**PLP\_FS: Prediction of lysine phosphoglycerylation sites in protein using support vector machine and fusion of multiple F\_Score feature selection**

Md. Sohrawordi1,2, Md. Ali Hossain1and Md. Al Mehedi Hasan1

*1* Dept. of Computer Science and Engineering, Rajshahi University of Engineering and Technology, Rajshahi, Bangladesh

2 Dept. of Computer Science and Engineering, Hajee Mohammad Danesh Science and Technology University, Dinajpur, Bangladesh

**Corresponding Author:**

Md. Sohrawordi

Department of Computer Science and Engineering

Hajee Mohammad Danesh Science and Technology University, Dinajpur, Bangladesh

Mobile: +8801722980888

Email: [mdsohrawordicse@gmail.com](mailto:mdsohrawordicse@gmail.com)

**Supplementary materials**

Supplementary Table 1: A list of Non-redundant Protein Containing Phosphoglycerylated Lysine Residues

|  |  |  |
| --- | --- | --- |
| **Serial Number** | **UniProt Accession** | **Position of Phosphoglycerylated Lysine** |
| 1 | P05387 | 99 |
| 2 | P06733 | 92,335,343,394 |
| 3 | P0DMV8 | 597 |
| 4 | P14174 | 78 |
| 5 | P14618 | 367 |
| 6 | P15259 | 100 |
| 7 | P17844 | 91 |
| 8 | P49006 | 144 |
| 9 | P51814 | 297 |
| 10 | P60174 | 256 |
| 11 | P61604 | 56 |
| 12 | P68104 | 330 |
| 13 | Q13541 | 69 |
| 14 | Q5JSZ5 | 386,398 |
| 15 | Q5RL73 | 351 |
| 16 | Q5T4S7 | 4685 |
| 17 | Q6PKG0 | 1083 |
| 18 | Q86UP2 | 292 |
| 19 | Q9BYV9 | 177 |
| 20 | Q9C026 | 347 |
| 21 | Q9H1E3 | 228 |
| 22 | Q9Y597 | 796 |
| 23 | Q9Y619 | 145 |
| 24 | C9J5S8 | 23 |
| 25 | O09172 | 263 |
| 26 | O55143 | 205 |
| 27 | O88844 | 224 |
| 28 | P02088 | 83,145 |
| 29 | P07724 | 236 |
| 30 | P11214 | 167 |
| 31 | P11881 | 569 |
| 32 | P11983 | 243 |
| 33 | P12710 | 96 |
| 34 | P12790 | 237 |
| 35 | P14824 | 442,445,446 |
| 36 | P14869 | 297 |
| 37 | P16015 | 64 |
| 38 | P16460 | 340 |
| 39 | P16879 | 161 |
| 40 | P18894 | 259,263 |
| 41 | P22752 | 96 |
| 42 | P23881 | 199 |
| 43 | P26040 | 60 |
| 44 | P26443 | 503 |
| 45 | P31786 | 55 |
| 46 | P35585 | 296 |
| 47 | P41216 | 544,552 |
| 48 | P47806 | 79 |
| 49 | P47857 | 678 |
| 50 | P48962 | 92,96 |
| 51 | P49722 | 92 |
| 52 | P61982 | 50 |
| 53 | P62814 | 48 |
| 54 | P62900 | 70 |
| 55 | P62908 | 201 |
| 56 | P97351 | 144 |
| 57 | P99028 | 83 |
| 58 | Q3UJB0 | 853 |
| 59 | Q3UL97 | 53 |
| 60 | Q3UNI1 | 697 |
| 61 | Q60932 | 33 |
| 62 | Q64FW2 | 55 |
| 63 | Q69ZK0 | 1141 |
| 64 | Q6P6M7 | 173 |
| 65 | Q6P8V7 | 47 |
| 66 | Q8BH04 | 262 |
| 67 | Q8BLF2 | 104 |
| 68 | Q8BMS1 | 214 |
| 69 | Q8C196 | 214,228,811,1479 |
| 70 | Q8VD72-2 | 33 |
| 71 | Q91V76 | 75 |
| 72 | Q920E5 | 57 |
| 73 | Q922Q1 | 66 |
| 74 | Q93092 | 314 |
| 75 | Q99L47 | 152 |
| 76 | Q9D020 | 123 |
| 77 | Q9DAY2 | 206 |
| 78 | Q9DB16 | 3 |
| 79 | Q9EP89 | 383 |
| 80 | Q9EQF5 | 256 |
| 81 | Q9EQP2 | 35 |
| 82 | Q9QWY8 | 675 |
| 83 | Q9QXX4 | 177 |
| 84 | A2AQC3 | 190 |
| 85 | A6ZI44 | 254 |
| 86 | B1AS29 | 367 |
| 87 | D3YYI5 | 252 |
| 88 | E9PZS8 | 56 |
| 89 | E9QNN1 | 1027 |
| 90 | G3UYJ7 | 249 |

