

Supporting Information

Technical Brief

Systematic study of protein sumoylation: development of a site-specific predictor of SUMOsp 2.0

Jian Ren¹, Xinjiao Gao¹, Changjiang Jin¹, Mei Zhu¹, Xiwei Wang¹, Andrew Shaw², Longping Wen¹,
Xuebiao Yao^{1**} and Yu Xue^{1*}

¹Hefei National Laboratory for Physical Sciences at Microscale and School of Life Sciences, University of Science & Technology of China, Hefei, Anhui 230027, China

²Department of Physiology and Cancer Biology Program, Morehouse School of Medicine, Atlanta, GA 30310, USA

Running title: Prediction of sumoylation sites

*Correspondence: Professor Yu Xue, Life Science School, USTC, Hefei 230027, China

E-mail: xueyu@ustc.edu.cn

Tel: +86-551-3607821;

**Additional corresponding author: Professor Xuebiao Yao, Email: yaorb@ustc.edu.cn

Abbreviations: **GPS**, Group-based Phosphorylation Scoring; **PSP**, potential sumoylation peptide; **MaM**, Matrix Mutation; **MS**, Mass Spectrometry; **OS**, Operating System; **JRE**, Java Runtime Environment; **Ac**, accuracy; **Sn**, sensitivity; **Sp**, specificity; **MCC**, Mathew correlation coefficient; **ROC**, Receiver Operating Characteristic; **SUMO-CS**, SUMO consensus sequence; **SVMs**, Support Vector Machines; **eNOS**, endothelial nitric oxide synthase; **ISE**, intronic splicing enhancer; **VEGF**, vascular endothelial growth factor

Supporting Information – Materials and Methods

Data preparation

Previously, we searched the PubMed with keywords of “SUMO” and “sumoylation”, and manually collected 239 experimentally verified sumoylation sites in 144 proteins [1]. This data set was used for SUMOsp 1.0 and SUMOsp 2.0 for performance comparison.

Also, by searching the additional research articles published online before Oct. 18th, 2007, the data set was expanded to include 355 sumoylation sites from 212 proteins. In this work, we arbitrarily took the 302 sumoylation sites from 181 proteins published before Feb., 2007 as the training data set. And the remnant 53 sites in 31 proteins were not included in training as an additional data set for performance evaluation. The protein sequences were retrieved from UniProt database [2].

As previously described [1, 3, 4], we regarded all known sumoylated lysine (K) residues as positive data (+), while all other non-sumoylated lysine residues were taken as negative data (-). The positive data (+) set for training might contain several homologous sites from homologous proteins. If the training data were highly redundant with too many homologous sites, the prediction accuracy would be overestimated. To avoid the overestimation, we clustered the protein sequences with a threshold of 40% identity by CD-HIT [5]. If two proteins were similar with $\geq 40\%$ identity, we re-aligned the proteins with BL2SEQ, a program in the BLAST package [6], and checked the results manually. If two sumoylation sites from two homologous proteins were at the same position after sequence alignment, only one item was reserved while the other was discarded. Finally, the non-redundant data set for training contained 279 positive sites and 7,004 negative sites. The training and the new data sets were available in Tables S1 and Table S2.

Algorithm

For SUMOsp 2.0, the basic two hypotheses of previous GPS (Group-based Phosphorylation Scoring) algorithm were reserved [1, 3, 4]. The first hypothesis is that if two short peptides share high sequence homology, they may also exhibit similar 3D structures and biological functions [3, 4]. Then we defined a *potential sumoylation peptide* $PSP(m, n)$ as a lysine (K) residue flanked by m residues upstream and n residues downstream. Here we used the amino acid substitution matrix BLOSUM62

as the initial matrix to calculate the similarity between two $PSP(m, n)$ peptides.

As previously described [1, 3, 4], for two amino acids a and b , let the substitution score between them in BLOSUM62 be $Score(a, b)$. The *similarity* between two $PSP(m, n)$ peptides A and B was defined as:

$$S(A, B) = \sum_{1 \leq i \leq (m+n)} Score(A[i], B[i])$$

If $S(A, B) < 0$, we simply redefined it as $S(A, B) = 0$.

The second and key hypothesis of GPS algorithm was that any given post-translational modification could recognize multiple sequence patterns/profiles/motifs for modification. Previously, in SUMOsp 1.0, the sumoylation sites were automatically clustered into three clusters (≤ 3) by different thresholds of peptides similarity [1]. Since it was time-consuming to find an optimal clustering result by exhaustively searching, we tried to obtain a near-optimized model and terminate the clustering process, if the prediction performance was not significantly increased ~~any more~~. In SUMOsp 2.0, the clustering process was much simpler. Based on experimental observations, we directly classified the known sumoylation sites into two clusters, including consensus (Type I) and non-consensus sites (Type II). The Type I sites should follow the $\Psi K X E$ (Ψ is A, I, L, M, P, F, or V, and X is any amino acid residue) [7-10], while Type II sites contained other non-canonical sites. Thus, the clustering procedure was based on experimental evidence rather than randomness.

In addition, we found that different amino acid matrices would generate different performances (published elsewhere). Thus, we suggested that these matrices might be optimized for sequence alignment or homologs searching, but not optimal for sumoylation sites prediction. In this regard, we developed a simple approach of Matrix Mutation (MaM) to further improve the prediction performance. First, the BLOSUM62 was chosen as the initial matrix, and the leave-one-out performance was calculated. Through exhaustive testing, we fixed the Sp at 85% to improve the Sn by randomly picking out an element of the matrix for mutation. The procedure was terminated when the Sn value could not be increased any more. The pseudocode of MaM was shown in Figure S1. From the result, both of the self-consistency and the leave-one-out validation were further improved (Table S3). We also chose three thresholds with high, medium and low stringencies for SUMOsp 2.0 (Table S3). The cut-off values were 0.17, 0.13, 0.1 and 3.33, 2.64, 2.26 for type I and type II sumoylation sites, respectively.

Collection of large-scale data sets of sumoylation proteomes in *S. cerevisiae* and *H. sapiens*

Currently, due to the great progress of Mass Spectrometry (MS) techniques in large-scale identifications, there were six and seven proteome-scale *in vivo* & *in vitro* experiments carried out in *S. cerevisiae* [11-16] and *H. sapiens* [17-23], respectively. As an application of SUMOsp 2.0 software, we compiled two integrated sumoylation proteome data sets from these large-scale experiments for *S. cerevisiae* and *H. sapiens*, respectively. The proteins sequences were retrieved from the Uniprot database according to the primary literature. And all data sets were manually curated before further analysis. The detailed information of sumoylation proteomes in *S. cerevisiae* and *H. sapiens* was available in Tables S4 and S5.

Software development and usage

In this work, the online service and local packages of SUMOsp 2.0 are implemented in JAVA and use a local CPU for computation. Thus, the calculating time is dependent on the users' computers. We tested the speed of SUMOsp 2.0 on numerous computers. Even on a laptop (e.g., IBM ThinkPad R51, 1.60GHz, 768MB), SUMOsp 2.0 can predict potential sumoylation sites for ~1000 proteins (average length ~1,000aa) within 3 minutes.

The SUMOsp 2.0 web server was tested on several internet browsers, including Internet Explorer 6.0, Netscape Browser 8.1.3 and Firefox 2 under Windows XP Operating System (OS), Mozilla Firefox 1.5 of Fedora Core 6 OS (Linux), and Safari 3.0 of Apple Mac OS X 10.4 (Tiger) and 10.5 (Leopard). For Windows and Linux systems, a latest version of Java Runtime Environment (JRE) package (JAVA 1.4.2 or later versions) of Sun Microsystems should be pre-installed for using the SUMOsp 2.0 program. However, for Mac OS, SUMOsp 2.0 could be directly visited without any additional packages. Furthermore, users could directly install the local packages of SUMOsp 2.0 on their own computers. Again, the local packages of SUMOsp 2.0 support three major Operating Systems (OS), including Windows, Unix/Linux and Mac.

For one protein, the sequence could be prepared either in raw or FASTA format. But for multiple proteins, the sequences should be prepared in FASTA format. Users could use "Ctrl+C & Ctrl+V" for windows and Linux, or "Command+C & Command+V" to copy and paste the protein sequences into the text box, and run the program by clicking the "Submit" button. Also, the prediction results could be copied and pasted into a new file, e.g., an Excel table, for further manipulation. The

protein sequence of human Tip60 was chosen (Figure S2) as an example for testing SUMOsp 2.0. Sumoylation of Tip60 plays an important role in the pathway of the p53-dependent DNA damage response [24]. With the high threshold, SUMOsp 2.0 predicted K72, K430 and K451 residues as highly potential sumoylation sites. Two of these sites (K430 and K451) had been verified *in vitro* and *in vivo* as real sumoylation sites in previous experiments [24].

Supporting Information – Supplementary Results

Evaluation of the prediction performance of SUMOsp 2.0

As previously described [1], four measurements: accuracy (Ac), sensitivity (Sn), specificity (Sp), and the Mathew correlation coefficient (MCC) were used to evaluate the prediction performance of SUMOsp 2.0. The four measurements can be calculated as follows:

$$Sn = \frac{TP}{TP + FN}, \quad Sp = \frac{TN}{TN + FP},$$

$$Ac = \frac{TP + TN}{TP + FP + TN + FN},$$

$$\text{and } MCC = \frac{(TP \times TN) - (FN \times FP)}{\sqrt{(TP + FN) \times (TN + FP) \times (TP + FP) \times (TN + FN)}}.$$

Here, the performances of self-consistency, leave-one-out validation and 4-, 6-, 8-, 10-fold cross-validations were performed on the training data set (279 positive sites and 7,004 negative sites) and the Receiver Operating Characteristic (ROC) curves were drawn (Figure S3). The results of leave-one-out validation and 4-, 6-, 8-, 10-fold cross-validations were very similar to those of the self-consistency performance. Thus, the validation results suggest that the SUMOsp 2.0 is a robust prediction system. The training data set is available in Table S1.

Also, we collected 53 experimentally verified sumoylation sites which were published after Feb, 2007. These sites were not included into the positive data for training as a new positive data set to test the performance of SUMOsp 2.0 (Table S2). Again, the non-sumoylated lysine residues in these 31 proteins were taken as a new negative data set (1,073 sites).

Comparison of SUMOsp 2.0 with other existing tools

In this work, we compared SUMOsp 2.0 to SUMOplot, SUMOsp 1.0 [1] and SUMOpre [25], with both the training data and the new data (Table S6). SUMOplot (www.abgent.com/doc/sumoplot) was the first web server for sumoylation site prediction. In the SUMOplot score system, a putative sumoylation peptide should follow the pattern of SUMO-CS (SUMO consensus sequence) or exhibit a similar hydrophobic profile with the consensus amino acid residues. To use SUMOplot, users could only input one protein sequence each time. Also, Liu *et al.* developed a predictor, which implemented

the Support Vector Machines (SVMs) algorithm, with the same data set of SUMOsp 1.0 [26]. However, the website is not available now. Additionally, Xu *et al.* used a statistical method to construct a high-profile program of SUMOpre [25].

