optimal conditions. To exclude potential

465 interferences, enantiomers were further confirmed by matching MS^2 spectra.

466

19

467 **Automatic data processing**

468 A total of 1,302 EAS-MS screening samples (31 proteins × 3 replicates × 14 pools = 1,302) were

469 completed in the current study, leading to 145 GB of raw mass spectrometry data. We have

470 established an automatic data processing workflow to detect EAS-MS hits (code is available on

471 https://github.com/huiUofT/eCPIN).

472 *1) Building the MS database for EAS-MS library*: we first injected each of the 14 chemical pools

473 to LC-MS analysis to build the database. In brief, raw mass spectrometry files were converted to

474 the mzXML format. Peak features from each chemical pool were detected with the 'XCMS' R

475 package at a mass tolerance of 2.5 ppm[29]. Only the peak features with peak intensity $> 10^5$, and

476 10 times higher abundances when compared to methanol were considered as library compounds

477 and retained for subsequent data analysis. Isotopic peaks and adducts were excluded by matching

478 chromatographic peaks and theoretical mass difference. The detected peak features were then

479 matched to the EAS-MS library with a mass tolerance of 3 ppm. Potential inferences and

480 misassigned compounds were further excluded via manual inspections. Then, a MS database of

481 8,210 EAS-MS library compounds were estalibshed, with *m/z*, retention time and SMILES

482 information recorded.

483 *2) Detecting compounds from EAS-MS features by matching to the MS database*: to detect library

484 compounds from the EAS-MS features, we matched each of the 8,210 EAS-MS compounds

485 against each EAS-MS samples. A mass tolerance of 3 ppm, and a retention time matching

486 window of 0.5 minutes were used. A final data matrix with 8,210 rows (corresponding to library

487 compounds) and 24 columns (corresponding to EAS-MS samples from each batch) was created.

488 *3) Detecting putative EAS-MS hits*: to detect putative EAS-MS hits, we calculated the enrichment

489 fold and *p* values of each library compound against each protein by using seven other proteins

490 from the same batch as the control (equation 1).

20

491 **Enrichment Fold = $A_{POI} \times N / \sum A_i$** (1)

492 A_{POI} represents the peak intensity of the putative hit compound enriched by the protein of interest;

493 N represents the number of background proteins from the same batch (N = 7); A_i represents the

494 peak intensity of the same hit compound enriched by the ith protein from the same batch.

495 Eventually, only the compounds with enrichment fold > 5, and *p* value < 0.05 were considered as

496 putative hit compounds. Each putative hit was further manually inspected by matching to

497 compound library, to exclude potential inteferences misassigned by the algorithms.

498

Quality Assurance and Quality Control (QA/QC) of EAS-MS

500 To assure the high-quality data for scalable EAS-MS screening, we conducted stringent QA/QC

501 from both protein and chemical perspectives.

502

Protein QA/QC

504 To assure the selected proteins were properly folded and compatible with our EAS-MS workflow,

505 we used SDS-PAGE to semi-quantitatively assess the recovery of each protein through our EAS-

506 MS procedure. In brief, SDS-PAGE analysis was conducted for each protein before and after

507 EAS-MS. The band intensity of each protein was compared, and those proteins with low

508 recoveries were excluded for EAS-MS analysis. The SDS-PAGE results for all proteins are

509 provided in Supplementary Fig. 12.

510

Chemical QA/QC

512 During LC-MS analysis, the EAS-MS chiral library compounds were injected every batch to

513 ensure the retention time, *m/z*, and intensity are not shifting for database matching. All the chiral

514 compounds were detected by the LC-MS system. A QC standard of small subset (~600

21

515   chemicals) was injected after every 12 samples to monitor instrument stability and LC (chiral)

516   separations. The instrument response deviation for analytes in the QC standard remained below

517   10% throughout the analysis. A solvent blank was injected after every 12 samples to monitor and

518   prevent carryover.

519

520   **Proteins used for EAS-MS and SPR**

521   Protein expression and purification summaries are detailed in Supplementary Table 5.

522

523   **Crystallization and structure determination**

524   Crystallization of EAS-MS hits with their respective targets are summarized in Supplementary

525   Table 6.

526

527   **Surface Plasmon Resonance**

528   SPR hit confirmation and validations were performed using a Biacore 8K instrument at 20 °C.

529   The biotinylated respective protein constructs were immobilized on the active flow cells of a

530   Streptavidin-coated SA sensor chip after initial conditioning of both reference and active flow

531   cells with 50 mM NaOH for 3×60 s with a flow rate of 10 μL/min. Each protein solution (30-

532   150 μg/mL) was injected through the respective active flow cells for 60-660 s with a flow rate of

533   5 μL/min to obtain a desired protein immobilization level. Equilibration was performed after

534   protein immobilization by flowing running buffer (50 mM HEPES, pH 7.5, 150 mM NaCl,

535   0.001% Tween 20, 0.2% PEG3350, 0.5 mM TCEP, and 3% DMSO) over the flow cells with a

536   flow rate of 50 μL/min until a stable baseline was observed. Initial start-up cycles, blank cycles

537   and wash (50% DMSO wash to flush the needles) steps were included in the SPR compound

538   analysis methods. The compound injections were performed over the reference and active flow

22

539   cells using multi-cycle kinetics at a flow rate of 40 μL/min with a 55 s association time and a 120

540   s dissociation time for dose-response titration. The stock compound concentration series was

541   performed in 100% DMSO with 2-3-fold serial dilution and prepared the samples using the

542   running buffer by maintaining final DMSO concentration of 3% (v/v) across the tested

543   concentration range. Initially, the compounds were titrated between 0-45 μM compound

544   concentrations and later followed up between 0-15 μM, 0-30 μM, 0-45 μM, and 0-90 μM

545   compound concentrations based on the affinity and solubility of the compounds. Solvent

546   correction cycles were also included across each run to adjust high bulk responses from the

547   solvent. Double referencing of the data was introduced by subtraction of the reference flow cell

548   and the respective zero compound concentration cycles. Affinity fitting was performed by

549   applying a 1:1 equilibrium binding model to the data using Biacore Intelligent analysis tool

550   provided with the Biacore Insight Evaluation software. The final figures were prepared using

551   GraphPad Prism.

552

553   **^{19}F NMR studies**

554   The binding of XS380871 to HAT1 was confirmed using ^{19}F NMR by looking for the

555   broadening and/or shifting of its ^{19}F resonance upon the addition of HAT1. Spectra of the

556   compound were acquired on a Bruker Avance III spectrometer operating at 600 MHz, equipped

557   with a QCI probe at 293K, and collected at 20 μM, with and without the presence of HAT1 (~34

558   μM, buffered in 50 mM Tris 7.0, 200 mM NaCl, 4 mM DTT). TFA (100 μM) was added as an

559   internal reference. 1k transients were acquired over a sweep width of 150 ppm; an exponential

560   window function (LB = 5 Hz) was applied prior to Fourier transformation.

561

562   **Statistics**

23

563   All statistical analyses were conducted using GraphPad PRISM (GraphPad Software, Inc., La

564   Jolla, CA). Data are presented as mean ± standard deviation. Statistical significance was

565   determined using Student's *t*-test, with a significance threshold of $p < 0.05$.

