### Additional Files

**Table S1 — Precision of annotation detection by extreme ubiquitous words**

<table>
<thead>
<tr>
<th>Word</th>
<th>Annotation</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>PZCD</td>
<td>DISULFID</td>
<td>6%</td>
</tr>
<tr>
<td>HBDS</td>
<td>DISULFID</td>
<td>10%</td>
</tr>
<tr>
<td>ZCDS</td>
<td>METAL</td>
<td>8%</td>
</tr>
<tr>
<td>UFQK</td>
<td>DISULFID</td>
<td>19%</td>
</tr>
<tr>
<td>GYUQ</td>
<td>DISULFID</td>
<td>4%</td>
</tr>
<tr>
<td>YBDS</td>
<td>DISULFID</td>
<td>8%</td>
</tr>
<tr>
<td>FQLG</td>
<td>DISULFID</td>
<td>5%</td>
</tr>
<tr>
<td>YZDS</td>
<td>DISULFID</td>
<td>5%</td>
</tr>
<tr>
<td>GUDO</td>
<td>-</td>
<td>0%</td>
</tr>
<tr>
<td>FFFI</td>
<td>DISULFID</td>
<td>6%</td>
</tr>
<tr>
<td>FQKG</td>
<td>DISULFID</td>
<td>19%</td>
</tr>
<tr>
<td>SLGI</td>
<td>REPEAT</td>
<td>5%</td>
</tr>
<tr>
<td>QLGI</td>
<td>DISULFID</td>
<td>7%</td>
</tr>
<tr>
<td>DRPI</td>
<td>MUTAGEN</td>
<td>1%</td>
</tr>
<tr>
<td>DSPI</td>
<td>DISULFID</td>
<td>2%</td>
</tr>
<tr>
<td>DSGI</td>
<td>DISULFID</td>
<td>2%</td>
</tr>
<tr>
<td>DSKG</td>
<td>DISULFID</td>
<td>8%</td>
</tr>
<tr>
<td>DSKH</td>
<td>DISULFID</td>
<td>8%</td>
</tr>
<tr>
<td>DOIP</td>
<td>-</td>
<td>0%</td>
</tr>
<tr>
<td>OIPI</td>
<td>-</td>
<td>0%</td>
</tr>
<tr>
<td>HBBQ</td>
<td>DISULFID</td>
<td>11%</td>
</tr>
<tr>
<td>BQGI</td>
<td>DISULFID</td>
<td>4%</td>
</tr>
<tr>
<td>SKGI</td>
<td>DISULFID</td>
<td>8%</td>
</tr>
<tr>
<td>DGPI</td>
<td>NP_BIND</td>
<td>11%</td>
</tr>
</tbody>
</table>
Table S2 — Analysis of UQHS fragments

Firstly, we compared UQHS fragments with Swiss-Prot annotations. Then, for unannotated fragments, we predicted Repeat regions using REP software (Andrade et al., 2000).

<table>
<thead>
<tr>
<th>Proteins</th>
<th>Word position</th>
<th>Swiss-Prot Annotation</th>
<th>REP</th>
</tr>
</thead>
<tbody>
<tr>
<td>1aor_A</td>
<td>66-72</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1chk_A</td>
<td>84-90</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1dce_A</td>
<td>470-476</td>
<td>-</td>
<td>LRR: 484-507</td>
</tr>
<tr>
<td>1dce_A</td>
<td>493-499</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1hye_A</td>
<td>52-58</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1nt4_A</td>
<td>107-113</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1ogq_A</td>
<td>108-114</td>
<td>REPEAT</td>
<td></td>
</tr>
<tr>
<td>1ogq_A</td>
<td>132-138</td>
<td>REPEAT</td>
<td></td>
</tr>
<tr>
<td>1ogq_A</td>
<td>156-162</td>
<td>REPEAT</td>
<td></td>
</tr>
<tr>
<td>1ogq_A</td>
<td>181-187</td>
<td>REPEAT</td>
<td></td>
</tr>
<tr>
<td>1ogq_A</td>
<td>204-210</td>
<td>REPEAT</td>
<td></td>
</tr>
<tr>
<td>1ogq_A</td>
<td>228-234</td>
<td>REPEAT</td>
<td></td>
</tr>
<tr>
<td>1ogq_A</td>
<td>251-257</td>
<td>REPEAT</td>
<td></td>
</tr>
<tr>
<td>1ogq_A</td>
<td>275-281</td>
<td>REPEAT</td>
<td></td>
</tr>
<tr>
<td>1omw_A</td>
<td>554-560</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1ozn_A</td>
<td>185-191</td>
<td>REPEAT</td>
<td></td>
</tr>
<tr>
<td>1ozn_A</td>
<td>209-215</td>
<td>REPEAT</td>
<td></td>
</tr>
<tr>
<td>1ozn_A</td>
<td>233-239</td>
<td>REPEAT</td>
<td></td>
</tr>
<tr>
<td>1ozn_A</td>
<td>64-70</td>
<td>REPEAT</td>
<td></td>
</tr>
<tr>
<td>1v1b_A</td>
<td>175-181</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1w4x_A</td>
<td>436-442</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2cl5_A</td>
<td>198-204</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Table S3 — Analysis of DODQ fragments