For prediction of sumoylation sites, the number of negative sites were much greater than positive sites (>25:1 in the training data and >20:1 in the new data). Thus, a higher *Sp* value was more important than a higher *Sn* score in improving the final *Ac* and *MCC*. The performance of SUMOsp 2.0 with high threshold was better than SUMOplot with high threshold, with both of higher *Sn* and *Sp* values (Table S6). The overall accuracy (*Ac*) of SUMOsp 2.0 with low threshold was better than SUMOplot with low threshold and SUMOsp 1.0 with medium stringency (Table S6). Again, when the *Sn* value (83.15%) of SUMOsp 2.0 was chosen to be identical with SUMOsp 1.0 under high threshold, the *Sp* value of SUMOsp 2.0 was higher (95.94% vs. 92.70%). Thus, we propose that the prediction accuracy of SUMOsp 2.0 is better than that of SUMOplot and SUMOsp 1.0. We also compared SUMOsp 2.0 to SUMOpre, a recently developed web server [25]. The performance of SUMOpre with a cut-off value of 0.3 was slightly better than SUMOsp 2.0 (Table S6). However, as previously described, there were about 23% real sumoylation sites to be non-consensus sites (Type II sites, not follow a Ψ KXE pattern) [1]. Thus, if the *Sn* value was too low (eg., <77%), only consensus sites might be predicted, while all of the non-consensus sites would be missed. We then manually checked the prediction results from SUMOpre under a 0.3 cut-off value and found that only consensus sites were predicted. Since protein sumoylation could also occur on non-consensus sites, the stringency should be relaxed so that at least some non-consensus sites could also be correctly predicted. When the threshold of SUMOpre was relaxed to 0.2, the performance of SUMOsp 2.0 was much better (Table S6). Also, the high *Sn* value (82.80%) of SUMOsp 2.0 can also predict non-consensus sumoylation sites accurately. Taken together, the prediction accuracy of SUMOsp 2.0 is much better than our and others previous tools.

A large-scale prediction of sumoylation sites in S. cerevisiae

Although there were six high-throughput experiments carried out to identify potential sumoylated substrates in *S. cerevisiae* [11-16], a fraction of these proteins might not be real targets for SUMOylation, as the exact sumoylation sites in most of these proteins have still not been identified. In this work, we combined these data sets into an integrated data set, with 562 unique

potential sumoylation proteins (Table S3). All the primary sequences of these proteins were retrieved from the UniProt database. We then used SUMOsp 2.0 with high threshold to predict the potential sumoylation sites in these proteins directly. Totally, there were 403 proteins (~71%) with at least one sumoylation site predicted. Several real sumoylated substrates might be missed in our predictions; however, we proposed that if a protein was detected in more experiments, its probability of being a real sumoylated protein would be higher.

Previously, biochemical identification of sumoylation substrates by their sites was also a popular approach. However, in *S. cerevisiae*, only 37 sumoylation sites have been biochemically identified in twenty proteins (in our training and testing data set, Tables S1 and S2). Sixteen of these sumoylated proteins were also detected in large-scale experiments (Table S7). Both the experimentally verified and predicted sumoylation sites of these sixteen proteins are listed in Table S7 along with other potentially sumoylated proteins which have been identified in at least four experiments. Among 31 known sites in 16 substrates, SUMOsp 2.0 could correctly predict 24 (77%) of them as positive hits (Table S7). For example, there were three Septins, including Shs1[11-16], Cdc11 and Cdc3 [11-13, 15, 16] identified in at least five experiments (Table S7). Septins are a family of well conserved superclass of P-loop GTPases that function in cytokinesis in animals and fungi [27, 28]. Septins are known to have diverse cellular roles, including cytoskeletal reorganization, membrane dynamics, vesicle trafficking, polarity determination, and exocytosis. There are at least seven members of Septins in *S. cerevisiae* and fourteen in *H. sapiens*. In *S. cerevisiae*, septins are required for the formation of 10-nm filaments, with which they assemble the actomyosin contractile ring at the cleavage furrow, that does not appear to be conserved in other organisms [27, 28]. Previously, there were two sumoylation sites (K426, K437) experimentally identified in Shs1 and one (K412) in Cdc11 [29, 30]. Although four major sumoylation sites (K4, K11, K30, K63) have been identified in Cdc3 (Table S7), it was predicted to contain at least seven sites (K4, K11, K30, K63, K287, K443, K465) (Table S7). The prediction results of SUMOsp 2.0 for Shs1 and Cdc11 are consistent with the experimental observations. In regards to Cdc3, the SUMOsp 2.0 only missed one site (K11). Thus, we were satisfied with the accuracy of SUMOsp 2.0 was satisfying for these known substrates.

Our predictions also generated useful information for other unknown sumoylated proteins. For example, Cet1 (O13297) protein (Table S7), the beta subunit of mRNA capping enzyme in budding

yeast, was proposed as a potential sumoylation protein in five experiments [11, 13-16]. Cet1 is essential for yeast cell growth and viability, and it can form a heterodimer with Ceg1, the alpha component of mRNA capping enzyme, which was only identified in one experiment [15] (Table S3). However, it's not known whether Cet1 and Ceg1 could be sumoylated. The SUMOsp 2.0 predicted Cet1 and Ceg1 as positive hits with three (K136, K150, K196) and two potential sumoylation sites (K115, K249), respectively. Thus, our predictions implicated that sumoylation might be involved in mRNA processes, which was consistent with our previous work [20]. Most interestingly, a hypothetical protein, Ymr111c (Q04461) was identified as a potential sumoylated protein in four large-scale experiments [11-13, 16] (Table S7). The SUMOsp 2.0 predicted it as a positive hit with three highly potential sumoylation sites (K123, K130 and K231). These predictions will be useful for further experimental manipulation.

A large-scale prediction of sumoylation sites in H. sapiens

Due to the complexity and dynamics of the human proteome, identification of the sumoylation proteome in *H. sapiens* was much more difficult than in *S. cerevisiae*. From the large-scale data set, there were 354 proteins identified as potentially sumoylated proteins in *H. sapiens* (See Table S4). In our training and testing data set (Tables S1 and Table S2), there were 248 biochemically identified sumoylation sites in 146 unique human proteins. Although seven large-scale analyses were performed in human [17-23], only 16 (~11%) biochemically known sumoylated proteins were detected.

Again, we used SUMOsp 2.0 with high threshold to predict potential sumoylation sites in these proteins. We list potentially sumoylated proteins which were identified in at least two experiments in Table 4 along with previously known sumoylated proteins. Only one protein (PTBP1) was identified in four experiments but missed by SUMOsp 2.0 (Table S8). Only nine proteins were identified in three experiments, while six of them were predicted as positive hits with at least one site (Table S8). From our results, we proposed that many potential targets might not be sumoylated. For example, although SUMO-1 was identified in three experiments [17, 19, 22], it was not self-sumoylated (Table S8).

Our results provide several useful insights for further experiments. For example, heterogeneous nuclear ribonucleoprotein M/hnRNP-M (HNRPM, P52272) [19, 21, 22] and

L/hnRNP-L (HNRPL, P14866) [18, 19, 22] were identified in three experiments as potential sumoylation candidates (Table S8). Both hnRNP-L and hnRNP-M bind to a variety of pre-mRNAs of RNA polymerase II transcripts directly, and regulate their splicing, translocation, stability, and expression [20, 31]. For example, hnRNP-L associates with the intronic splicing enhancer (ISE) of the human endothelial nitric oxide synthase (eNOS) gene, and regulates its mRNA splicing [32]. Moreover, hnRNP-L can modulate the mRNA expression of human vascular endothelial growth factor (VEGF) by binding to VEGF mRNA under hypoxia [33]. However, hnRNP-M prefers to associate with poly(G) and poly(U) RNA homopolymers *in vitro* and plays important roles in 3'-end maturation of pre-mRNAs [31]. SUMOsp 2.0 predicted both proteins as positive hits, with five and one potential sumoylation sites in hnRNP-M (K17, K37, K120, K685 and K698) and hnRNP-L (K233), respectively (Table S8). These prediction results could be greatly helpful for research.

Supporting Information – Supplementary References

- [1] Xue, Y., Zhou, F., Fu, C., Xu, Y., Yao, X., SUMOsp: a web server for sumoylation site prediction. *Nucleic Acids Res* 2006, *34*, W254-257.
- [2] The Universal Protein Resource (UniProt) 2009. *Nucleic acids research* 2009, *37*, D169-174.
- [3] Xue, Y., Zhou, F., Zhu, M., Ahmed, K., *et al.*, GPS: a comprehensive www server for phosphorylation sites prediction. *Nucleic Acids Res* 2005, *33*, W184-187.
- [4] Zhou, F. F., Xue, Y., Chen, G. L., Yao, X., GPS: a novel group-based phosphorylation predicting and scoring method. *Biochem Biophys Res Commun* 2004, *325*, 1443-1448.
- [5] Li, W., Godzik, A., Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics (Oxford, England)* 2006, *22*, 1658-1659.
- [6] Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., *et al.*, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997, *25*, 3389-3402.
- [7] Geiss-Friedlander, R., Melchior, F., Concepts in sumoylation: a decade on. *Nature reviews* 2007, *8*, 947-956.
- [8] Rodriguez, M. S., Dargemont, C., Hay, R. T., SUMO-1 conjugation in vivo requires both a consensus modification motif and nuclear targeting. *J Biol Chem* 2001, *276*, 12654-12659.
- [9] Gill, G., SUMO and ubiquitin in the nucleus: different functions, similar mechanisms? *Genes Dev* 2004, *18*, 2046-2059.
- [10] Seeler, J. S., Dejean, A., Nuclear and unclear functions of SUMO. *Nat Rev Mol Cell Biol* 2003, *4*, 690-699.
- [11] Hannich, J. T., Lewis, A., Kroetz, M. B., Li, S. J., *et al.*, Defining the SUMO-modified proteome by multiple approaches in *Saccharomyces cerevisiae*. *J Biol Chem* 2005, *280*, 4102-4110.
- [12] Zhou, W., Ryan, J. J., Zhou, H., Global analyses of sumoylated proteins in *Saccharomyces cerevisiae*. Induction of protein sumoylation by cellular stresses. *J Biol Chem* 2004, *279*, 32262-32268.
- [13] Wohlschlegel, J. A., Johnson, E. S., Reed, S. I., Yates, J. R., 3rd, Global analysis of protein sumoylation in *Saccharomyces cerevisiae*. *J Biol Chem* 2004, *279*, 45662-45668.
- [14] Wykoff, D. D., O'Shea E, K., Identification of sumoylated proteins by systematic immunoprecipitation of the budding yeast proteome. *Mol Cell Proteomics* 2005, *4*, 73-83.

- [15] Panse, V. G., Hardeland, U., Werner, T., Kuster, B., Hurt, E., A proteome-wide approach identifies sumoylated substrate proteins in yeast. *J Biol Chem* 2004, 279, 41346-41351.
- [16] Denison, C., Rudner, A. D., Gerber, S. A., Bakalarski, C. E., *et al.*, A Proteomic Strategy for Gaining Insights into Protein Sumoylation in Yeast. *Mol Cell Proteomics* 2004.
- [17] Zhao, Y., Kwon, S. W., Anselmo, A., Kaur, K., White, M. A., Broad spectrum identification of cellular small ubiquitin-related modifier (SUMO) substrate proteins. *J Biol Chem* 2004, 279, 20999-21002.
- [18] Manza, L. L., Codreanu, S. G., Stamer, S. L., Smith, D. L., *et al.*, Global shifts in protein sumoylation in response to electrophile and oxidative stress. *Chem Res Toxicol* 2004, 17, 1706-1715.
- [19] Vertegaal, A. C., Ogg, S. C., Jaffray, E., Rodriguez, M. S., *et al.*, A proteomic study of SUMO-2 target proteins. *J Biol Chem* 2004, 279, 33791-33798.
- [20] Li, T., Evdokimov, E., Shen, R. F., Chao, C. C., *et al.*, Sumoylation of heterogeneous nuclear ribonucleoproteins, zinc finger proteins, and nuclear pore complex proteins: a proteomic analysis. *Proc Natl Acad Sci U S A* 2004, 101, 8551-8556.
- [21] Gocke, C. B., Yu, H., Kang, J., Systematic identification and analysis of mammalian small ubiquitin-like modifier substrates. *J Biol Chem* 2005, 280, 5004-5012.
- [22] Rosas-Acosta, G., Russell, W. K., Deyrieux, A., Russell, D. H., Wilson, V. G., A Universal Strategy for Proteomic Studies of SUMO and Other Ubiquitin-like Modifiers. *Mol Cell Proteomics* 2005, 4, 56-72.
- [23] Ganesan, A. K., Kho, Y., Kim, S. C., Chen, Y., *et al.*, Broad spectrum identification of SUMO substrates in melanoma cells. *Proteomics* 2007, 7, 2216-2221.
- [24] Cheng, Z., Ke, Y., Ding, X., Wang, F., *et al.*, Functional characterization of TIP60 sumoylation in UV-irradiated DNA damage response. *Oncogene* 2008, 27, 931-941.
- [25] Xu, J., He, Y., Qiang, B., Yuan, J., *et al.*, A novel method for high accuracy sumoylation site prediction from protein sequences. *BMC bioinformatics* 2008, 9, 8.
- [26] Liu, B., Li, S., Wang, Y., Lu, L., *et al.*, Predicting the protein SUMO modification sites based on Properties Sequential Forward Selection (PSFS). *Biochem Biophys Res Commun* 2007, 358, 136-139.
- [27] Finger, F. P., Reining in cytokinesis with a septin corral. *Bioessays* 2005, 27, 5-8.