566

567   **DATA AVAILABILITY**

568   Atomic coordinates and structure factors for all crystal structures have been deposited in the

569   Protein Data bank under the accession codes: 9EJO, 9EJP, 9EJQ, 9EKP, 9MJG.

570

571   **ACKNOWLEDGEMENTS**

587   Winship Cancer Institute. AME holds the Temerty Nexus Chair of Health Technology and

588   Innovation at the University of Toronto.

589

590   **AUTHOR CONTRIBUTIONS**

591   XW, JS, DY, AB, SA, PJB, VS, LH, HP, AME: EAS-MS method development

592   JS, XW: EAS-MS screening

593   HZ, PG, EG, MK, LKR, DM, CL: Protein production and QC

594   ShA, MK, FL: SPR, DSF confirmation

595   SH: F-NMR characterization

596   HZ, UHC, AD: X-ray crystallography

597   MS, CVS: Druggability assessment

598   SA, VS, PJB: Project management

599   WWY, RH, TSB, SM, ACJ, SK, LH, HP, AME, CHA: Supervision, review and editing

600   XW, JS, LH, HP: Writing, reviewing and editing with input from all

601

602   **COMPETING INTERESTS**

603   The authors declare no competing interests.

604
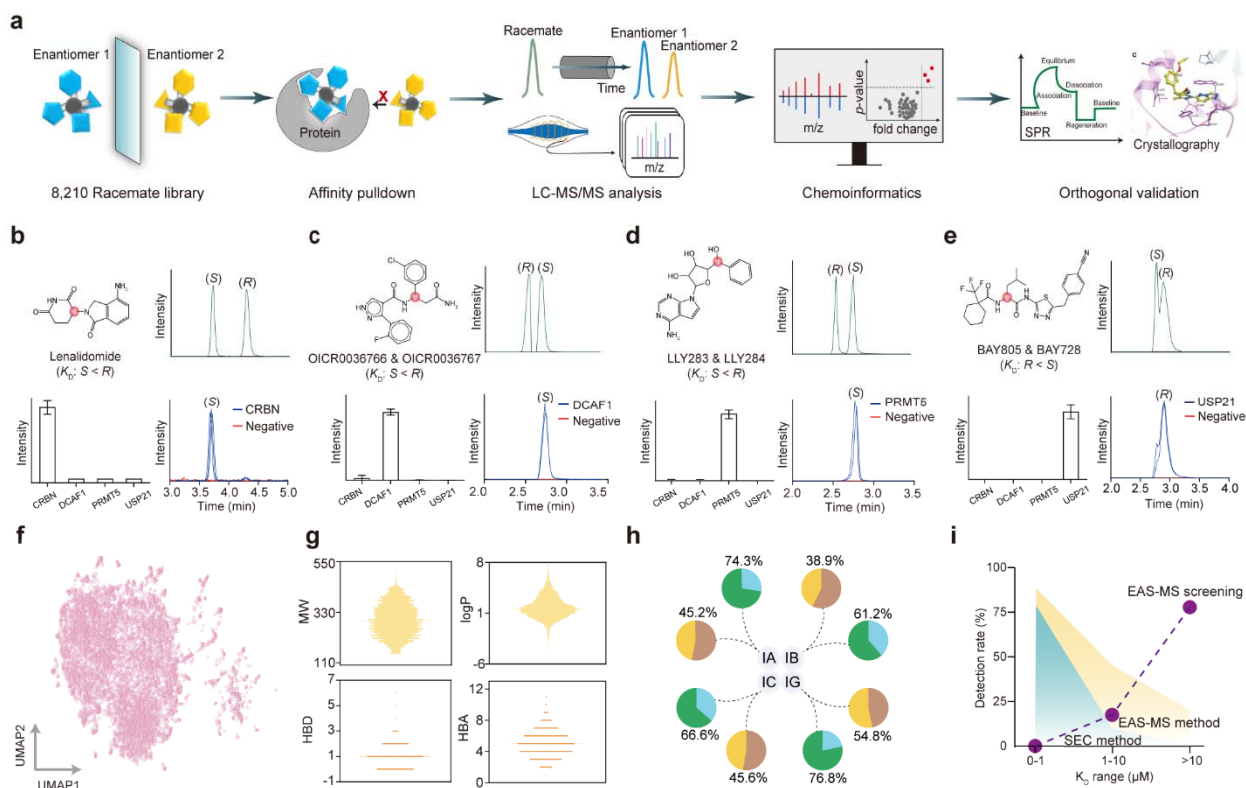
Fig. 1 | Development of the EAS-MS platform for scalable protein ligandability discovery.

(**a**) Schematic representation of the EAS-MS screening workflow. (**b-e**) Benchmarking the EAS-MS platform using four positive control ligands binding to CRBN, DCAF1, PRMT5, and USP21, respectively. (**f**) Chemical space of the EAS-MS library. A two-dimensional uniform manifold approximation and projection (UMAP) visualization of SMILES descriptor. (**g**) Molecular properties of the EAS-MS chemical library: molecular weight (MW), LogP, hydrogen bond acceptors (HBA) and donors (HBD). (**h**) Chiral separation of the EAS-MS library using four chiral columns (IA, IB, IC, and IG), and two mobile phases (acetonitrile/methanol, yellow; water/methanol, green). (**i**) Detection rate of 102 ligands against 8 proteins by SEC and EAS-MS methods, respectively. Purple dots represent the K_D distribution of chemical ligands discovered by our own EAS-MS screening.