Firstly, we compared DODQ fragments with Swiss-Prot annotations. Then, for unannotated fragments, we predicted residues involved in a calcium-binding sites using SitePredict software (Bordner et al., 2008). Only residues predicted with a score higher than 0.5 and matching with studied DODQ-fragments are indicated.

1: pdb codes.

<table>
<thead>
<tr>
<th>Proteins 1</th>
<th>Word position</th>
<th>Swiss-Prot Annotation</th>
<th>SitePredict</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a12_A</td>
<td>376-382</td>
<td>-</td>
<td>373,382</td>
</tr>
<tr>
<td>1dab_A</td>
<td>524-530</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>1ij5_A</td>
<td>231-237</td>
<td>CA_BIND</td>
<td></td>
</tr>
<tr>
<td>1ij5_A</td>
<td>266-272</td>
<td>CA_BIND</td>
<td></td>
</tr>
<tr>
<td>1ij5_A</td>
<td>333-339</td>
<td>CA_BIND</td>
<td></td>
</tr>
<tr>
<td>1k1x_A</td>
<td>393-399</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>1k9u_A</td>
<td>14-20</td>
<td>CA_BIND</td>
<td></td>
</tr>
<tr>
<td>1k9u_A</td>
<td>49-55</td>
<td>CA_BIND</td>
<td></td>
</tr>
<tr>
<td>1l1d_A</td>
<td>437-443</td>
<td>-</td>
<td>440,439,441,437</td>
</tr>
<tr>
<td>1omr_A</td>
<td>111-117</td>
<td>CA_BIND</td>
<td></td>
</tr>
<tr>
<td>1qs_A</td>
<td>67-73</td>
<td>CA_BIND</td>
<td></td>
</tr>
<tr>
<td>1qus_A</td>
<td>238-244</td>
<td>-</td>
<td>240,242,241,239,243,238,244</td>
</tr>
<tr>
<td>1qv1_A</td>
<td>31-37</td>
<td>CA_BIND</td>
<td></td>
</tr>
<tr>
<td>1wdc_B</td>
<td>29-35</td>
<td>CA_BIND</td>
<td></td>
</tr>
<tr>
<td>1wde_B</td>
<td>98-104</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1wde_C</td>
<td>95-101</td>
<td>-</td>
<td>98,96,97,100</td>
</tr>
<tr>
<td>1xoc_A</td>
<td>347-353</td>
<td>-</td>
<td>349,348,350,351,347,352,353</td>
</tr>
<tr>
<td>2h61_A</td>
<td>62-68</td>
<td>-</td>
<td>63,65,64,62,66,67,68,</td>
</tr>
<tr>
<td>2pvb_A</td>
<td>52-58</td>
<td>CA_BIND</td>
<td></td>
</tr>
<tr>
<td>2pvb_A</td>
<td>91-97</td>
<td>CA_BIND</td>
<td></td>
</tr>
<tr>
<td>2scp_A</td>
<td>105-111</td>
<td>CA_BIND</td>
<td></td>
</tr>
<tr>
<td>2scp_A</td>
<td>139-145</td>
<td>CA_BIND</td>
<td></td>
</tr>
<tr>
<td>2scp_A</td>
<td>17-23</td>
<td>CA_BIND</td>
<td></td>
</tr>
</tbody>
</table>
Table S4 — Analysis of UODO-unannotated fragments

Firstly, we compared UODO fragments with Swiss-Prot annotations. Then, for unannotated fragments, we predicted residues involved in an ATP/GTP-binding sites using SitePredict software (Bordner et al., 2008). Only residues predicted with a score higher than 0.5 and matching with studied UODO-fragments are indicated.