**Supplementary Table 2:** The performances of PLP\_FS for various window sizes on 10-fold cross-validation.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Window Size** | **Acc (%)** | **Sn (%)** | **Sp (%)** | **MCC (%)** | **AUC (%)** |
| 7 | 91.48 | 88.34 | 93.41 | 81.92 | 98.34 |
| 9 | 89.35 | 88.35 | 90.01 | 77.85 | 97.48 |
| 11 | 90.56 | 93.20 | 89.07 | 80.43 | 98.89 |
| 13 | 89.47 | 95.14 | 85.41 | 79.49 | 98.49 |
| 15 | 91.01 | 93.20 | 89.71 | 81.42 | 98.83 |
| 17 | 94.93 | 94.17 | 95.34 | 88.92 | 98.67 |
| 19 | 96.35 | 95.14 | 97.08 | 92.21 | 98.41 |
| 21 | **98.88** | **99.02** | **98.78** | **97.64** | **99.96** |
| 23 | 97.48 | 98.06 | 97.14 | 94.65 | 99.92 |
| 25 | 95.53 | 97.08 | 94.28 | 90.24 | 99.71 |
| 27 | 93.52 | 96.15 | 92.06 | 86.64 | 99.35 |
| 29 | 94.96 | 96.12 | 94.28 | 89.42 | 99.65 |
| 31 | 94.24 | 93.20 | 94.85 | 87.72 | 99.51 |

**Supplementary Table 3:** A List of Optimal Features

|  |  |  |  |
| --- | --- | --- | --- |
| **Feature Type** | **Features** | **F\_score** | **Importance Order** |
| Amino Acid Composition (AAC) | Amino\_acid\_L | 201.712331 | 1 |
| Amino\_acid\_D | 166.339760 | 2 |
| Amino\_acid\_K | 69.588370 | 3 |
| Binary Encoding (BE) | Binary168 | 14.61274132 | 1 |
| Binary331 | 5.361580728 | 2 |
| Binary440 | 4.187893042 | 3 |
| Binary120 | 3.194729478 | 4 |
| Binary402 | 3.150841253 | 5 |
| Binary56 | 3.150841253 | 6 |
| Binary393 | 2.794236447 | 7 |
| Binary377 | 2.297223366 | 8 |
| Binary237 | 2.284011352 | 9 |
| Binary273 | 1.61834123 | 10 |
| Binary398 | 1.612372175 | 11 |
| Binary378 | 1.36852948 | 12 |
| Binary108 | 1.265432555 | 13 |
| Binary116 | 1.08739695 | 14 |
| Binary302 | 1.030507015 | 15 |
| Binary19 | 0.997829858 | 16 |
| Binary40 | 0.979614015 | 17 |
| Binary382 | 0.958440855 | 18 |
| Binary278 | 0.623361849 | 19 |
| Binary112 | 0.603930656 | 20 |
| Binary34 | 0.403568048 | 21 |
| Binary98 | 0.344902679 | 22 |
| Binary358 | 0.292834198 | 23 |
| Binary65 | 0.173436094 | 24 |
| Binary334 | 0.113140188 | 25 |
| Binary256 | 0.101623537 | 26 |
| Binary145 | 0.020602882 | 27 |
| Binary12 | 0.016370295 | 28 |
| Binary274 | 0.016370295 | 29 |
| Binary419 | 4.72147E-06 | 30 |
| Composition of k-spaced amino acid pairs (CKSAAP) | K\_space\_SxV | 12.54684 | 1 |
| K\_space\_IxI | 4.707032 | 2 |
| K\_space\_CxxxS | 4.228953 | 3 |
| K\_space\_CxxL | 3.091942 | 4 |
| K\_space\_CxxA | 2.346721 | 5 |
| K\_space\_VxxxS | 1.988003 | 6 |
| K\_space\_MxxxA | 1.893173 | 7 |
| K\_space\_GxxP | 1.872536 | 8 |
| K\_space\_GxxxY | 1.225293 | 9 |
| K\_space\_AK | 1.162084 | 10 |
| K\_space\_NxxxC | 1.018368 | 11 |
| K\_space\_CxxxxV | 1.018368 | 12 |
| K\_space\_SxxxA | 0.99783 | 13 |
| K\_space\_RxY | 0.958441 | 14 |
| K\_space\_RxxN | 0.925934 | 15 |
| K\_space\_SxxD | 0.625608 | 16 |
| K\_space\_AxxxS | 0.463614 | 17 |
| K\_space\_CxxC | 0.450787 | 18 |
| K\_space\_ExG | 0.368759 | 19 |
| K\_space\_ExP | 0.354221 | 20 |
| K\_space\_LI | 0.310216 | 21 |
| K\_space\_YxC | 0.278034 | 22 |
| K\_space\_LxxxG | 0.228093 | 23 |
| K\_space\_KxxxxV | 0.141249 | 24 |
| K\_space\_KxxxxT | 0.137104 | 25 |
| K\_space\_ExA | 0.115739 | 26 |
| K\_space\_VxV | 0.055118 | 27 |
| K\_space\_IxxxxK | 0.005249 | 28 |
| K\_space\_PxxD | 0.001683 | 29 |
| K\_space\_AxR | 0.000005 | 30 |
| K\_space\_DxxxE | 0.000003 | 31 |