26
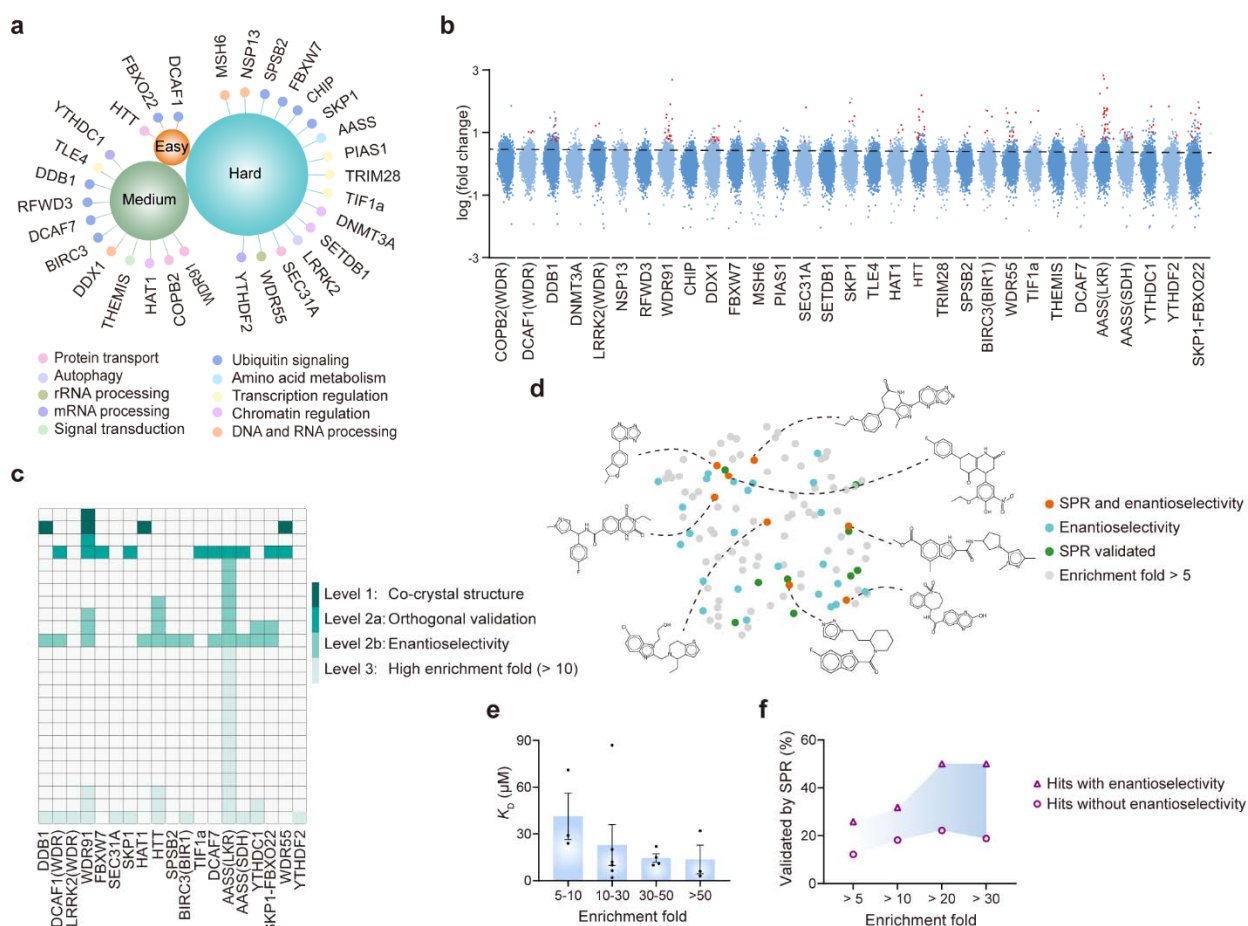
Fig. 2 | Scalable chemical ligand profiling across 31 diverse proteins. (**a**) Overview of the 31 screened proteins. 'High': drug-like density (DLID) > 1, 'Medium': 0.5 < DLID < 1, 'Low': DLID < 0.5. (**b**) Scatter plot of detected compounds interacting with all proteins. Dots above the line represent hits with enrichment fold > 5. Red dots indicate hits with enrichment fold > 5 and *p*-value < 0.05. (**c**) Confidence 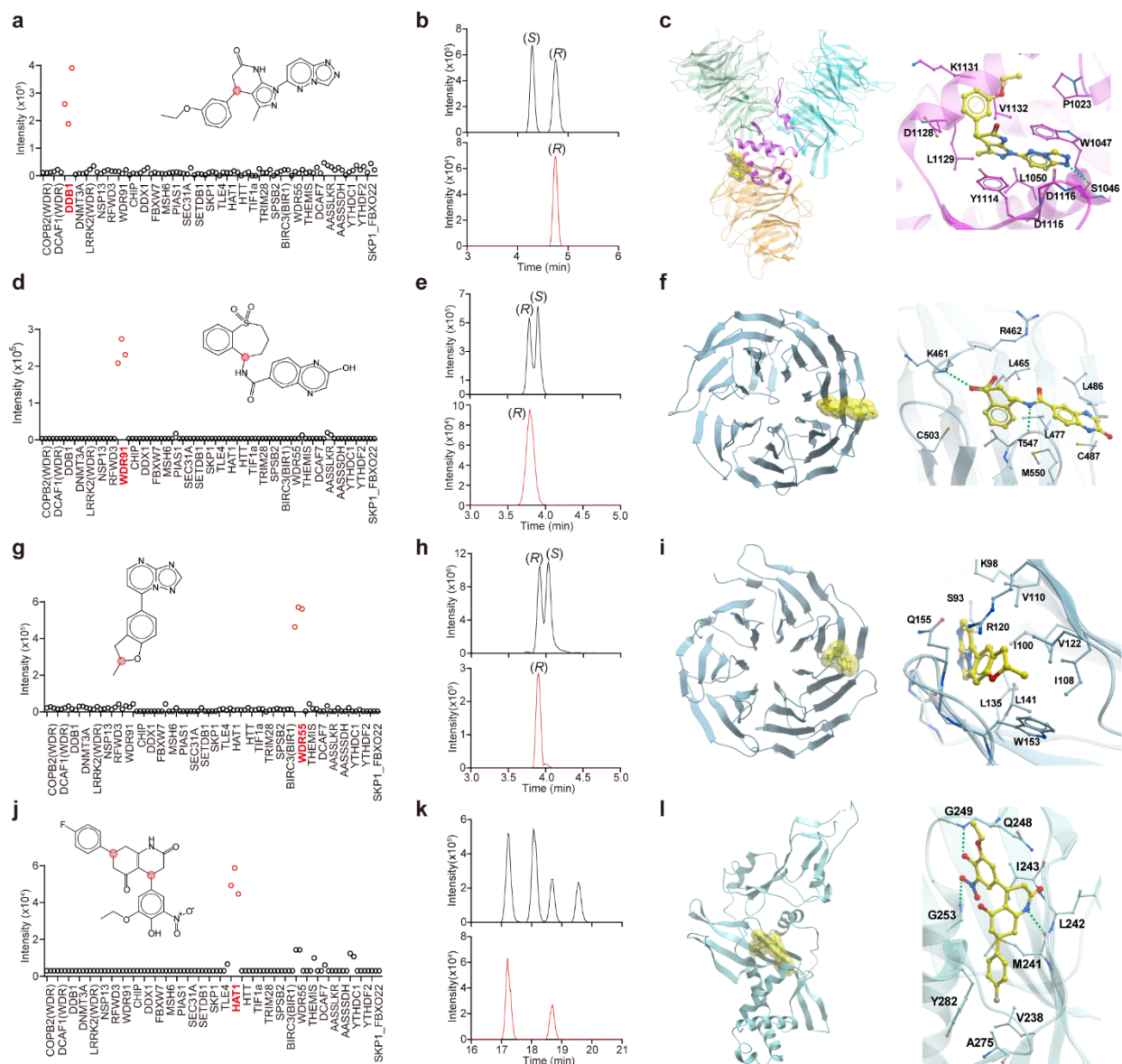levels of all identified hits. (**d**) UMAP visualization of the chemical space for hits with enrichment fold > 5, based on SMILES descriptors. (**e**) Correlation between K_D and fold change. (**f**) SPR validation rate of hits across different enrichment fold ranges.

627

Fig. 3 | Identification of enantioselective hits for DDB1, WDR91, WDR55, and HAT1. (a) Scatter plot showing XS381952 pulled down by DDB1. (**b**) Chiral chromatogram of the XS381952 enantiomers. (**c**) Co-crystal structure of DDB1 with XS381952. Hydrogen bond between the protein and the ligand is shown as green dashed line (**d**) Scatter plot showing XS838489 pulled down by WDR91. (**e**) Chiral chromatogram of the XS838489 enantiomers. (**f**) Co-crystal structure of WDR91 with XS838489. Hydrogen bonds between the protein and the ligand are shown as green dashed lines (**g**) Scatter plot showing XS381774 pulled down by WDR55. (**h**) Chiral chromatogram of the XS381774 enantiomers. (**i**) Co-crystal structure of

28

636   WDR55 with XS381774. (**j**) Scatter plot showing XS380871 pulled down by HAT1. (**k**) Chiral

637   chromatogram of the XS380871 enantiomers. (**l**) Co-crystal structure of HAT1 with XS380871.