1: pdb codes.

<table>
<thead>
<tr>
<th>Proteins</th>
<th>Word position</th>
<th>Swiss-Prot Annotation</th>
<th>SitePredict</th>
</tr>
</thead>
<tbody>
<tr>
<td>1c9k_A</td>
<td>7-13</td>
<td>NP_BIND</td>
<td></td>
</tr>
<tr>
<td>1cnz_A</td>
<td>290-296</td>
<td>NP_BIND</td>
<td></td>
</tr>
<tr>
<td>1cr1_A</td>
<td>313-319</td>
<td>NP_BIND</td>
<td></td>
</tr>
<tr>
<td>1d2n_A</td>
<td>552-558</td>
<td>NP_BIND</td>
<td></td>
</tr>
<tr>
<td>1e2k_A</td>
<td>57-63</td>
<td>NP_BIND</td>
<td></td>
</tr>
<tr>
<td>1e6c_A</td>
<td>10-16</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1f9v_A</td>
<td>475-481</td>
<td>NP_BIND</td>
<td></td>
</tr>
<tr>
<td>1fzq_A</td>
<td>25-31</td>
<td>NP_BIND</td>
<td></td>
</tr>
<tr>
<td>1g6h_A</td>
<td>41-47</td>
<td>NP_BIND</td>
<td></td>
</tr>
<tr>
<td>1g7s_A</td>
<td>13-19</td>
<td>NP_BIND</td>
<td></td>
</tr>
<tr>
<td>1h72_C</td>
<td>90-96</td>
<td>NP_BIND</td>
<td></td>
</tr>
<tr>
<td>1hgs_A</td>
<td>351-357</td>
<td>NP_BIND</td>
<td></td>
</tr>
<tr>
<td>1htw_A</td>
<td>41-47</td>
<td>NP_BIND</td>
<td></td>
</tr>
<tr>
<td>1ihu_A</td>
<td>16-22</td>
<td>NP_BIND</td>
<td></td>
</tr>
<tr>
<td>1ihu_A</td>
<td>335-341</td>
<td>NP_BIND</td>
<td></td>
</tr>
<tr>
<td>1iic_A</td>
<td>177-183</td>
<td>-</td>
<td>177,178,179,180,181,182,183</td>
</tr>
<tr>
<td>1in4_A</td>
<td>59-65</td>
<td>NP_BIND</td>
<td></td>
</tr>
<tr>
<td>1jjv_A</td>
<td>10-16</td>
<td>NP_BIND</td>
<td></td>
</tr>
<tr>
<td>1kk1_A</td>
<td>18-24</td>
<td>NP_BIND</td>
<td></td>
</tr>
<tr>
<td>1kkh_A</td>
<td>109-115</td>
<td>NP_BIND</td>
<td></td>
</tr>
<tr>
<td>1ko7_A</td>
<td>152-158</td>
<td>NP_BIND</td>
<td></td>
</tr>
<tr>
<td>1kvk_A</td>
<td>138-144</td>
<td>NP_BIND</td>
<td></td>
</tr>
<tr>
<td>1l2t_A</td>
<td>39-45</td>
<td>NP_BIND</td>
<td></td>
</tr>
<tr>
<td>1l8q_A</td>
<td>120-126</td>
<td>NP_BIND</td>
<td></td>
</tr>
<tr>
<td>1lnz_A</td>
<td>166-172</td>
<td>NP_BIND</td>
<td></td>
</tr>
<tr>
<td>1m0w_A</td>
<td>385-391</td>
<td>-</td>
<td>385,386,387</td>
</tr>
<tr>
<td>1mja_A</td>
<td>311-317</td>
<td>-</td>
<td>312,313,314,315,316,317</td>
</tr>
<tr>
<td>1mkv_A</td>
<td>9-15</td>
<td>NP_BIND</td>
<td></td>
</tr>
<tr>
<td>1mky_A</td>
<td>188-194</td>
<td>NP_BIND</td>
<td></td>
</tr>
<tr>
<td>1nri_B</td>
<td>46-52</td>
<td>NP_BIND</td>
<td></td>
</tr>
<tr>
<td>1odf_A</td>
<td>39-45</td>
<td>NP_BIND</td>
<td></td>
</tr>
<tr>
<td>1ogo_X</td>
<td>75-81</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1p6x_A</td>
<td>33-39</td>
<td>NP_BIND</td>
<td></td>
</tr>
<tr>
<td>1ptm_A</td>
<td>303-309</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1puj_A</td>
<td>128-134</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1qde_A</td>
<td>66-72</td>
<td>NP_BIND</td>
<td></td>
</tr>
<tr>
<td>1qhx_A</td>
<td>11-17</td>
<td>NP_BIND</td>
<td></td>
</tr>
<tr>
<td>1qsm_A</td>
<td>99-105</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1r2q_A</td>
<td>28-34</td>
<td>NP_BIND</td>
<td></td>
</tr>
<tr>
<td>1svi_A</td>
<td>31-37</td>
<td>NP_BIND</td>
<td></td>
</tr>
</tbody>
</table>