- [28] Longtine, M. S., Bi, E., Regulation of septin organization and function in yeast. *Trends Cell Biol* 2003, 13, 403-409.
- [29] Johnson, E. S., Blobel, G., Cell cycle-regulated attachment of the ubiquitin-related protein SUMO to the yeast septins. *J Cell Biol* 1999, 147, 981-994.
- [30] Rallabhandi, P., Hashimoto, K., Mo, Y. Y., Beck, W. T., *et al.*, Sumoylation of topoisomerase I is involved in its partitioning between nucleoli and nucleoplasm and its clearing from nucleoli in response to camptothecin. *J Biol Chem* 2002, 277, 40020-40026.
- [31] Datar, K. V., Dreyfuss, G., Swanson, M. S., The human hnRNP M proteins: identification of a methionine/arginine-rich repeat motif in ribonucleoproteins. *Nucleic Acids Res* 1993, 21, 439-446.
- [32] Hui, J., Stangl, K., Lane, W. S., Bindereif, A., HnRNP L stimulates splicing of the eNOS gene by binding to variable-length CA repeats. *Nat Struct Biol* 2003, 10, 33-37.
- [33] Shih, S. C., Claffey, K. P., Regulation of human vascular endothelial growth factor mRNA stability in hypoxia by heterogeneous nuclear ribonucleoprotein L. *J Biol Chem* 1999, 274, 1359-1365.

Supporting Information – Supplementary Figures

Figure S1 – The pseudocode of Matrix Mutation (MaM). In this work, we fixed the Specificity (Sp) value as 85% to improve the Sensitivity (Sn) value by this approach.

The pseudocode of Matrix Mutation (MaM):

```
Fix the Sp as 85%
The BLOSUM62 matrix is chosen as the initialized matrix
Set the default mutation times to 0
While the mutation times less than 10000
    Randomly pick an element of the matrix for mutation
    Increase or decrease the value of the element
    Re-calculate the score of Leave-one-out with the mutated matrix
    If the score is increased
        Continue to mutate the element
    Else
        Give up the mutation
    Endif
    Score the number of mutation times
Endwhile
Return the mutated matrix
```

Figure S2 – The snapshot of SUMOsp 2.0 interface. As an example, the protein sequence of human Tip60 was chosen. With the high threshold, SUMOsp 2.0 predicted K72, K430 and K451 residues as positive hits. And two sites (K430 and K451) were *in vitro* and *in vivo* verified in our previous experiments [24].

Group-based Prediction System -- SUMOsp 2.0

File Tools About

Predicted Sites

Position	Peptide	Score	Cutoff	Type
72	QFPKKEA	0.521	0.17	Type: Ψ-K-X-E
430	MGLKSES	1.886	0.17	Type: Ψ-K-X-E
451	TSIKKED	4.351	0.17	Type: Ψ-K-X-E

Enter sequence(s) in FASTA format

>Example (Human Tip60 protein)
 MAEVGEIIEGCRLPVLRNRNQNEDWPLAEILSVKDISGRKLFVHYIDFNKRLDEWWTHERLDLKKIQFPKKEAKTPTKNGLPGRPGSPEREVPASAQ
 ASGKTLPIPVQITLRFNLPKEREAIPIGGEPDQPLSSSSCLQPNHRSTKRKVEWSPATPVPSETAPASVFPQNGAARRAVAAGPGRKRKSNCLGTDSDS
 QDSSDGIPSA PRMTGSLVSDRSHDDIVTRMKNIECIELGRHRLKPWFSPYPQELTLPVLYLCEFCCLKYGRSLKCLQRHLTKCDLRHPPGNEIYRKGT
 ISFFEIDGRKNKSYSQNLCLLAKCFLDHKTLYYDTPFLFYVMTEYDCKGFHIVGYFSKEKESTEDYNVACILTLPYQRRGYGKLLIEFSYELSKVEGKTG
 TPEKPLSDLGLLSYRSYWSQTILEILMGLKSESGERPQITINEISEITSIKKEDVISTLQYLNINYYKGQYILTLSEDIVDGHARAMLKRLLRIDSKCLHFTPK
 DWSKRGKW

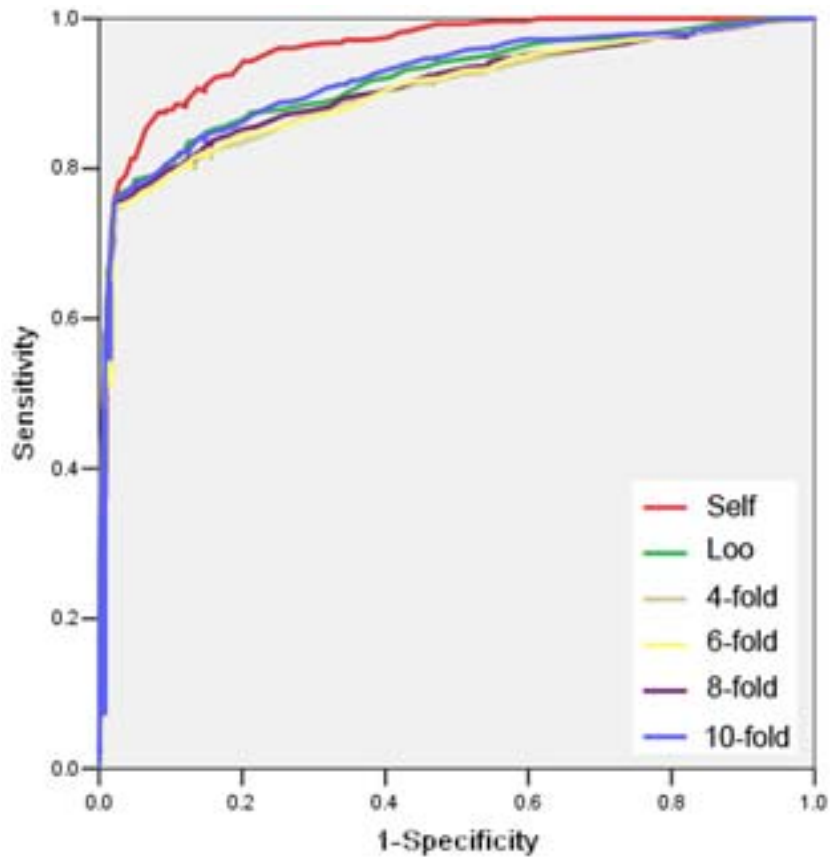
Threshold

High
 Medium
 Low
 All

Console

Example Clear Submit

Figure S3 – The Receiver Operating Characteristic (ROC) curves of the self-consistency (Self), the leave-one-out validation (Loo), and 4-, 6-, 8-, 10-fold cross-validations (4-fold, 6-fold, 8-fold and 10-fold).



Supporting Information – Supplementary Tables

Table S1 - The non-redundant training data set of SUMOsp 2.0. The data set contained 279 biochemically verified sumoylation sites from 166 distinct proteins. a. The UniProt Accession number for each protein.

UniProt ^a	Position	Organism	PMID
Q9SUI1	258	<i>A. thaliana</i>	11581165
O56136	84	<i>Adeno-associated virus</i>	15527853
O56136	447	<i>Adeno-associated virus</i>	15527853
P03116	514	<i>Bovine papillomavirus type 1</i>	11005821;12149243
P15330-2	382	<i>D. melanogaster</i>	11756545
P33244-2	418	<i>D. melanogaster</i>	15713642
Q9U1H5	438	<i>D. melanogaster</i>	15788563
P03206	12	<i>Epstein-Barr virus</i>	11160742;16112644
P03209	19	<i>Epstein-Barr virus</i>	15229220
P03209	213	<i>Epstein-Barr virus</i>	15229220
P03209	517	<i>Epstein-Barr virus</i>	15229220
O00180	274	<i>H. sapiens</i>	15820677
O00257	492	<i>H. sapiens</i>	17027752
O00327	259	<i>H. sapiens</i>	16109848
O00541	517	<i>H. sapiens</i>	11071894
O15169	857	<i>H. sapiens</i>	12223491
O15169	860	<i>H. sapiens</i>	12223491
O15525	14	<i>H. sapiens</i>	16738329
O60315	391	<i>H. sapiens</i>	16061479
O60315	866	<i>H. sapiens</i>	16061479
O60812	237	<i>H. sapiens</i>	15082759
O75030	289	<i>H. sapiens</i>	15507434;16029420
O75030	423	<i>H. sapiens</i>	15507434;16029420
O95600	67	<i>H. sapiens</i>	16617055
P00441	76	<i>H. sapiens</i>	16828461
P01100	265	<i>H. sapiens</i>	16055710
P04150	277	<i>H. sapiens</i>	12144530
P04150	293	<i>H. sapiens</i>	12144530
P04150	703	<i>H. sapiens</i>	12144530
P04637	386	<i>H. sapiens</i>	11867732
P05549	10	<i>H. sapiens</i>	12072434
P06400	720	<i>H. sapiens</i>	15806172
P06401	388	<i>H. sapiens</i>	12529333;12943706;17053081
P08235	89	<i>H. sapiens</i>	14500761

P08235	399	<i>H. sapiens</i>	14500761
P08235	428	<i>H. sapiens</i>	14500761
P08235	494	<i>H. sapiens</i>	17105732
P08235	953	<i>H. sapiens</i>	17105732
P10242	503	<i>H. sapiens</i>	12631292;16162816
P10242	527	<i>H. sapiens</i>	12631292;16162816
P10275	386	<i>H. sapiens</i>	12177000;15031320
P10275	520	<i>H. sapiens</i>	12177000;15031320
P10636	657	<i>H. sapiens</i>	16464864
P11387	103	<i>H. sapiens</i>	11709553
P11387	117	<i>H. sapiens</i>	11709553;12439742
P11387	153	<i>H. sapiens</i>	11709553;12439742
P11387	328	<i>H. sapiens</i>	11709553
P11387	436	<i>H. sapiens</i>	11709553
P11831	147	<i>H. sapiens</i>	12788062
P12757	50	<i>H. sapiens</i>	17202138
P12757	383	<i>H. sapiens</i>	17202138
P14921	15	<i>H. sapiens</i>	16319071;16862185
P15976	137	<i>H. sapiens</i>	15173587
P16220	285	<i>H. sapiens</i>	12552083
P16220	304	<i>H. sapiens</i>	12552083
P17676	174	<i>H. sapiens</i>	12810706
P18031	335	<i>H. sapiens</i>	17159996
P18031	347	<i>H. sapiens</i>	17159996
P19419	230	<i>H. sapiens</i>	14992729;15210726
P19419	249	<i>H. sapiens</i>	14992729;15210726
P19419	254	<i>H. sapiens</i>	15210726
P19544	73	<i>H. sapiens</i>	15520190
P19544	177	<i>H. sapiens</i>	15520190
P19793	108	<i>H. sapiens</i>	16912044
P20941	33	<i>H. sapiens</i>	16421094
P23497	297	<i>H. sapiens</i>	11792325
P23769	222	<i>H. sapiens</i>	12750312
P23769	389	<i>H. sapiens</i>	12750312
P25963	21	<i>H. sapiens</i>	14613580
P27540	245	<i>H. sapiens</i>	12354770
P28360	9	<i>H. sapiens</i>	16678795
P28360	127	<i>H. sapiens</i>	16678795
P29374	418	<i>H. sapiens</i>	16479010
P29374	444	<i>H. sapiens</i>	16479010
P29590	65	<i>H. sapiens</i>	9756909;12149243
P29590	160	<i>H. sapiens</i>	9756909;11413191;11704854;12149243
P29590	490	<i>H. sapiens</i>	9756909;12149243
P35520	211	<i>H. sapiens</i>	17087506