638   Hydrogen bonds between the protein and the ligand are shown as green dashed lines.

639
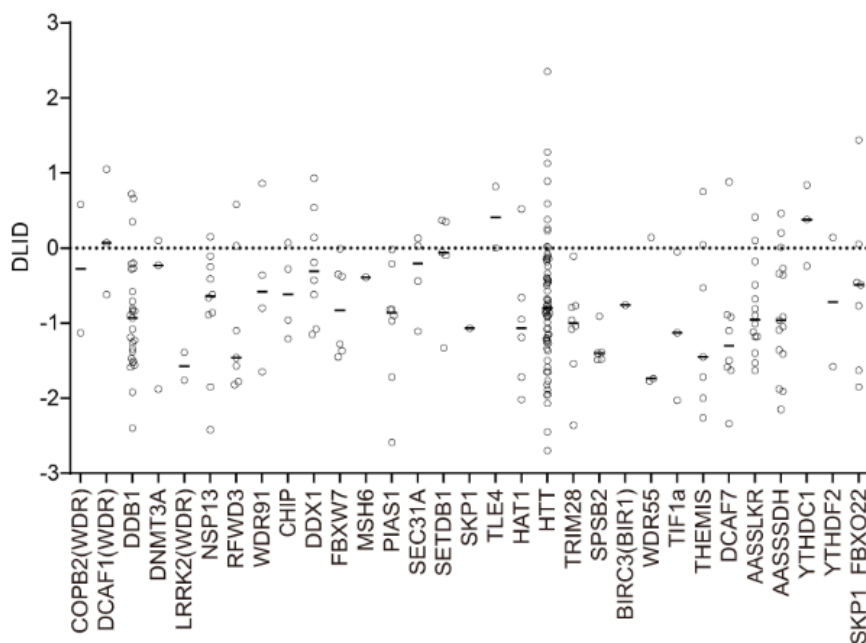
**Supplementary Fig. 1 | Predicted ligandability of 31 selected proteins.** Drug-like density (DLID) of each protein was predicted. The dots from each protein represents the predicted DLID score for each pocket in that protein. To define the ligandability of proteins, we used the maximal DLID score for the pocket from each protein. For instance, the maximal DLID score for HTT is 2.35 for the top drug-like pockets compared to other pockets, due to its large molecular size and surface. Proteins with low, medium and high ligandability were defined, as having maximal DLID scores < 0.5, 0.5-1, and > 1, respectively.
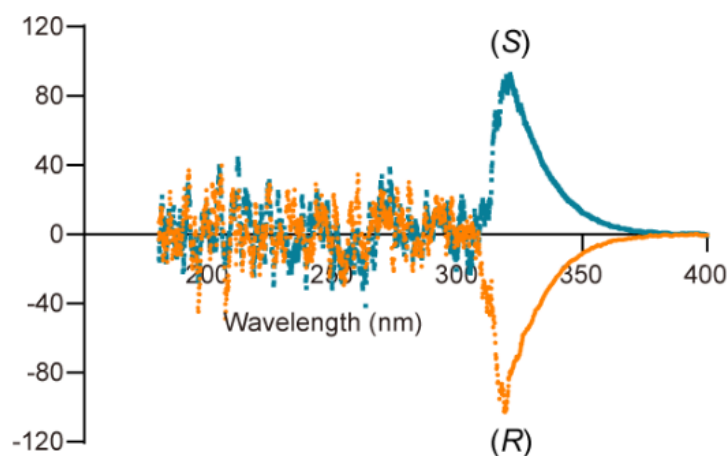
647
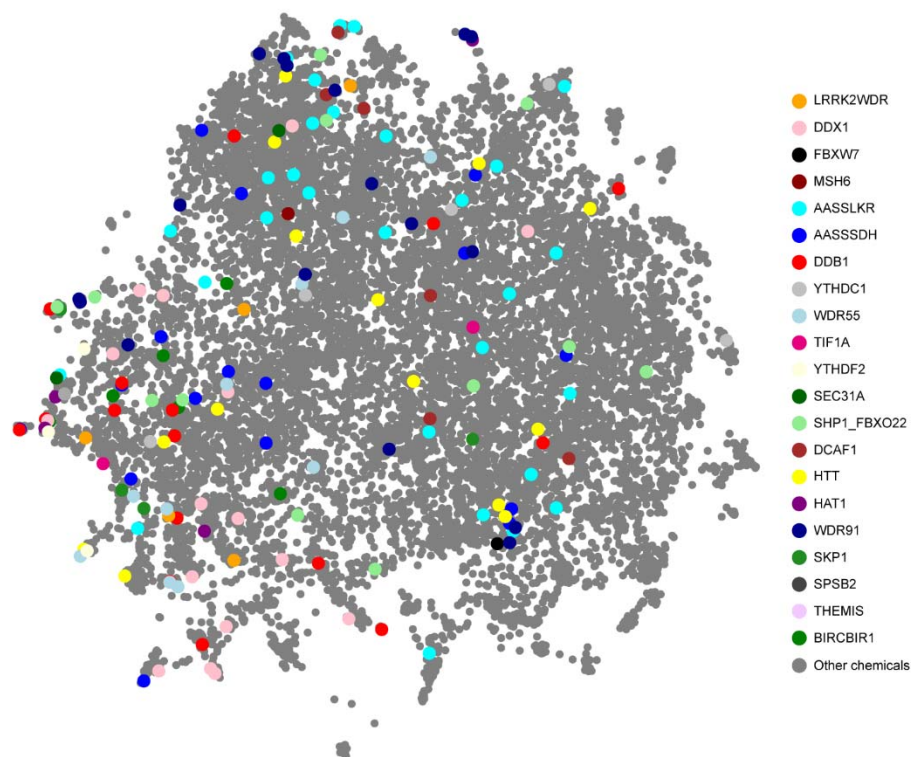
Supplementary Fig. 2 | Distribution of EAS-MS hits across the chemical space. UMAP

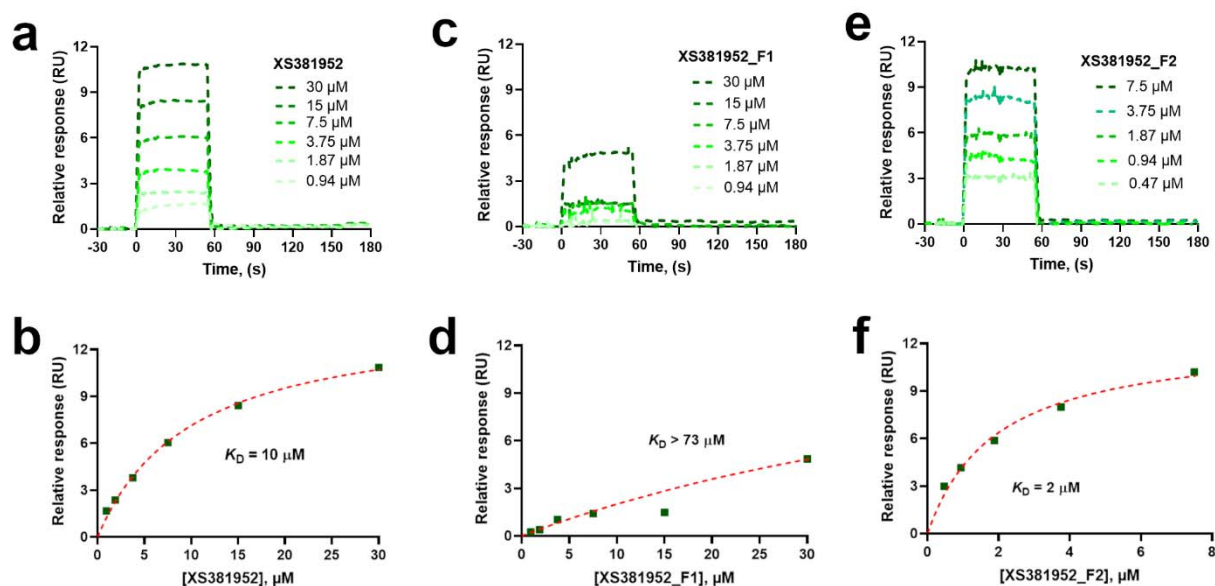visualization based on SMILES descriptor highlights the chemical distribution of identified hits.

650

651

Supplementary Fig. 3 | ECD analysis of two enantiomers of XS381952 binding to DDB1.

The two enantiomers were purified through the CHIRALPAK IA column.