Continued on next page
<table>
<thead>
<tr>
<th>Proteins [1]</th>
<th>Word position</th>
<th>Swiss-Prot Annotation</th>
<th>SitePredict</th>
</tr>
</thead>
<tbody>
<tr>
<td>1sxj_A</td>
<td>354-360</td>
<td>NP_BIND</td>
<td></td>
</tr>
<tr>
<td>1sxj_C</td>
<td>54-60</td>
<td>NP_BIND</td>
<td></td>
</tr>
<tr>
<td>1sxj_D</td>
<td>66-72</td>
<td>NP_BIND</td>
<td></td>
</tr>
<tr>
<td>1sxj_E</td>
<td>44-50</td>
<td>NP_BIND</td>
<td></td>
</tr>
<tr>
<td>1u0j_A</td>
<td>335-341</td>
<td>NP_BIND</td>
<td></td>
</tr>
<tr>
<td>1u0l_A</td>
<td>171-177</td>
<td>NP_BIND</td>
<td></td>
</tr>
<tr>
<td>1w0d_A</td>
<td>276-282</td>
<td>NP_BIND</td>
<td></td>
</tr>
<tr>
<td>1w0m_A</td>
<td>145-151</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1w1w_A</td>
<td>34-40</td>
<td>NP_BIND</td>
<td></td>
</tr>
<tr>
<td>1w7j_A</td>
<td>164-170</td>
<td>NP_BIND</td>
<td></td>
</tr>
<tr>
<td>1ye8_A</td>
<td>8-14</td>
<td>NP_BIND</td>
<td></td>
</tr>
<tr>
<td>1z2a_A</td>
<td>17-23</td>
<td>NP_BIND</td>
<td></td>
</tr>
<tr>
<td>2bcg_Y</td>
<td>16-22</td>
<td>NP_BIND</td>
<td></td>
</tr>
<tr>
<td>2bm0_A</td>
<td>20-26</td>
<td>NP_BIND</td>
<td></td>
</tr>
<tr>
<td>2c78_A</td>
<td>19-25</td>
<td>NP_BIND</td>
<td></td>
</tr>
<tr>
<td>2cxx_A</td>
<td>25-31</td>
<td>NP_BIND</td>
<td></td>
</tr>
<tr>
<td>2cxx_A</td>
<td>9-15</td>
<td>-</td>
<td>9,10,11,12,13,14,15</td>
</tr>
<tr>
<td>2d0t_A</td>
<td>155-161</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>2e3b_A</td>
<td>74-80</td>
<td>NP_BIND</td>
<td></td>
</tr>
<tr>
<td>2ilq_A</td>
<td>106-112</td>
<td>NP_BIND</td>
<td></td>
</tr>
<tr>
<td>3tmk_A</td>
<td>13-19</td>
<td>NP_BIND</td>
<td></td>
</tr>
</tbody>
</table>

[1] Proteins: Additional information about the proteins is available elsewhere.
Table S5 — Analysis of EIJU fragments

Firstly, we compared EIJU fragments with Swiss-Prot annotations. Then, for unannotated fragments, we predicted residues involved in a NAD(P)-binding sites using SitePredict software (Bordner et al., 2008). Only residues predicted with a score higher than 0.5 and matching with studied EIJU-fragments are indicated.

1: pdb codes.