P35712	404	<i>H. sapiens</i>	16442531
P35712	417	<i>H. sapiens</i>	16442531
P36508	411	<i>H. sapiens</i>	15280358
P36956	123	<i>H. sapiens</i>	12615929
P36956	418	<i>H. sapiens</i>	12615929
P37231	107	<i>H. sapiens</i>	15229330;15507114
P37231	395	<i>H. sapiens</i>	16127449
P37840	96	<i>H. sapiens</i>	16464864
P37840	102	<i>H. sapiens</i>	16464864
P41161	89	<i>H. sapiens</i>	15857832
P41161	263	<i>H. sapiens</i>	15857832
P41161	293	<i>H. sapiens</i>	15857832
P41161	350	<i>H. sapiens</i>	15857832
P41212	99	<i>H. sapiens</i>	12626745;15107848
P41970	162	<i>H. sapiens</i>	15580297
P42224	703	<i>H. sapiens</i>	12855578;14596924;15761017
P42575	77	<i>H. sapiens</i>	15882978
P42858	6	<i>H. sapiens</i>	15064418
P42858	9	<i>H. sapiens</i>	15064418
P42858	15	<i>H. sapiens</i>	15064418
P43354	91	<i>H. sapiens</i>	14559918
P43354	577	<i>H. sapiens</i>	14559918
P43694	365	<i>H. sapiens</i>	15337742
P46060	524	<i>H. sapiens</i>	15355965
P49715	161	<i>H. sapiens</i>	12511558
P49716	120	<i>H. sapiens</i>	12161447
P54132	317	<i>H. sapiens</i>	15829507
P54132	331	<i>H. sapiens</i>	15829507
P54132	344	<i>H. sapiens</i>	15829507
P54132	347	<i>H. sapiens</i>	15829507
P54253	16	<i>H. sapiens</i>	15824120
P54253	194	<i>H. sapiens</i>	15824120
P54253	610	<i>H. sapiens</i>	15824120
P54253	697	<i>H. sapiens</i>	15824120
P54253	746	<i>H. sapiens</i>	15824120
P55055	410	<i>H. sapiens</i>	17218271
P55055	448	<i>H. sapiens</i>	17218271
P55265	418	<i>H. sapiens</i>	16120648
P55854	11	<i>H. sapiens</i>	12506199
P56524	559	<i>H. sapiens</i>	12032081;16166628
P56817	275	<i>H. sapiens</i>	12506199
P57682	10	<i>H. sapiens</i>	15684403
P57682	198	<i>H. sapiens</i>	15684403
P61086	14	<i>H. sapiens</i>	15723079

P63279	153	<i>H. sapiens</i>	15272016
P78347	221	<i>H. sapiens</i>	15016812
P78347	240	<i>H. sapiens</i>	15016812
P78347	456	<i>H. sapiens</i>	15016812
P78347	991	<i>H. sapiens</i>	15016812
Q00613	298	<i>H. sapiens</i>	11514557;12646186;12665592
Q01543	67	<i>H. sapiens</i>	16148010
Q02447	120	<i>H. sapiens</i>	12419227
Q02447	551	<i>H. sapiens</i>	15247228;15494207
Q03188	534	<i>H. sapiens</i>	15272016
Q03188	721	<i>H. sapiens</i>	15272016
Q03188	746	<i>H. sapiens</i>	15272016
Q03933	82	<i>H. sapiens</i>	11278381;15023536
Q05193	376	<i>H. sapiens</i>	15123615
Q05397	152	<i>H. sapiens</i>	14500712
Q06413	391	<i>H. sapiens</i>	16478538
Q07666	96	<i>H. sapiens</i>	16568089
Q08211	76	<i>H. sapiens</i>	15312759
Q08211	120	<i>H. sapiens</i>	15312759
Q09472	1020	<i>H. sapiens</i>	12718889;12887893;15632193
Q09472	1024	<i>H. sapiens</i>	12718889;12887893;15632193
Q12772	464	<i>H. sapiens</i>	12615929
Q13285	119	<i>H. sapiens</i>	15192080;15192092;15713642
Q13285	194	<i>H. sapiens</i>	15192080;15192092;15713642
Q13363	428	<i>H. sapiens</i>	12679040;12769861
Q13422	58	<i>H. sapiens</i>	15767674
Q13422	241	<i>H. sapiens</i>	15767674
Q13426	210	<i>H. sapiens</i>	16478998
Q13485	113	<i>H. sapiens</i>	12621041;12740389;15028714
Q13485	159	<i>H. sapiens</i>	12621041;12740389;14514699;15028714;15637079
Q13547	444	<i>H. sapiens</i>	11960997;12393750
Q13547	476	<i>H. sapiens</i>	11960997;12393750
Q13569	330	<i>H. sapiens</i>	11889051;15823533;15959518
Q14191	496	<i>H. sapiens</i>	15355988
Q15054	258	<i>H. sapiens</i>	16934752
Q15054	444	<i>H. sapiens</i>	16934752
Q15542	14	<i>H. sapiens</i>	15637059
Q15596	239	<i>H. sapiens</i>	12060666
Q15596	731	<i>H. sapiens</i>	12060666
Q15596	788	<i>H. sapiens</i>	12060666
Q15744	121	<i>H. sapiens</i>	15661739
Q15788	732	<i>H. sapiens</i>	12529333
Q15788	774	<i>H. sapiens</i>	12529333
Q16514	19	<i>H. sapiens</i>	15637059

Q16621	368	<i>H. sapiens</i>	16287851
Q16665	391	<i>H. sapiens</i>	15465032
Q16665	477	<i>H. sapiens</i>	15465032
Q16849	754	<i>H. sapiens</i>	16622421
Q5UBU5	571	<i>H. sapiens</i>	17101795
Q86YP4	30	<i>H. sapiens</i>	16738318
Q86YP4	487	<i>H. sapiens</i>	16738318
Q8N2W9	35	<i>H. sapiens</i>	15831457
Q8N4C6	1641	<i>H. sapiens</i>	16154161
Q8N4C6	1680	<i>H. sapiens</i>	16154161
Q8TBT7	38	<i>H. sapiens</i>	14972687
Q8WXI9	33	<i>H. sapiens</i>	16738318
Q92793	998	<i>H. sapiens</i>	16287980
Q92793	1033	<i>H. sapiens</i>	16287980
Q92793	1056	<i>H. sapiens</i>	16287980
Q969V6	499	<i>H. sapiens</i>	16098147
Q969V6	576	<i>H. sapiens</i>	16098147
Q969V6	624	<i>H. sapiens</i>	16098147
Q99497	130	<i>H. sapiens</i>	12761214;15976810
Q99607	657	<i>H. sapiens</i>	16904644
Q99683	535	<i>H. sapiens</i>	16142216
Q99683	1083	<i>H. sapiens</i>	16142216
Q99683	1114	<i>H. sapiens</i>	16142216
Q9BYV9	202	<i>H. sapiens</i>	15060166
Q9BYV9	276	<i>H. sapiens</i>	15060166
Q9BYV9	421	<i>H. sapiens</i>	15060166
Q9BYV9	580	<i>H. sapiens</i>	15060166
Q9H2X6	32	<i>H. sapiens</i>	15766567
Q9H2X6	1191	<i>H. sapiens</i>	12149243
Q9H3D4	588	<i>H. sapiens</i>	15539951
Q9H3D4	676	<i>H. sapiens</i>	15539951;15611636
Q9NSC2	1086	<i>H. sapiens</i>	12200128
Q9NZG4	429	<i>H. sapiens</i>	16479010
Q9NZG4	440	<i>H. sapiens</i>	16479010
Q9UBW7	963	<i>H. sapiens</i>	17027752
Q9UER7	630	<i>H. sapiens</i>	12150977
Q9UER7	631	<i>H. sapiens</i>	12150977
Q9UIS9	499	<i>H. sapiens</i>	17066076
Q9UIS9	538	<i>H. sapiens</i>	17066076
Q9UNR9	560	<i>H. sapiens</i>	14516784
Q9UPW6	233	<i>H. sapiens</i>	14701874
Q9UPW6	350	<i>H. sapiens</i>	14701874
Q9Y230	456	<i>H. sapiens</i>	16699503
Q9Y4L2	244	<i>H. sapiens</i>	15208321

Q9Y4L2	263	<i>H. sapiens</i>	15208321
Q9Y6K9	277	<i>H. sapiens</i>	14651848
Q9Y6K9	309	<i>H. sapiens</i>	14651848
Q9Y6Q9	723	<i>H. sapiens</i>	16760465
Q9Y6Q9	786	<i>H. sapiens</i>	16760465
Q9Y6Q9	1194	<i>H. sapiens</i>	16760465
P04591	475	<i>HIV-1</i>	15613319
P03243	104	<i>Human adenovirus 5</i>	11553772
P13202	450	<i>Human herpesvirus 5</i>	11602710
Q7M6S5	175	<i>Human herpesvirus 5</i>	10684265;11264375;12149243
Q7M6S5	180	<i>Human herpesvirus 5</i>	10684265;11264375;12149243
Q6XA64	802	<i>Human herpesvirus 6</i>	15105549
O92597	158	<i>Human herpesvirus 8</i>	16014952
P05627	229	<i>M. musculus</i>	10788439;11867732;16055710
P05627	257	<i>M. musculus</i>	16055710
P27782	25	<i>M. musculus</i>	11731474
P27782	267	<i>M. musculus</i>	11731474
P48432	247	<i>M. musculus</i>	17097055
Q60591	684	<i>M. musculus</i>	15117942
Q60591	897	<i>M. musculus</i>	15117942
Q60974	152	<i>M. musculus</i>	16421255
Q60974	260	<i>M. musculus</i>	16421255
Q60974	1117	<i>M. musculus</i>	16421255
Q60974	1330	<i>M. musculus</i>	16421255
Q64127	724	<i>M. musculus</i>	11313457
Q64127	742	<i>M. musculus</i>	11313457
P00445	19	<i>S. cerevisiae</i>	15596868
P00445	69	<i>S. cerevisiae</i>	15596868
P04456	61	<i>S. cerevisiae</i>	15542864
P05750	212	<i>S. cerevisiae</i>	15596868
P06778	11	<i>S. cerevisiae</i>	17013376
P06778	220	<i>S. cerevisiae</i>	17013376
P06786	1220	<i>S. cerevisiae</i>	12086615
P06786	1246	<i>S. cerevisiae</i>	12086615
P06786	1277	<i>S. cerevisiae</i>	12086615
P11978	1128	<i>S. cerevisiae</i>	15542864
P15873	127	<i>S. cerevisiae</i>	12226657;15121847
P15873	164	<i>S. cerevisiae</i>	12226657;15121847;15542864
P16649	270	<i>S. cerevisiae</i>	15596868
P18412	54	<i>S. cerevisiae</i>	16306045
P21538	807	<i>S. cerevisiae</i>	15542864
P32457	4	<i>S. cerevisiae</i>	12149243;12761287
P32457	11	<i>S. cerevisiae</i>	12149243;12761287
P32457	30	<i>S. cerevisiae</i>	12149243;12761287