654

Supplementary Fig. 4 | Representative SPR sensorgrams (**a**, **c**, **e**) and the response vs concentration plots (**b**, **d**, **f**) for the DDB1 specific hit XS381952 (**a**, **b**) and its enantiomers XS381952_F1 (**c**, **d**) and XS381952_F2 (**e**, **f**) after separation, fraction 1 (100% pure) and fraction 2 (85% pure), respectively. The response at equilibrium for each concentration were plotted against the compound concentration and affinity fitted by applying 1:1 equilibrium binding model to extract the K_D of the interaction. The compounds were tested in duplicates. Data from one replicate is shown here for clarity.
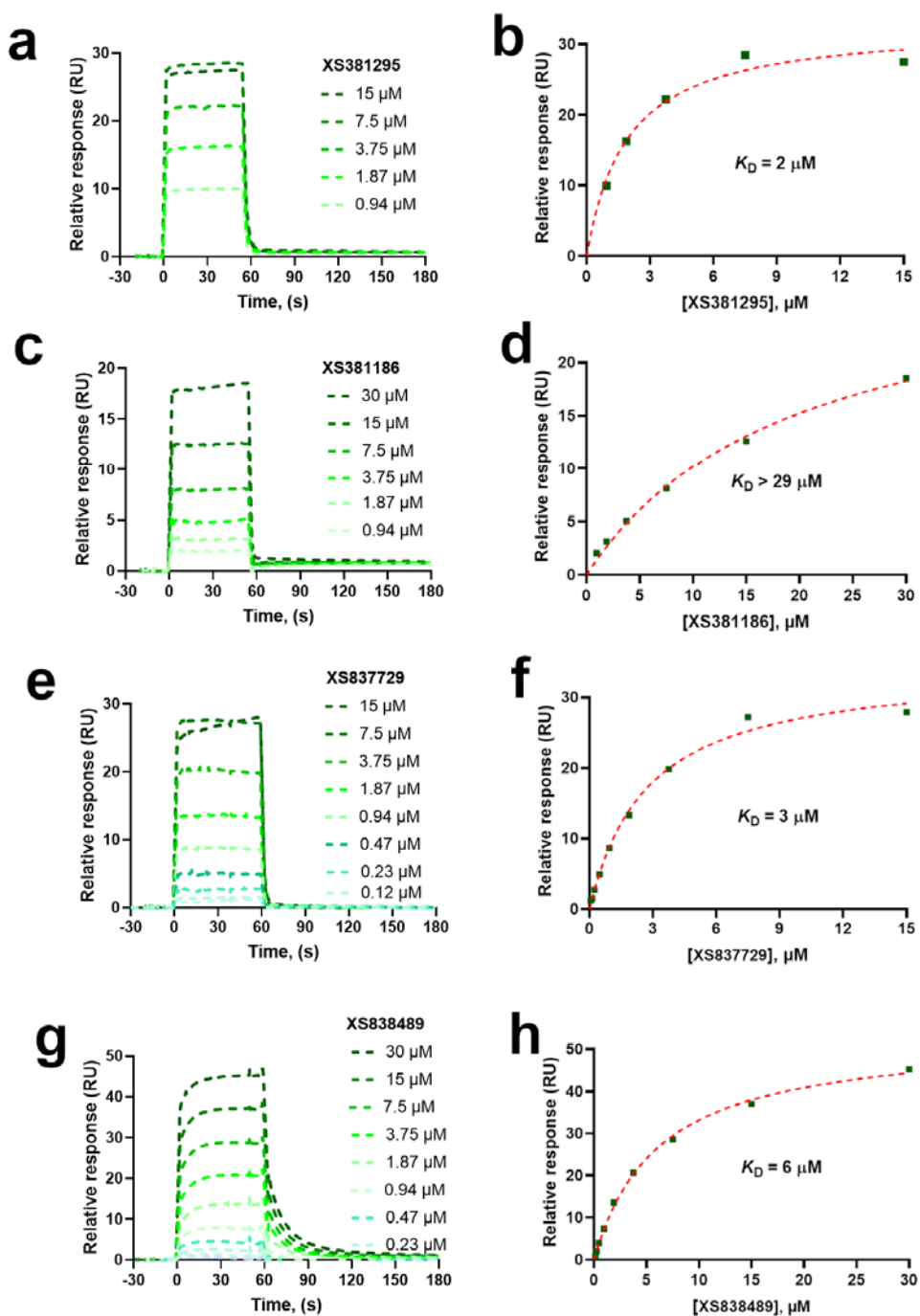
Supplementary Fig. 5 | Representative SPR sensorgrams (**a, c, e,** and **g**) and the response vs concentration plots (**b, d, f**, and **h**) for the four WDR91-specific EAS-MS hits: XS381295a (**a, b**), XS381186a (**c, d**), XS3837729 (**e, f**), and XS838489 (**g, h**). The response at equilibrium for each concentration were plotted against the compound concentration and affinity fitted by

668   applying 1:1 equilibrium binding model to extract the K_D of the interaction. The compounds

669   were tested in duplicates. Data from one replicate is shown here for clarity.
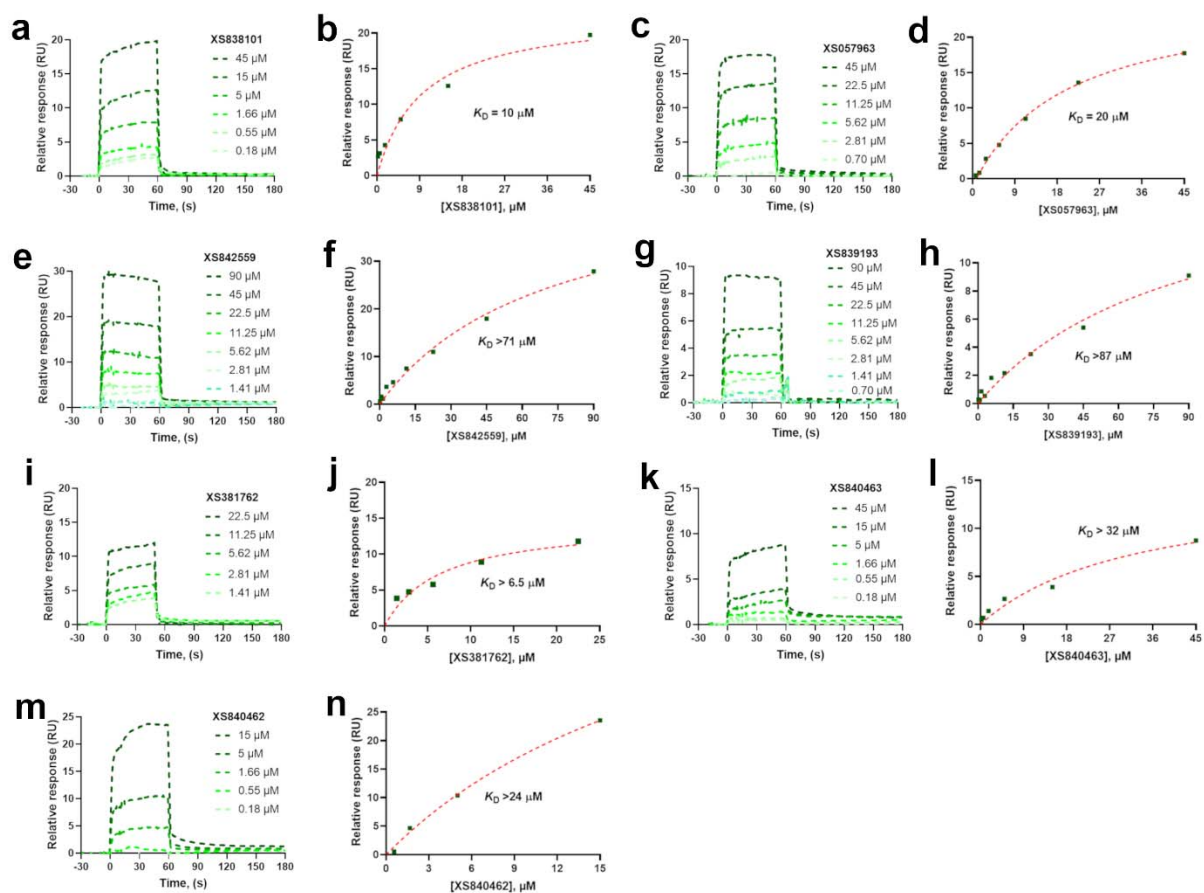
670