<table>
<thead>
<tr>
<th>Proteins1</th>
<th>Word position</th>
<th>Swiss-Prot Annotation</th>
<th>SitePredict</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a0c_A</td>
<td>388-394</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1aoz_A</td>
<td>535-541</td>
<td>NP_BIND</td>
<td>-</td>
</tr>
<tr>
<td>1cyd_A</td>
<td>14-20</td>
<td>NP_BIND</td>
<td>-</td>
</tr>
<tr>
<td>1edo_A</td>
<td>24-30</td>
<td>NP_BIND</td>
<td>-</td>
</tr>
<tr>
<td>1gee_A</td>
<td>14-20</td>
<td>NP_BIND</td>
<td>-</td>
</tr>
<tr>
<td>1h5q_A</td>
<td>18-24</td>
<td>-</td>
<td>18,19,20,21,22,23,24</td>
</tr>
<tr>
<td>1hdc_A</td>
<td>13-19</td>
<td>NP_BIND</td>
<td>-</td>
</tr>
<tr>
<td>1iy8_A</td>
<td>20-26</td>
<td>NP_BIND</td>
<td>-</td>
</tr>
<tr>
<td>1k2w_A</td>
<td>12-18</td>
<td>NP_BIND</td>
<td>-</td>
</tr>
<tr>
<td>1m93_B</td>
<td>256-262</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1oy5_A</td>
<td>21-27</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1ve9_A</td>
<td>108-114</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1w6g_A</td>
<td>562-568</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1x7g_A</td>
<td>13-19</td>
<td>NP_BIND</td>
<td>-</td>
</tr>
<tr>
<td>2ae2_A</td>
<td>16-22</td>
<td>-</td>
<td>16,17,18,19,20,21,22</td>
</tr>
</tbody>
</table>
Table S6 — Analysis of UGRU fragments

The structural word UGRU is seen 37 times in the initial data set and is strongly over-represented in the “S-adenosyl-L-methionine-dependent methyltransferase” superfamily (SCOP id=53335). When we restrict the analysis to proteins present in the PDB/UniProt Mapping database (= 1487 proteins composing the annotation data set), it is seen only 12 times. Only four of these 12 fragments (precision=33%) are annotated as “Binding”, resulting in the fact that this word is not defined as functional word. The manual analysis of the functional annotations of the 29 UGRU fragments seen in the initial data set and in the “S-adenosyl-L-methionine-dependent methyltransferase” superfamily through the Swiss-Prot web interface (http://www.uniprot.org/uniprot/) revealed that 12 fragments are actually S-adenosyl-L-methionine (SAH/SAM) binding sites. Furthermore eight of 17 unannotated UGRU fragments are extracted from proteins co-crystallized with SAH/SAM, and the UGRU fragments are actually involved in the binding site. Thus, 69% of the 29 UGRU fragments are indeed involved in SAH/SAM-binding sites.

Firstly, we compared UGRU fragments with Swiss-Prot annotations. Then, for unannotated fragments, we extracted residues involved in the SAH-binding sites using Ligplot (Wallace et al., 1995). Only SAH-binding residues matching with UGRU-fragments are presented.

<table>
<thead>
<tr>
<th>Proteins</th>
<th>Position</th>
<th>Swiss-Prot annotation</th>
<th>Ligand</th>
<th>Residues involved in binding site</th>
</tr>
</thead>
<tbody>
<tr>
<td>13i_A</td>
<td>39-45</td>
<td></td>
<td>SAH</td>
<td>44,45</td>
</tr>
<tr>
<td>1kyz_A</td>
<td>206-212</td>
<td>Binding site</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1nv8_A</td>
<td>127-133</td>
<td></td>
<td>SAM-MEQ</td>
<td>129</td>
</tr>
<tr>
<td>2gh1_A</td>
<td>45-51</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1dus_A</td>
<td>61-67</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1nw3_A</td>
<td>161-167</td>
<td></td>
<td>SAM</td>
<td>161,163</td>
</tr>
<tr>
<td>1ri5_A*</td>
<td>70-76</td>
<td>Binding site</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1rjd_A</td>
<td>103-109</td>
<td>Binding site</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1im8_A</td>
<td>61-67</td>
<td></td>
<td>SAI</td>
<td>63,65</td>
</tr>
<tr>
<td>2aoi_A</td>
<td>58-64</td>
<td>Binding site</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1fp1_D</td>
<td>215-221</td>
<td>Binding site</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2fk8_A</td>
<td>79-85</td>
<td></td>
<td>SAM</td>
<td>81,83</td>
</tr>
<tr>
<td>2fyt_A</td>
<td>278-284</td>
<td>Binding site</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1kvy_A</td>
<td>42-48</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1qyr_A</td>
<td>43-49</td>
<td>Binding site</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1jsx_A</td>
<td>71-77</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1ne2_A</td>
<td>54-60</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2ex4_A</td>
<td>68-74</td>
<td>Binding site</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1iw4_A</td>
<td>52-58</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1or8_A</td>
<td>76-82</td>
<td>Binding site</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1zkd_A</td>
<td>86-92</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1y8c_A</td>
<td>44-50</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1wz7n_A</td>
<td>47-53</td>
<td></td>
<td>SAH</td>
<td>49</td>
</tr>
<tr>
<td>1dl5_A</td>
<td>81-87</td>
<td></td>
<td>SAH</td>
<td></td>
</tr>
<tr>
<td>1zq9_A</td>
<td>62-68</td>
<td>Binding site</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1vl5_A</td>
<td>48-54</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1xxl_A</td>
<td>18-24</td>
<td>Binding site</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1yzh_A</td>
<td>44-50</td>
<td>Binding site</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1qzz_A</td>
<td>188-194</td>
<td></td>
<td>SAM</td>
<td>190</td>
</tr>
</tbody>
</table>