P32457	63	<i>S. cerevisiae</i>	12149243;12761287;15542864
P32457	287	<i>S. cerevisiae</i>	12761287
P32457	443	<i>S. cerevisiae</i>	12761287
P32457	465	<i>S. cerevisiae</i>	12761287
P32458	412	<i>S. cerevisiae</i>	12149243
P32504	556	<i>S. cerevisiae</i>	16923829
P32504	651	<i>S. cerevisiae</i>	16923829
P32504	779	<i>S. cerevisiae</i>	16923829
P40460	231	<i>S. cerevisiae</i>	16923829
Q03330	25	<i>S. cerevisiae</i>	16411780
Q04322	498	<i>S. cerevisiae</i>	15542864
Q07657	426	<i>S. cerevisiae</i>	12149243
Q07657	437	<i>S. cerevisiae</i>	12149243
Q07979	322	<i>S. cerevisiae</i>	15596868
Q07979	328	<i>S. cerevisiae</i>	15596868
Q12216	438	<i>S. cerevisiae</i>	12761287
Q12216	446	<i>S. cerevisiae</i>	12761287
O42934	198	<i>S. pombe</i>	16168376
O60016	109	<i>S. pombe</i>	16168376
O60016	160	<i>S. pombe</i>	16168376
P40381	103	<i>S. pombe</i>	16168376
P21063	95	<i>Vaccinia virus</i>	15800065
Q90YL1	61	<i>X. laevis</i>	16256735
Q90YL1	253	<i>X. laevis</i>	16256735
Q90YL1	365	<i>X. laevis</i>	16256735

Table S2 – An additional data set for performance evaluation. To evaluate the prediction performance of SUMOsp 2.0, we also collected 53 sumoylation sites from 31 proteins which were published after Feb., 2007. These sites were not included in training as an independent data set.

UniProt	Position	Organism	PMID
Q9LSE2	393	<i>A. thaliana</i>	17416732
P18356	227	<i>Dengue virus type 2</i>	17265167
P18356	573	<i>Dengue virus type 2</i>	17265167
O95365	61	<i>H. sapiens</i>	17595526
P05455	41	<i>H. sapiens</i>	17646655
P06748	263	<i>H. sapiens</i>	17951246
P10914	275	<i>H. sapiens</i>	17942705
P10914	299	<i>H. sapiens</i>	17942705
P17544	118	<i>H. sapiens</i>	17264123
P22460	221	<i>H. sapiens</i>	17261810
P22460	536	<i>H. sapiens</i>	17261810
P25490	288	<i>H. sapiens</i>	17353273
P34896	38	<i>H. sapiens</i>	17446168
P43004	571	<i>H. sapiens</i>	17823119
P43268	96	<i>H. sapiens</i>	17585876
P43268	226	<i>H. sapiens</i>	17585876
P43268	260	<i>H. sapiens</i>	17585876
P43268	322	<i>H. sapiens</i>	17585876
Q05516	242	<i>H. sapiens</i>	17498654
Q05516	387	<i>H. sapiens</i>	17498654
Q05516	396	<i>H. sapiens</i>	17498654
Q13002	886	<i>H. sapiens</i>	17486098
Q13263	554	<i>H. sapiens</i>	17298944
Q13263	575	<i>H. sapiens</i>	17298944
Q13263	676	<i>H. sapiens</i>	17298944
Q13263	750	<i>H. sapiens</i>	17298944
Q13263	779	<i>H. sapiens</i>	17298944
Q13263	804	<i>H. sapiens</i>	17298944
Q14526	333	<i>H. sapiens</i>	17283066
Q2QGD7	660	<i>H. sapiens</i>	17696781
Q92797	361	<i>H. sapiens</i>	17923699
Q92797	483	<i>H. sapiens</i>	17923699
Q92993	430	<i>H. sapiens</i>	17704809
Q92993	451	<i>H. sapiens</i>	17704809
Q96EB6	734	<i>H. sapiens</i>	17934453
Q9NP62	156	<i>H. sapiens</i>	17646165
Q9UHF7	1192	<i>H. sapiens</i>	17391059

Q9UHF7	1201	<i>H. sapiens</i>	17391059
Q9UKF6	462	<i>H. sapiens</i>	17923699
Q9UKF6	465	<i>H. sapiens</i>	17923699
Q9UKF6	545	<i>H. sapiens</i>	17923699
Q9UQR1	115	<i>H. sapiens</i>	17940278
Q9UQR1	356	<i>H. sapiens</i>	17940278
Q9Y458	63	<i>H. sapiens</i>	17846996
Q9Y5Q3	32	<i>H. sapiens</i>	17548468
Q9Y5Q3	297	<i>H. sapiens</i>	17548468
P20263	118	<i>M. musculus</i>	17496161;17525163
P46099	56	<i>M. musculus</i>	17938210
P97447-2	144	<i>M. musculus</i>	17509614
P97447-2	171	<i>M. musculus</i>	17509614
Q61656	53	<i>M. musculus</i>	17369852
Q04110	5	<i>S. cerevisiae</i>	17888002
Q04110	101	<i>S. cerevisiae</i>	17888002

Table S3 - The performance of SUMOsp 2.0 was further improved by Matrix Mutation (MaM). Both of self-consistency and leave-one-out validation were performed on the training data. *a.* the performance before matrix mutation; *b.* the performance after matrix mutation.

SUMOsp 2.0	Threshold	Self-consistency				Leave-one-out			
		<i>Ac</i>	<i>Sn</i>	<i>Sp</i>	<i>MCC</i>	<i>Ac</i>	<i>Sn</i>	<i>Sp</i>	<i>MCC</i>
Before MaM^a	High	95.11%	80.65%	95.69%	0.5655	94.95%	76.34%	95.69%	0.5393
	Medium	89.43%	88.17%	89.48%	0.4362	89.14%	80.65%	89.48%	0.3975
	Low	82.68%	92.47%	82.29%	0.3550	82.41%	85.30%	82.29%	0.3226
After MaM^b	High	96.17%	80.65%	96.79%	0.6173	96.02%	76.70%	96.79%	0.5927
	Medium	92.52%	88.17%	92.69%	0.5083	92.23%	80.65%	92.69%	0.4667
	Low	85.81%	92.47%	85.55%	0.3947	85.55%	85.66%	85.55%	0.3624

Table S4 – A large-scale prediction of sumoylation sites in *S. cerevisiae*. There were six high-throughput experiments carried out to identify 562 potential sumoylated substrates (15-20), while 403 (~71%) of them were predicted with at least one sumoylation site. To date, there were only 37 sumoylation sites dissected in twenty sumoylated proteins from biochemical experiments (Tables S1 and S2), while sixteen of these targets were also detected in the large-scale experiments. The biochemically verified sumoylated proteins were shown in grey background. *a.* Exp., the number of experiments in which the protein was detected; *b.* Predicted sites, the number of predicted sites by SUMOsp 2.0.

Protein Name	ORF Name	UniProt	Denison <i>et al.</i>	Wohlschlegel <i>et al.</i>	Panse <i>et al.</i>	Zhou <i>et al.</i>	Wykoff <i>et al.</i>	Hannich <i>et al.</i>	Exp. ^a	Pre. Sites ^b
Shs1	YDL225W	Q07657	T	T	T	T	T	T	6	2
Cdc11	YJR076C	P32458	T	T	T	T	T	T	6	1
Tup1	YCR084C	P16649	T	T	T	T	T	T	6	1
Cdc3	YLR314C	P32457	T	T	T	T		T	5	6
Tfg1	YGR186W	P41895	T	T	T	T		T	5	5
Cet1	YPL228W	O13297	T	T	T		T	T	5	3
Pol30	YBR088C	P15873	T	T	T	T		T	5	3
Aos1	YPR180W	Q06624	T	T	T	T	T		5	1
Rsc58	YLR033W	Q07979	T	T		T	T	T	5	1
Rpc53	YDL150W	P25441	T	T		T		T	4	6
Rpa190	YOR341W	P10964	T		T	T		T	4	5
Stb3	YDR169C	Q12427	T	T	T			T	4	4
Toa1	YOR194C	P32773	T	T	T	T			4	4
Cdc48	YDL126C	P25694	T		T	T		T	4	3
Hsc82	YMR186W	P15108	T		T	T		T	4	3
Rsc8	YFR037C	P43609	T	T	T		T		4	3
Ymr111c	YMR111C	Q04461	T	T		T		T	4	3
Bop3	YNL042W	P53958	T	T	T			T	4	2

Eft2	YDR385W	P32324	T		T	T		T	4	2
Pgk1	YCR012W	P00560	T		T	T		T	4	2
Ssa1	YAL005C	P10591	T		T	T		T	4	2
Uba2	YDR390C	P52488	T	T	T	T			4	2
Cdc19	YAL038W	P00549	T		T	T		T	4	1
Rps3	YNL178W	P05750	T		T	T	T		4	1
Sod1	YJR104C	P00445	T			T	T	T	4	1
Tif2	YJL138C	P10081	T		T	T		T	4	1
Adh1	YOL086C	P00330	T		T	T		T	4	0
Fba1	YKL060C	P14540	T		T	T		T	4	0
Rhr2	YIL053W	P41277	T		T	T		T	4	0
Ubc9	YDL064W	P50623	T	T		T	T		4	0
Isw1	YBR245C	P38144	T	T				T	3	11
Spp41	YDR464W	P38904	T	T		T			3	11
Top1	YOL006C	P04786	T	T				T	3	10
Bir1	YJR089W	P47134	T	T		T			3	6
Pob3	YML069W	Q04636	T	T		T			3	6
Snf2	YOR290C	P22082	T	T				T	3	6
Top2	YNL088W	P06786	T	T				T	3	6
Fpr3	YML074C	P38911	T	T				T	3	5
Dbp9	YLR276C	Q06218	T	T				T	3	4
Gal11	YOL051W	P19659	T	T			T		3	4
Sth1	YIL126W	P32597	T	T			T		3	4
Ysh1	YLR277C	Q06224		T			T	T	3	4
Hpc2	YBR215W	Q01448	T	T				T	3	3
Hsp82	YPL240C	P02829	T		T	T			3	3
Pdc1	YLR044C	P06169	T		T			T	3	3
Rsc2	YLR357W	Q06488	T	T			T		3	3
Sko1	YNL167C	Q02100	T	T				T	3	3
Spt7	YBR081C	P35177	T	T			T		3	3

Taf8	YML114C	Q03750	T	T			T		3	3
Tef2	YBR118W	P02994			T	T		T	3	3
Fpr4	YLR449W	Q06205	T	T				T	3	2
Rap1	YNL216W	P11938	T	T				T	3	2
Rps8a	YBL072C	P05754	T			T		T	3	2
Smc5	YOL034W	Q08204	T	T				T	3	2
Spt5	YML010W	P27692	T	T		T			3	2
Hmo1	YDR174W	Q03973	T			T		T	3	1
Htb1	YDR224C	P02293	T	T		T			3	1
Rpa43	YOR340C	P46669	T	T		T			3	1
Rpc37	YKR025W	P36121	T	T	T				3	1
Sdc1	YDR469W	Q03323	T	T				T	3	1
Sfh1	YLR321C	Q06168	T	T			T		3	1
Abf1	YKL112W	P14164	T			T		T	3	0
Act1	YFL039C	P60010	T		T			T	3	0
Cpr1	YDR155C	P14832	T				T	T	3	0
Eno1	YGR254W	P00924	T			T		T	3	0
Gpm1	YKL152C	P00950	T			T		T	3	0
Mrp8	YKL142W	P35719	T	T		T			3	0
Npl6	YMR091C	P32832	T	T			T		3	0
Rpl28	YGL103W	P02406	T			T		T	3	0
Rxt2	YBR095C	P38255	T	T			T		3	0
Siz1	YDR409W	Q04195	T	T		T			3	0
Stm1	YLR150W	P39015		T		T		T	3	0
Taf14	YPL129W	P35189	T	T	T				3	0
Tdh3	YGR192C	P00359			T	T		T	3	0
Cbf5	YLR175W	P33322		T				T	2	18
Ebp2	YKL172W	P36049	T	T					2	12
Mlp2	YIL149C	P40457	T	T					2	10
Ngg1	YDR176W	P32494	T	T					2	9