671

34
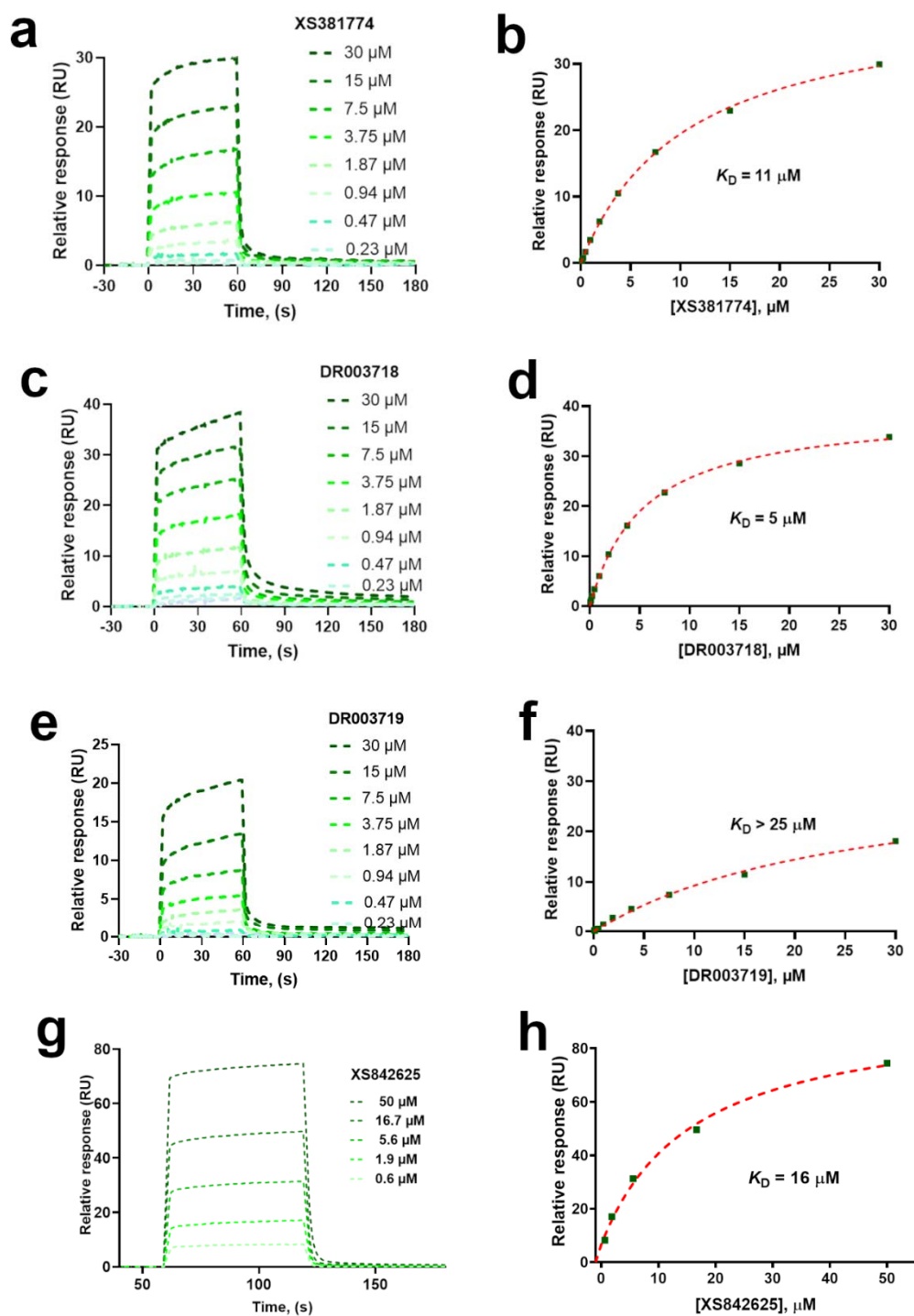
Supplementary Fig. 6 | Representative SPR sensorgrams (**a**, **c**, **e**, and **e**) and the response vs concentration plots (**b**, **d**, **f**, and **h**) for the WDR55 specific hit XS381774 (**a**, **b**) and its separated enantiomers DR003718 (**c**, **d**) and DR003719 (**e**, **f**), as well as XS842625 (**g**, **h**), respectively. The response at equilibrium for each concentration were plotted against the compound concentration and affinity fitted by applying 1:1 equilibrium binding model to calculate the K_D of the interaction. The compounds were tested in duplicates. Data from one replicate is shown here for clarity.
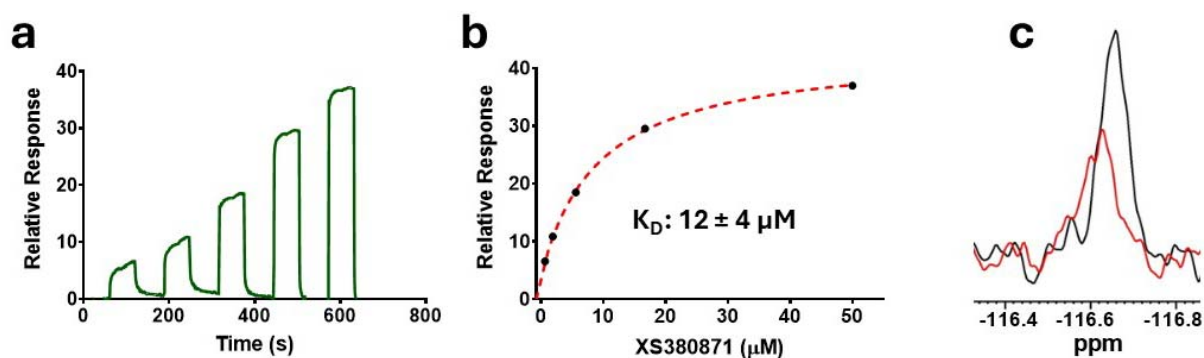

Supplementary Fig. 7 | Representative SPR sensorgram (**a**) and the steady-state response vs concentration plot (**b**) for the HAT1 specific hit XS380871. The response at equilibrium for each concentration were plotted against the compound concentration and affinity fitted by applying 1:1 equilibrium binding model (red dashed line), yielding a K_D value of 12 ± 4 µM. The experiments were performed in triplicate (n = 3). Data from one replicate is shown here for clarity. (**c**) Overlay of ^{19}F spectra of 20 µM XS380871 alone (black), and in the presence of ~34 µM of HAT1 (red).