1: pdb codes.
Table S7 — Analysis of ZCLH fragments

Firstly, we compared ZCLH fragments with Swiss-Prot annotations. Then, for unannotated fragments, we extracted residues involved in a catalytic site using the CSA database (Porter et al., 1994). Only catalytic site residues matching with ZCLH-fragments are presented.  

1: pdb codes.

<table>
<thead>
<tr>
<th>Proteins</th>
<th>position</th>
<th>Annotation</th>
<th>Residues in catalytic site</th>
</tr>
</thead>
<tbody>
<tr>
<td>1auo_A</td>
<td>168-174</td>
<td>Active site</td>
<td></td>
</tr>
<tr>
<td>1c4s_A</td>
<td>235-241</td>
<td>-</td>
<td>235</td>
</tr>
<tr>
<td>1fj2_A</td>
<td>169-175</td>
<td>Active site</td>
<td></td>
</tr>
<tr>
<td>1isp_A</td>
<td>133-139</td>
<td>Active site</td>
<td></td>
</tr>
<tr>
<td>1j1i_A</td>
<td>233-239</td>
<td>-</td>
<td>233</td>
</tr>
<tr>
<td>luxo_A</td>
<td>137-143</td>
<td>Active site</td>
<td></td>
</tr>
<tr>
<td>1vlq_A</td>
<td>274-280</td>
<td>-</td>
<td>274</td>
</tr>
<tr>
<td>1wm1_A</td>
<td>268-274</td>
<td>Active site</td>
<td></td>
</tr>
</tbody>
</table>
Table S8 — Computation of a random sensitivity for each functional word

Random sensitivity values for functional word and the associated functional annotation. In order to analyze the significance of sensitivity values obtained for each pair functional word / annotation, we compare these values to ones computed in a random dataset. The random data set is built by randomly assigning a functional annotation for each word. Using this random data set, we compute the random recall between the each pairs functional word / annotation. We ran this procedure 10,000 times and we compute the average random recall values. \(^1\) and \(^2\): are computed on the 10,000 simulations.

<table>
<thead>
<tr>
<th>Word</th>
<th>Swiss-Prot annotation</th>
<th>Sensitivity</th>
<th>Random sensitivity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean(^1)</td>
<td>Standard deviation(^2)</td>
</tr>
<tr>
<td>DODQ</td>
<td>Calcium-binding sites</td>
<td>95</td>
<td>0.35</td>
</tr>
<tr>
<td>ZDOD</td>
<td>Calcium-binding sites</td>
<td>64</td>
<td>0.27</td>
</tr>
<tr>
<td>YUOD</td>
<td>ATP/GTP-binding sites</td>
<td>29</td>
<td>3.3</td>
</tr>
<tr>
<td>UODO</td>
<td>ATP/GTP-binding sites</td>
<td>40</td>
<td>4.25</td>
</tr>
<tr>
<td>OEIJ</td>
<td>NAD(P)-binding sites</td>
<td>7</td>
<td>1.57</td>
</tr>
<tr>
<td>EIJU</td>
<td>NAD(P)-binding sites</td>
<td>10</td>
<td>2.12</td>
</tr>
<tr>
<td>RUDO</td>
<td>SAM/SAH-binding sites</td>
<td>44</td>
<td>0.7</td>
</tr>
</tbody>
</table>