Smc4	YLR086W	Q12267		T			T	2	9
Ils1	YBL076C	P09436	T			T		2	8
Vps72	YDR485C	Q03388	T	T				2	8
Nup2	YLR335W	P32499		T			T	2	7
Ris1	YOR191W	Q08562		T			T	2	7
Set1	YHR119W	P38827	T	T				2	7
Hap1	YLR256W	P12351	T	T				2	6
Rpa34	YJL148W	P47006	T	T				2	6
Sse1	YPL106C	P32589				T	T	2	6
Swi3	YJL176C	P32591	T	T				2	6
Tye7	YOR344C	P33122	T	T				2	6
Chd1	YER164W	P32657	T	T				2	5
Ntg1	YAL015C	P31378		T			T	2	5
Nto1	YPR031W	Q12311	T	T				2	5
Swc1	YAL011W	P31376	T	T				2	5
Swr1	YDR334W	Q05471	T	T				2	5
Cst6	YIL036W	P40535		T			T	2	4
Gcn5	YGR252W	Q03330	T	T				2	4
Orc3	YLL004W	P54790	T	T				2	4
Rsc1	YGR056W	P53236	T	T				2	4
Sum1	YDR310C	P46676	T	T				2	4
Trp5	YGL026C	P00931				T	T	2	4
Bdp1	YNL039W	P46678	T	T				2	3
Brf1	YGR246C	P29056	T	T				2	3
Gfa1	YKL104C	P14742		T		T		2	3
Gin4	YDR507C	Q12263	T	T				2	3
Hir2	YOR038C	P32480	T	T				2	3
Hms1	YOR032C	Q12398	T	T				2	3
Leo1	YOR123C	P38439		T		T		2	3
Met4	YNL103W	P32389	T	T				2	3

Nut1	YGL151W	P53114	T						T	2	3
Pgi1	YBR196C	P12709	T					T		2	3
Ret1	YOR207C	P22276		T	T					2	3
Rpc82	YPR190C	P32349		T	T					2	3
Rpl6a	YML073C	Q02326	T	T						2	3
Sir3	YLR442C	P06701	T	T						2	3
Sir4	YDR227W	P11978	T	T						2	3
Smc1	YFL008W	P32908	T	T						2	3
Tef1	YPR080W	P02994	T			T				2	3
Ulp1	YPL020C	Q02724	T	T						2	3
Yef3	YLR249W	P16521				T	T			2	3
Adk1	YDR226W	P07170		T					T	2	2
Asf2	YDL197C	P32448	T	T						2	2
Bdf1	YLR339C	O94085	T	T						2	2
Bud4	YJR092W	P47136	T	T						2	2
Cct3	YJL014W	P39077	T			T				2	2
Dep1	YAL013W	P31385	T	T						2	2
Grs1	YBR121C	P38088				T	T			2	2
Hda1	YNL021W	P53973				T			T	2	2
Ies1	YFL013C	P43579	T	T						2	2
Iki1	YHR187W	P38874	T					T		2	2
Leu1	YGL009C	P07264		T					T	2	2
Mhp1	YJL042W	P43638		T					T	2	2
Ncp1	YHR042W	P16603	T	T						2	2
Net1	YJL076W	P47035	T	T						2	2
Nfi1	YOR156C	Q12216	T					T		2	2
Reb1	YBR049C	P21538	T	T						2	2
Rpc40	YPR110C	P07703		T	T					2	2
Rpo21	YDL140C	P04050	T					T		2	2
Rps10b	YMR230W	P46784	T	T						2	2

Rsc4	YKR008W	Q02206	T	T			2	2
Shm2	YLR058C	P37291	T		T		2	2
Tof2	YKR010C	Q02208	T	T			2	2
Utp9	YHR196W	P38882	T	T			2	2
Arg1	YOL058W	P22768		T		T	2	1
Ash1	YKL185W	P34233	T	T			2	1
Asn2	YGR124W	P49090	T			T	2	1
Cbf1	YJR060W	P17106	T	T			2	1
Cin5	YOR028C	P40917	T	T			2	1
Clf1	YLR117C	Q12309	T	T			2	1
Cti6	YPL181W	Q08923	T	T			2	1
Fin1	YDR130C	Q03898	T	T			2	1
Gcn4	YEL009C	P03069	T	T			2	1
Gcr1	YPL075W	P07261	T	T			2	1
Gnd1	YHR183W	P38720	T		T		2	1
Gsy2	YLR258W	P27472	T			T	2	1
Hta1	YDR225W	P04911	T	T			2	1
Ies4	YOR189W	Q08561	T	T			2	1
Ipp1	YBR011C	P00817	T			T	2	1
Mes1	YGR264C	P00958		T	T		2	1
Msn1	YOL116W	P22148	T	T			2	1
Pab1	YER165W	P04147		T	T		2	1
Pho23	YNL097C	P50947	T	T			2	1
Pho8	YDR481C	P11491	T	T			2	1
Prp45	YAL032C	P28004	T	T			2	1
Psa1	YDL055C	P41940	T		T		2	1
Rpa135	YPR010C	P22138	T	T			2	1
Rpb4	YJL140W	P20433	T	T			2	1
Rpl3	YOR063W	P14126	T			T	2	1
Rpl4a	YBR031W	P10664	T			T	2	1

Rps0a	YGR214W	P32905				T		T	2	1
Rps1a	YLR441C	P33442	T			T			2	1
Rsa3	YLR221C	Q05942	T	T					2	1
Rtf1	YGL244W	P53064	T	T					2	1
Rvb1	YDR190C	Q03940	T	T					2	1
Scs2	YER120W	P40075				T	T		2	1
Snf1	YDR477W	P06782	T					T	2	1
Ssb1	YDL229W	P11484				T	T		2	1
Tfa2	YKR062W	P36145		T				T	2	1
Tpi1	YDR050C	P00942						T	2	1
Vid21	YDR359C	Q06337	T	T					2	1
Vps71	YML041C	Q03433	T	T					2	1
Yal027w	YAL027W	P39735	T	T					2	1
Yer064c	YER064C	P40041	T	T					2	1
Ylr455w	YLR455W	Q06188	T	T					2	1
Apa1	YCL050C	P16550		T	T				2	0
Arp9	YMR033W	Q05123	T	T					2	0
Bfr2	YDR299W	Q06631		T				T	2	0
Brn1	YBL097W	P38170	T	T					2	0
Cdc10	YCR002C	P25342				T	T		2	0
Erg13	YML126C	P54839				T		T	2	0
Faa1	YOR317W	P30624	T			T			2	0
Gcr2	YNL199C	Q01722	T	T					2	0
Ldb7	YBL006C	P38210	T	T					2	0
Mcd1	YDL003W	Q12158	T	T					2	0
Mmt2	YPL224C	Q08970	T					T	2	0
Nma1	YLR328W	Q06178	T					T	2	0
Rpl7a	YGL076C	P05737	T					T	2	0
Rpo26	YPR187W	P20435	T	T					2	0
Rps18a	YDR450W	P35271	T					T	2	0

Rsp5	YER125W	P39940	T	T			2	0
Sgf73	YGL066W	P53165	T	T			2	0
Swc4	YGR002C	P53201	T	T			2	0
Tsa1	YML028W	P34760	T			T	2	0
Uaf30	YOR295W	Q08747	T	T			2	0
Ydl025c	YDL025C	Q12100	T			T	2	0
Ygl157w	YGL157W	P53111		T	T		2	0
Mlp1	YKR095W	Q02455	T				1	22
Nop58	YOR310C	Q12499				T	1	16
Spa2	YLL021W	P23201				T	1	14
Tra1	YHR099W	P38811				T	1	10
Rpl2a	YFR031C	P38989	T				1	9
Esc1	YMR219W	Q03661				T	1	8
Ifh1	YLR223C	P39520		T			1	8
Abp1	YCR088W	P15891		T			1	7
Smc3	YJL074C	P47037	T				1	7
Snt1	YCR033W	P25357		T			1	7
Chc1	YGL206C	P22137		T			1	6
Ctr9	YOL145C	P89105		T			1	6
Nop7	YGR103W	P53261		T			1	6
Srs2	YJL092W	P12954				T	1	6
Ydl156w	YDL156W	Q12510		T			1	6
Bms1	YPL217C	Q08965		T			1	5
Bud3	YCL014W	P25558	T				1	5
Crz1	YNL027W	P53968				T	1	5
Gcd2	YGR083C	P12754			T		1	5
Ino80	YGL150C	P53115		T			1	5
Mot1	YPL082C	P32333			T		1	5
Prp22	YER013W	P24384		T			1	5
Taf7	YMR227C	Q05021		T			1	5

Tfc3	YAL001C	P34111		T				1	5
Tfg2	YGR005C	P41896			T			1	5
Bfr1	YOR198C	P38934		T				1	4
Cft2	YLR115W	Q12102			T			1	4
Cki1	YLR133W	P20485				T		1	4
Cyr1	YJL005W	P08678					T	1	4
Ded81	YHR019C	P38707			T			1	4
Elg1	YOR144C	Q12050					T	1	4
Epl1	YFL024C	P43572		T				1	4
His4	YCL030C	P00815			T			1	4
Hof1	YMR032W	Q05080	T					1	4
Hrp1	YOL123W	Q99383					T	1	4
Kar2	YJL034W	P16474			T			1	4
Mrd1	YPR112C	Q06106					T	1	4
Pxr1	YGR280C	P53335		T				1	4
Rco1	YMR075W	Q04779				T		1	4
Rpn1	YHR027C	P38764			T			1	4
Sap1	YER047C	P39955					T	1	4
Sli15	YBR156C	P38283		T				1	4
Svf1	YDR346C	Q05515		T				1	4
Ume1	YPL139C	Q03010			T			1	4
Zpr1	YGR211W	P53303					T	1	4
Ccr4	YAL021C	P31384				T		1	3
Cdc12	YHR107C	P32468	T					1	3
Cyc8	YBR112C	P14922			T			1	3
Esf1	YDR365C	Q06344		T				1	3
Hal5	YJL165C	P38970		T				1	3
Ilv2	YMR108W	P07342			T			1	3
Ist2	YBR086C	P38250					T	1	3
Krr1	YCL059C	P25586		T				1	3

Loc1	YFR001W	P43586		T			1	3
Mga2	YIR033W	P40578		T			1	3
Nop12	YOL041C	Q08208	T				1	3
Nop6	YDL213C	Q07623		T			1	3
Rom2	YLR371W	P51862		T			1	3
Rpl6B	YLR448W	P05739		T			1	3
Rrp6	YOR001W	Q12149		T			1	3
Sgs1	YMR190C	P35187		T			1	3
Sse2	YBR169C	P32590			T		1	3
Ste12	YHR084W	P13574		T			1	3
Sub2	YDL084W	Q07478				T	1	3
Taf13	YML098W	P11747			T		1	3
Ths1	YIL078W	P04801	T				1	3
Ubp3	YER151C	Q01477			T		1	3
Ubx4	YMR067C	P54730				T	1	3
Ufd1	YGR048W	P53044				T	1	3
Upc2	YDR213W	Q12151		T			1	3
Ybr159w	YBR159W	P38286			T		1	3
Ycf1	YDR135C	P39109		T			1	3
Yck2	YNL154C	P23292	T				1	3
Ycr016w	YCR016W	P25617				T	1	3
Ydr071c	YDR071C	Q12447				T	1	3
Ymr269w	YMR269W	Q03525		T			1	3
Ynl134c	YNL134C	P53912		T			1	3
Ynl311c	YNL311C	P42843			T		1	3
Yol109w	YOL109W	Q08245				T	1	3
Ypk1	YKL126W	P12688		T			1	3
Ade6	YGR061C	P38972				T	1	2
Ald4	YOR374W	P46367		T			1	2
Arp7	YPR034W	Q12406		T			1	2