Supplementary Fig. 8 | Representative SPR sensorgrams (**a**, **c**, **e**, **g**, **i**, **k**, and **m**) and the response vs concentration plots (**b**, **d**, **f**, **h**, **j**, **l**, and **n**) for the SKP1 (**a**, **b**), DCAF1 (**c**, **d**), DCAF7 (**e**, **f**), FBXW7 (**g**, **h**), FBXO22 (**i**, **j**), AASS-LKR-GFP (**k**, **l**), and AASS-SDH-GFP (**m**, **n**) EAS-MS hits. The responses at equilibrium for each concentration were plotted against the compound concentration and affinity fitted by applying 1:1 equilibrium binding model to extract the K_D of the interaction. Where the K_D values are less/not reliable due to either sub stoichiometric binding or suboptimal concentration ranges for the dose-response titration were annotated by (>) sign. The compounds were tested in duplicates. Data from one replicate is shown here for clarity.
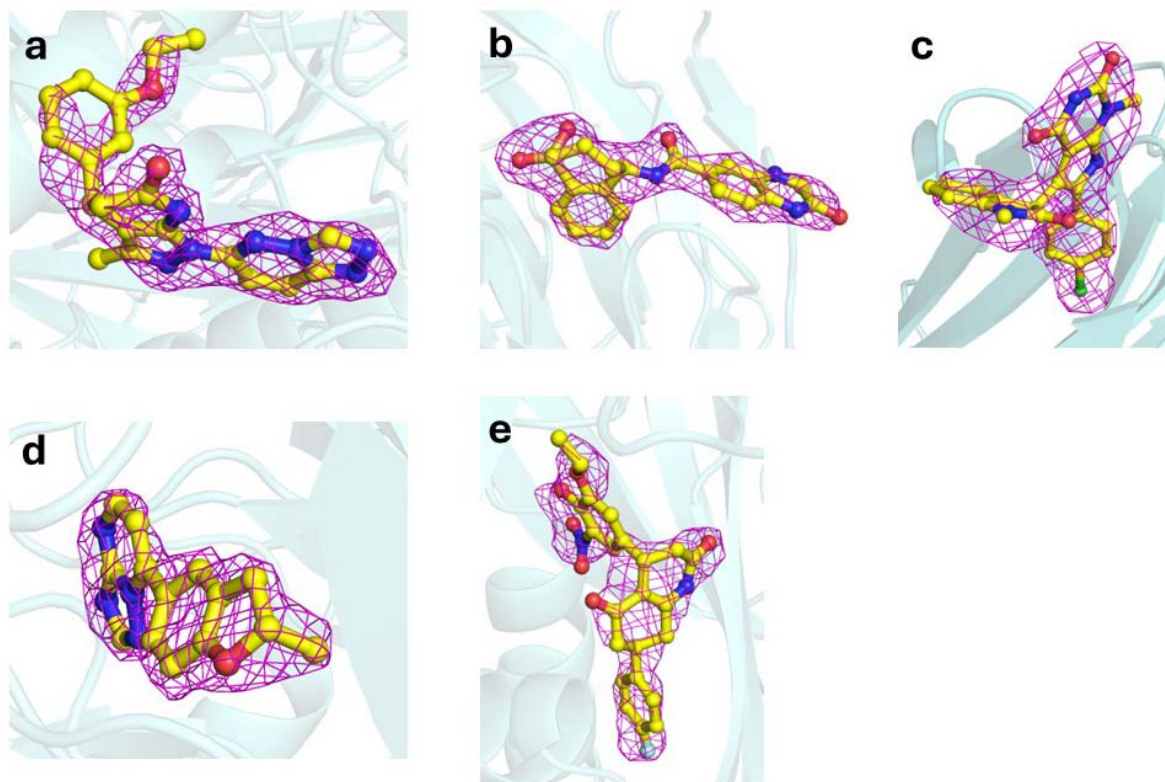
Supplementary Fig. 9 | The mFo-dFc electron density omit-maps for the EAS-MS hits (shown in yellow sticks) in the crystal structures of (**a**) DDB1_XS381952, (**b**) WDR91_XS838489, (**c**) WDR91_XS381295, (**d**) WDR55_XS381774, and (**e**) HAT1_XS380871 displayed as magenta meshes and contoured at 3σ.
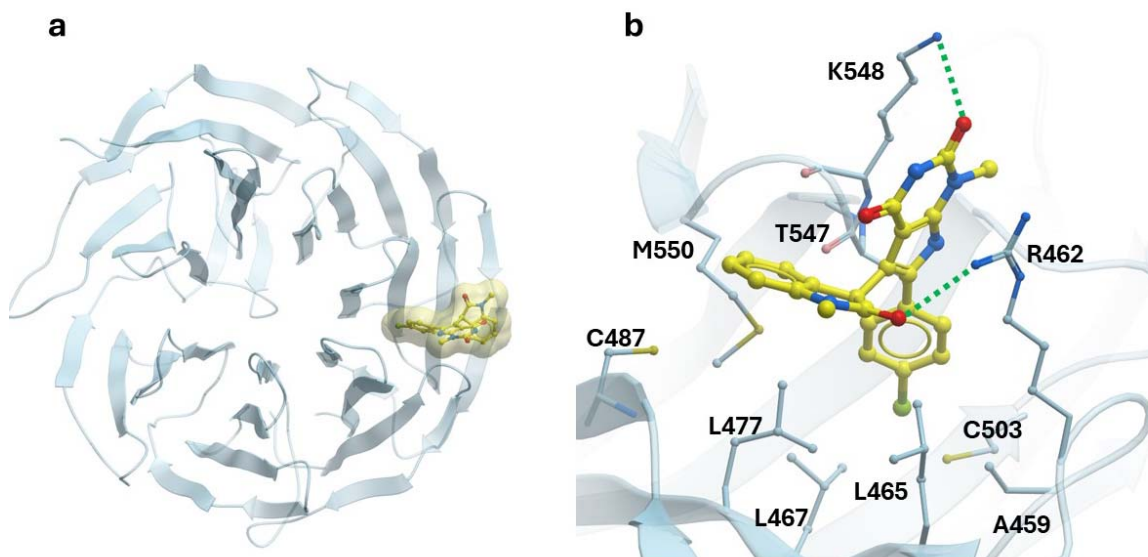
Supplementary Fig. 10 | Co-crystal structure of WDR91 in complex with XS381295. (**a**) Cartoon representation of WDR91 (cyan) bound to ligand (yellow). (**b**) Close-up view of the XS381295 binding site. Hydrogen bonds between the protein and the ligand are shown as black green dashed lines.
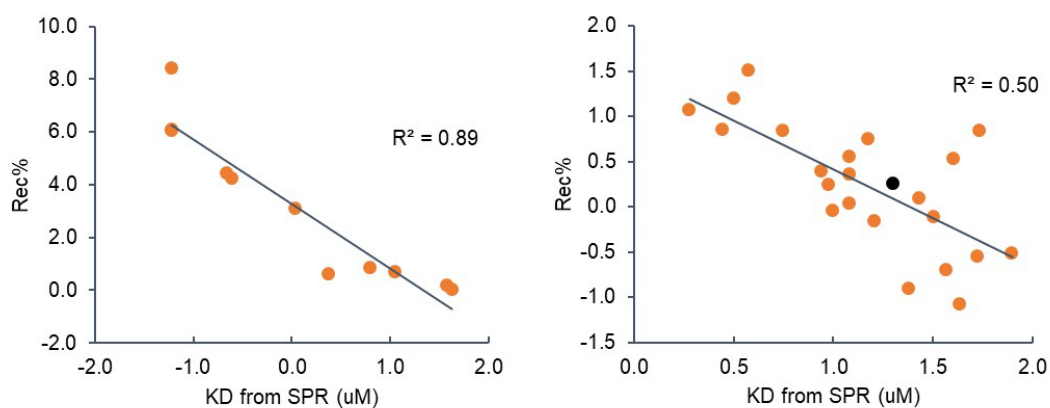
Supplementary Fig. 11 | Strong correlations between hit MS signals and K_D values of confirmed hits for single protein (WDR5, left) and all 11 proteins (right).