Cbf2	YGR140W	P32504		T			1	2
Ccc1	YLR220W	P47818				T	1	2
Ceg1	YGL130W	Q01159			T		1	2
Dot6	YER088C	P40059				T	1	2
Erb1	YMR049C	Q04660		T			1	2
Erg1	YGR175C	P32476				T	1	2
Erg9	YHR190W	P29704			T		1	2
Gal7	YBR018C	P08431	T				1	2
Gcn1	YGL195W	P33892		T			1	2
Gcv2	YMR189W	P49095		T			1	2
Hsp42	YDR171W	Q12329				T	1	2
Iws1	YPR133C	Q06505		T			1	2
Ixr1	YKL032C	P33417				T	1	2
Mak16	YAL025C	P10962		T			1	2
Mcm21	YDR318W	Q06675		T			1	2
Nhp10	YDL002C	Q03435		T			1	2
Nop2	YNL061W	P40991		T			1	2
Pdr5	YOR153W	P33302		T			1	2
Pet127	YOR017W	P32606				T	1	2
Rad16	YBR114W	P31244		T			1	2
Rck2	YLR248W	P38623		T			1	2
Rfc1	YOR217W	P38630				T	1	2
Rgr1	YLR071C	P19263	T				1	2
Rnr2	YJL026W	P09938	T				1	2
Rpb2	YOR151C	P08518		T			1	2
Rpl16b	YNL069C	P26785	T				1	2
Rpl8A	YHL033C	P17076	T				1	2
Rpo31	YOR116C	P04051			T		1	2
Rvb2	YPL235W	Q12464		T			1	2
Sin3	YOL004W	P22579		T			1	2

Snf12	YNR023W	P53628		T			1	2
Snf5	YBR289W	P18480		T			1	2
Srl2	YLR082C	Q12020		T			1	2
Ssa2	YLL024C	P10592				T	1	2
Ssc1	YJR045C	P12398				T	1	2
Ssd1	YDR293C	P24276				T	1	2
Tal1	YLR354C	P15019	T				1	2
Tec1	YBR083W	P18412	T				1	2
Tef4	YKL081W	P36008	T				1	2
Tom70	YNL121C	P07213		T			1	2
Tri1	YMR233W	Q05024				T	1	2
Ubp15	YMR304W	P50101	T				1	2
Ybl054w	YBL054W	P34219				T	1	2
Ydr266c	YDR266C	Q05580		T			1	2
Ydr322w	YDR322W	Q06678				T	1	2
Yhr020w	YHR020W	P38708				T	1	2
Yjr011c	YJR011C	P47086				T	1	2
Ynl035c	YNL035C	P53962		T			1	2
Ypk2	YMR104C	P18961	T				1	2
Ypr143w	YPR143W	Q06511		T			1	2
Ada2	YDR448W	Q02336		T			1	1
Ade12	YNL220W	P80210				T	1	1
Adh3	YMR083W	P07246				T	1	1
Adh6	YMR318C	Q04894	T				1	1
Alf1	YNL148C	P53904				T	1	1
Ape3	YBR286W	P37302		T			1	1
Ara1	YBR149W	P38115		T			1	1
Aro2	YGL148W	P28777				T	1	1
Arp8	YOR141C	Q12386				T	1	1
Axl2	YIL140W	P38928				T	1	1

Bdh1	YAL060W	P39714		T				1	1
Bem2	YER155C	P39960					T	1	1
Brx1	YOL077C	Q08235		T				1	1
Bur6	YER159C	P40096		T				1	1
Cct8	YJL008C	P47079			T			1	1
Cdc123	YLR215C	Q05791		T				1	1
Cpa2	YJR109C	P03965			T			1	1
Cyc1	YJR048W	P00044	T					1	1
Dak1	YML070W	P54838		T				1	1
Dem1	YBR163W	P38289					T	1	1
Dsl1	YNL258C	P53847					T	1	1
Egd2	YHR193C	P38879					T	1	1
Erg20	YJL167W	P08524	T					1	1
Fir1	YER032W	P40020					T	1	1
Fmp12	YHL021C	P23180		T				1	1
Fob1	YDR110W	O13329		T				1	1
Frs1	YLR060W	P15624		T				1	1
Gal10	YBR019C	P04397	T					1	1
Gal2	YLR081W	P13181	T					1	1
Gas5	YOL030W	Q08193		T				1	1
Gcd11	YER025W	P32481			T			1	1
Gcs1	YDL226C	P35197			T			1	1
Gcy1	YOR120W	P14065	T					1	1
Gnd2	YGR256W	P53319			T			1	1
Gos1	YHL031C	P38736		T				1	1
Gpi16	YHR188C	P38875					T	1	1
Hcr1	YLR192C	Q05775					T	1	1
Hrk1	YOR267C	Q08732	T					1	1
Hsp60	YLR259C	P19882			T			1	1
Hta2	YBL003C	P04912		T				1	1

Ilv3	YJR016C	P39522			T			1	1
Lap3	YNL239W	Q01532	T					1	1
Lcb1	YMR296C	P25045			T			1	1
Leu2	YCL018W	P04173			T			1	1
Lys21	YDL131W	Q12122					T	1	1
Map1	YLR244C	Q01662		T				1	1
Med4	YOR174W	Q12343					T	1	1
Mis1	YBR084W	P09440		T				1	1
Ngr1	YBR212W	P32831					T	1	1
Nip1	YMR309C	P32497	T					1	1
Nog1	YPL093W	Q02892		T				1	1
Opy1	YBR129C	P38271		T				1	1
Oxa1	YER154W	P39952		T				1	1
Pdb1	YBR221C	P32473		T				1	1
Pgm2	YMR105C	P37012	T					1	1
Pho88	YBR106W	P38264					T	1	1
Pmi40	YER003C	P29952					T	1	1
Por1	YNL055C	P04840					T	1	1
Ppz1	YML016C	P26570					T	1	1
Pre10	YOR362C	P21242		T				1	1
Pre2	YPR103W	P30656		T				1	1
Pre9	YGR135W	P23638					T	1	1
Prp12	YMR302C	P32843					T	1	1
Pyc1	YGL062W	P11154	T					1	1
Rcs1	YGL071W	P22149	T					1	1
Rpa12	YJR063W	P32529		T				1	1
Rpa49	YNL248C	Q01080					T	1	1
Rpl13a	YDL082W	Q12690	T					1	1
Rpl18b	YNL301C	P07279					T	1	1
Rpl19b	YBL027W	P05735					T	1	1

Rpl29	YFR032C-A	P05747				T	1	1
Rpl35a	YDL191W	P39741		T			1	1
Rpl9b	YNL067W	P51401		T			1	1
Rpp2b	YDR382W	P02400	T				1	1
Rps10a	YOR293W	Q08745		T			1	1
Rps13	YDR064W	P05756				T	1	1
Rps9b	YBR189W	P05755				T	1	1
Rpt2	YDL007W	P40327		T			1	1
Rsc9	YML127W	Q03124				T	1	1
Sam1	YLR180W	P10659			T		1	1
Sbe22	YHR103W	P38814				T	1	1
Sec53	YFL045C	P07283			T		1	1
Spt15	YER148W	P13393		T			1	1
Spt8	YLR055C	P38915		T			1	1
Srb4	YER022W	P32569				T	1	1
Srp40	YKR092C	P32583	T				1	1
Ssb2	YNL209W	P40150			T		1	1
Ssp120	YLR250W	P39931				T	1	1
Sui2	YJR007W	P20459		T			1	1
Sup45	YBR143C	P12385		T			1	1
Sur2	YDR297W	P38992				T	1	1
Taf3	YPL011C	Q12297		T			1	1
Tdh1	YJL052W	P00360	T				1	1
Tif5	YPR041W	P38431		T			1	1
Trr1	YDR353W	P29509		T			1	1
Ura4	YLR420W	P20051		T			1	1
Wss1	YHR134W	P38838				T	1	1
Ybl005w-b	YBL005W-B	Q12490				T	1	1
Ydr336w	YDR336W	Q05473					1	1
Yfr006w	YFR006W	P43590		T			1	1

Yil110w	YIL110W	P40481	T				1	1
Ynr021w	YNR021W	P53723		T			1	1
Ypl158c	YPL158C	Q99299				T	1	1
Yra1	YDR381W	Q12159		T			1	1
Ade16	YLR028C	P54113		T			1	0
Adh2	YMR303C	P00331				T	1	0
Arc35	YNR035C	P53731		T			1	0
Arc40	YBR234C	P38328		T			1	0
Aro4	YBR249C	P32449				T	1	0
Arp4	YJL081C	P80428		T			1	0
Atf2	YGR177C	P53296				T	1	0
Bmh1	YER177W	P29311	T				1	0
Bna5	YLR231C	Q05979		T			1	0
Caf4	YKR036C	P36130				T	1	0
Car2	YLR438W	P07991	T				1	0
Cdc14	YFR028C	Q00684		T			1	0
Cks1	YBR135W	P20486		T			1	0
Cot1	YOR316C	P32798		T			1	0
Cpd1	YGR247W	P53314				T	1	0
Cwp1	YKL096W	P28319		T			1	0
Dig1	YPL049C	Q03063		T			1	0
Dpm1	YPR183W	P14020				T	1	0
Eno2	YHR174W	P00925				T	1	0
Erg10	YPL028W	P41338				T	1	0
Fmp35	YIL157C	P40452				T	1	0
Gal1	YBR020W	P04385	T				1	0
Gas3	YMR215W	Q03655		T			1	0
Gcd7	YLR291C	P32502				T	1	0
Gid7	YCL039W	P25569				T	1	0
Gpd1	YDL022W	Q00055				T	1	0

Gpd2	YOL059W	P41911				T	1	0
Gsp1	YLR293C	P32835	T				1	0
Gsp2	YOR185C	P32836	T				1	0
Gtt1	YIR038C	P40582	T				1	0
Has1	YMR290C	Q03532	T				1	0
Hem2	YGL040C	P05373	T				1	0
Hhf1	YBR009C	P02309	T				1	0
Hhf2	YNL030W	P02309	T				1	0
Hom2	YDR158W	P13663		T			1	0
Hor2	YER062C	P40106				T	1	0
Hsp26	YBR072W	P15992			T		1	0
Htb2	YBL002W	P02294	T				1	0
Hyp2	YEL034W	P23301	T				1	0
Kic1	YHR102W	P38692				T	1	0
Ktr1	YOR099W	P27810	T				1	0
Lhp1	YDL051W	P33399				T	1	0
Med7	YOL135C	Q08278				T	1	0
Met17	YLR303W	P06106	T				1	0
Mlh1	YMR167W	P38920				T	1	0
Ndc1	YML031W	P32500	T				1	0
Nhp2	YDL208W	P32495				T	1	0
Nit3	YLR351C	P49954				T	1	0
Nop1	YDL014W	P15646		T			1	0
Orm2	YLR350W	Q06144				T	1	0
Pbi2	YNL015W	P01095				T	1	0
Pbp1	YGR178C	P53297				T	1	0
Pex8	YGR077C	P53248				T	1	0
Pgu1	YJR153W	P47180				T	1	0
Pma1	YGL008C	P05030				T	1	0
Rep2	Rep2	P03872	T				1	0