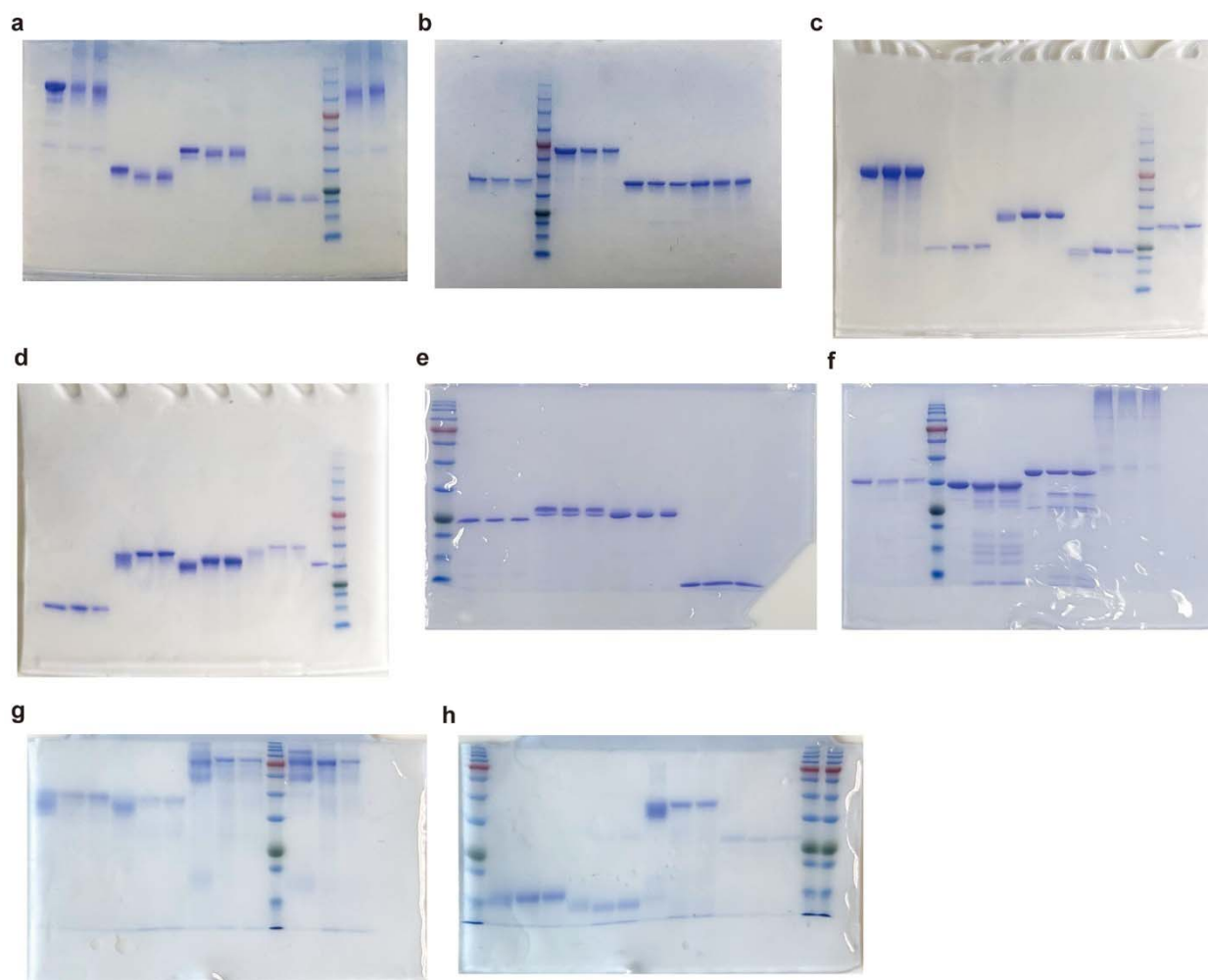**Supplementary Fig. 12 | Uncropped SDS-PAGE images.** Each protein is shown in three lanes: total protein amount and two pulldown samples. (**a**) DDB1, COPB2(WDR), DCAF1(WDR), and DNMT3A. (**b**) LRRK2(WDR), NSP13, WDR91, and RFWD3. (**c**) DDX1, SETDB1, SEC31A, and SKP1. (**d**) MSH6, FBXW7, CHIP, and PIAS1. (**e**) TIF1A, TRIM28, SPSB2, and BIRC3(BIR1). (**f**) WDR55, TLE4, HAT1, and HTT. (**g**) THEMIS, DCAF7, AASS(LKR), and AASS(SDH). (**h**) YTHDC1, YTHDF2, and SKP1-FBXO22.

Supplementary Table 3. Data collection and refinement statistics

	WDR91_XS381295	WDR91_XS838489	DDB1_XS381952

PDB code	9EJO	9EJP	9EJQ
Data collection			
Space group	$C222_1$	$C222_1$	$P2_12_12_1$
Cell dimensions			
a, b, c (Å)	78.63, 132.38, 119.60	77.61, 132.55, 119.80	62.64, 124.70, 167.72
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Resolution (Å)	50.0-2.40 (2.44-2.40)*	50.0-2.22 (2.26-2.22)*	50.0-1.87(1.90-1.87)*
R_{sym} or R_{merge}	0.086 (0.966)	0.062 (0.904)	0.068 (0.940)
CC1/2	0.999 (0.716)	1.001 (0.802)	0.999 (0.666)
$I / \sigma I$	18.9 (1.9)	31.0 (2.7)	18.8 (1.9)
Completeness (%)	99.9 (100.0)	100.0 (100.0)	91.2 (94.1)
Redundancy	6.0 (6.0)	10.8 (10.8)	4.3 (4.7)
Refinement			
Resolution (Å)	27.70-2.40 (2.46-2.40)	44.67-2.22 (2.28-2.22)	25.10-1.87 (1.92-1.87)
No. reflections	23892	29329	94637
R_{work} / R_{free}(%)	19.9/23.8	20.3/24.5	16.4/22.0
No. atoms	2650	2682	9508
Protein	2535	2561	8739
Ligand/ion	30	27	29
Water	84	94	624
B-factors	48.0	48.5	31.5
Protein	48.1	48.5	30.8
Ligand/ion	55.5	57.6	45.2
Water	41.2	46.5	37.9
R.m.s. deviations			
Bond lengths (Å)	0.006	0.005	0.007
Bond angles (°)	1.295	1.342	1.282

722   *Values in parentheses are for highest-resolution shell.

723

Supplementary Table 3. Data collection and refinement statistics (continued)

	WDR55_XS381774	HAT1_XS380871
PDB code	9EKP	9MJG
Data collection		
Space group	P2$_1$	P1
Cell dimensions		
a, b, c (Å)	78.31, 58.41, 87.22	78.89, 87.84, 118.38
α, β, γ (°)	90.0, 96.3, 90.0	84.38, 80.07, 76.76
Resolution (Å)	34.81-1.95 (2.00-1.95)*	50.0-2.58 (2.62-2.58)*
R$_{sym}$ or R$_{merge}$	0.082 (0.685)	0.127 (0.653)
CC1/2	0.998 (0.774)	0.984 (0.602)
I / σI	14.1 (2.4)	8.3 (1.3)
Completeness (%)	97.5 (95.9)	96.0 (76.5)
Redundancy	4.9 (5.0)	2.8 (2.3)
Refinement		
Resolution (Å)	34.83-1.95 (2.00-1.95)	49.45-2.58 (2.65-2.58)
No. reflections	53046	87207
R_{work} / R_{free}(%)	15.6/20.6	23.5/27.0
No. atoms	5101	20479
Protein	4696	20130
Ligand/ion	38	256
Water	339	93
B-factors	31.5	46.6
Protein	30.9	47.8
Ligand/ion	24.5	37.9
Water	39.4	26.0

R.m.s. deviations		
Bond lengths (Å)	0.005	0.005
Bond angles (°)	1.289	1.266

725   *Values in parentheses are for highest-resolution shell.

726

727   **Supplementary Table 4. Detailed instrumental parameters of LC-MS.**

Parameter setting of the mass spectrometer	
Ion Source:	
Spray Voltage: 3500 V (pos); 2500 V (neg)	Ion Transfer Tube Temperature: 325
Sheath Gas: 30 Arb	Auxiliary Gas: 10 Arb
Vaporize Temperature: 350	Ion Source Type: H-ESI
Full scan parameters:	
Scan range: 150-520 m/z	Resolution: 60,000
S-lens RF (%): 70	Automatic Gain Control (AGC) Target: Standard
DIA mode parameters for MS^2:	
Precursor Mass Range:	Resolution: 22,500
Isolation Window (m/z): 18	Window Overlap: 1 m/z
AGC Target: Standard	HCD Collision Energy (%): 15, 30, 60
Loop Control: N	N (Number of Spectra): 7

728