Rox3	YBL093C	P25046			T		1	0
Rpd4b	YHR203C	P05753				T	1	0
Rpl15a	YLR029C	P05748				T	1	0
Rpl20b	YOR312C	P47913	T				1	0
Rpl21a	YBR191W	Q02753				T	1	0
Rpl25	YOL127W	P04456	T				1	0
Rpl27b	YDR471W	P38706				T	1	0
Rpl2a	YFR031C-A	P05736				T	1	0
Rpl30	YGL030W	P14120			T		1	0
Rpl31a	YDL075W	P04649		T			1	0
Rpl31b	YLR406C	P04649		T			1	0
Rpl32	YBL092W	P38061				T	1	0
Rps11a	YBR048W	P26781				T	1	0
Rps17b	YDR447C	P14127			T		1	0
Rps24a	YER074W	P26782				T	1	0
Rps25b	YLR333C	P07282			T		1	0
Rps4a	YJR145C	P05753	T				1	0
Rps5	YJR123W	P26783				T	1	0
Rps7b	YNL096C	P48164	T				1	0
Rsc6	YCR052W	P25632		T			1	0
Rtn1	YDR233C	Q04947				T	1	0
Sah1	YER043C	P39954	T				1	0
Sam4	YPL273W	Q08985		T			1	0
Sas10	YDL153C	Q12136		T			1	0
Sdh1	YKL148C	Q00711			T		1	0
Sec14	YMR079W	P24280		T			1	0
Sft2	YBL102W	P38166			T		1	0
Sis1	YNL007C	P25294		T			1	0
Skn7	YHR206W	P38889				T	1	0
Sna2	YDR525W-A	P56508			T		1	0

Snu13	YEL026W	P39990	T			1	0
Sro9	YCL037C	P25567			T	1	0
Srp72	YPL210C	P38688	T			1	0
Swi4	YER111C	P25302	T			1	0
Taf9	YMR236W	Q05027			T	1	0
Tbf1	YPL128C	Q02457	T			1	0
Tdh2	YJR009C	P00358	T			1	0
Tfc6	YDR362C	Q06339			T	1	0
Tfc7	YOR110W	Q12415	T			1	0
Tif34	YMR146C	P40217			T	1	0
Tkl1	YPR074C	P23254			T	1	0
Trm8	YDL201W	Q12009			T	1	0
Tub1	YML085C	P09733		T		1	0
Tub2	YFL037W	P02557		T		1	0
Tuf1	YOR187W	P02992		T		1	0
Ubc6	YER100W	P33296			T	1	0
Ubi4	YLL039C	P61864	T			1	0
Ura1	YKL216W	P28272	T			1	0
Ura7	YBL039C	P28274		T		1	0
Vma2	YBR127C	P16140		T		1	0
Vph1	YOR270C	P32563	T			1	0
Yap7	YOL028C	Q08182			T	1	0
Yar010c	YAR010C	O13528			T	1	0
Ybr238c	YBR238C	P38330			T	1	0
Ydr026c	YDR026C	Q12457	T			1	0
Ydr140w	YDR140W	Q03920			T	1	0
Yel007w	YEL007W	P40002			T	1	0
Yer010c	YER010C	P40011			T	1	0
Yil056w	YIL056W	P40522	T			1	0
Yju3	YKL094W	P28321	T			1	0

YII032c	YLL032C	Q07834		T	1	0
Ymr122w	YMR122W	Q3E842	T		1	0
Yor251c	YOR251C	Q08686		T	1	0
Yor352w	YOR352W	Q08816	T		1	0

Table S5 - A large-scale prediction of sumoylation sites in *H. sapiens*. There were seven large-scale experiments to identify 354 potential sumoylated targets, and 238 (~67%) of them were predicted with at least one site. To date there were 248 biochemically identified sumoylation sites identified in 146 unique human proteins (Tables S1 and S2), while only sixteen of them were also detected in the large-scale experiments. The biochemically verified sumoylated proteins were shown in grey background.

Protein Name	UniProt	Vertegaal <i>et al.</i>	Rosas-Acosta <i>et al.</i>	Zhao <i>et al.</i>	Manza <i>et al.</i>	Li <i>et al.</i>	Gocke <i>et al.</i>	Ganesan <i>et al.</i>	Exp.	Pre. sites
PTBP1	Q9BUQ0	T	T		T			T	4	0
HNRPM	P52272	T	T				T		3	5
TOP2B	Q02880				T	T	T		3	5
SF3A2	Q15428	T	T					T	3	2
HNRPF	P52597				T	T		T	3	1
HNRPL	P14866	T	T		T				3	1
SMT3H2	P61956	T	T	T					3	1
RPL15	P61313	T	T		T				3	0
RPL27A	P46776	T	T					T	3	0
SUMO1	P63165	T	T	T					3	0
CHD4	Q14839				T		T		2	15
SAFB2	Q14151	T	T						2	15
TOP1	P11387		T	T					2	13
PARP1	P09874		T				T		2	7
SART1	O43290	T					T		2	5
GTF2I	P78347			T			T		2	4
NOP5	Q9Y2X3			T	T				2	4
TMPO	P42167	T			T				2	4
DKFZp434D1319	Q9NSV0	T	T						2	3
ZNF281	Q9Y2X9	T	T						2	3
ATXN2L	Q8WWM7	T	T						2	2
KIAA1321	Q9P2M5	T	T						2	2

PML	P29590			T			T		2	2
RANGAP1	P46060			T			T		2	2
SFPQ	P23246			T			T		2	2
TKT	P29401			T			T		2	2
TRIM28	Q13263			T		T			2	2
BAT1	Q13838			T			T		2	1
CHERP	O00302	T		T					2	1
FASN	P49327			T			T		2	1
HNRPA1	P09651						T	T	2	1
KCTD1	Q719H9					T		T	2	1
NONO	Q15233	T		T					2	1
SNRP116	Q15029	T					T		2	1
ACTB	P60709	T		T					2	0
GPI	P06744			T			T		2	0
KHSRP	Q92945							T	2	0
RAN	P62826			T			T		2	0
RPL8	P25120	T		T					2	0
RPS18	P25232			T			T		2	0
SDF2L1	Q9HCN8	T		T					2	0
SF1	Q15637	T		T					2	0
YY1	P25490	T		T					2	0
MYH11	P35749						T		1	14
MDN1	Q9NU22						T		1	13
CASP8AP2	Q9NZV9							T	1	12
PLEC1	Q15149-1	T							1	12
COL16A1	Q07092							T	1	10
FLJ33887	Q8N260							T	1	10
MYH9	P35579						T		1	10
NCOR2	Q9Y618-1	T							1	10
OBSCN	Q5VST9							T	1	10

RANBP2	P49792			T				1	10
RBM25	P49756						T	1	9
JARID1A	P29375					T		1	8
KIAA0947	Q9Y2F5							T	1
LRPPRC	P42704				T			1	8
PRPF8	Q6P2Q9	T						1	8
TCERG1	O14776						T	1	8
DDX24	Q9GZR7						T	1	7
PRKDC	P78527-2				T			1	7
DKC1	O60832				T			1	6
DKFZp686J17211	Q68DK7						T	1	6
RLF	Q13129	T						1	6
WNK1	Q9H4A3						T	1	6
ZCWCC3	Q14149			T				1	6
ZNF291	Q9BY12					T		1	6
ADD3	Q9UEY8						T	1	5
ASH1L	Q9NR48	T						1	5
C20orf14	O94906					T		1	5
GTF2IRD2A	Q6EKI8							T	1
KIAA0284	Q9Y4F5							T	1
KIAA1596	Q9HCH6						T	1	5
LRP10	Q5KTZ5							T	1
NUP88	Q99567	T						1	5
SYMPK	Q92797						T	1	5
XRCC5	P13010						T	1	5
ZNF262	Q5VZL5				T			1	5
ASCC3L1	O75643	T						1	4
CTNNA1	P35221						T	1	4
DHX9	Q08211					T		1	4
DKFZp434A112	Q9H0E3						T	1	4

ETV1	P50549				T	1	4	
EXOSC10	Q01780			T		1	4	
HELZ	P42694	T				1	4	
HNRPU	Q9BQ09				T	1	4	
MAP3K1	Q13233				T	1	4	
PAICS	P22234			T		1	4	
PDIA4	P13667			T		1	4	
RP11-93N17.1	Q5T1K0				T	1	4	
VIL2	P15311				T	1	4	
WDR33	Q9C0J8	T				1	4	
ZFP106	Q9H2Y7	T				1	4	
ZFP91	Q96JP5		T			1	4	
ZNF255	Q8IZC8		T			1	4	
AOF2	O60341				T	1	3	
CALU	O43852				T	1	3	
CANX	P27824				T	1	3	
DDX42	Q86XP3		T			1	3	
DEK	P35659				T	1	3	
DHX15	O43143	T				1	3	
DKFZp586A0419	Q9Y3V8		T			1	3	
G22P1	Q6FG89				T	1	3	
GEMIN5	Q8TEQ6	T				1	3	
HISPPD1	Q8TB50					T	1	3
ILF3	Q12906		T			1	3	
KIAA0423	Q9Y4F4					T	1	3
KIAA0764	O94864				T	1	3	
MOZ2	Q8WYB5				T	1	3	
OXS1	O95747					T	1	3
PGRMC1	Q6IB11				T	1	3	
POLR3E	Q9NVU0			T		1	3	

PPP1R10	O00405		T				1	3
PSMD4	P55036				T		1	3
PTPRB	P23467					T	1	3
SF3B1	O75533				T		1	3
SF3b125	Q96BK1		T				1	3
SYDE2	Q5VT97					T	1	3
TOP2A	P11388				T		1	3
TPM3	Q5VU66				T		1	3
ZNF226	Q9NYT6					T	1	3
82-FIP	Q7Z417		T				1	2
ACTN4	O43707		T				1	2
AP1G1	Q8IY97				T		1	2
APP	P05067					T	1	2
C1QBP	Q07021				T		1	2
C20orf27	Q9GZN8					T	1	2
CFL1	P23528				T		1	2
COG1	Q8WTW3					T	1	2
CREB5	Q02930		T				1	2
DKFZp434P1215	Q9UIX2	T					1	2
FBS1	Q9HAH7		T				1	2
GATA2	P23769	T					1	2
HIVEP2	P31629					T	1	2
HSPCA	P07900		T				1	2
HSPCB	P08238		T				1	2
HYOU1	Q9Y4L1				T		1	2
KIAA1401	Q9P2E6		T				1	2
KRT8	P05787	T					1	2
LDHA	P00338				T		1	2
LMNA	P02545-1	T					1	2
LRPAP1	P30533		T				1	2

NPM1	P06748		T				1	2		
NUP98	P52948					T	1	2		
PKP1	Q13835						T	1	2	
PLOD1	Q02809	T						1	2	
POLR3A	Q8IW34			T				1	2	
RNASE3L	Q9NRR4-2	T						1	2	
RPL23A	P62750					T		1	2	
RPS6	P62753					T		1	2	
SATB1	Q01826						T	1	2	
SETBP1	Q9Y6X0							T	1	2
SMTN	Q8WWW1							T	1	2
SOX3	P41225							T	1	2
SPATS2L	Q9NUQ6							T	1	2
SREBF2	Q12772						T		1	2
SSRP1	Q08945							T	1	2
ST13	P50502					T			1	2
TSNAX	Q99598							T	1	2
XRCC1	P18887							T	1	2
ZBTB1	Q9Y2K1	T							1	2
ZNF221	Q9UK13							T	1	2
ZNF24	P17028							T	1	2
ZNF6	Q6NX42							T	1	2
ZNF7	P17097	T							1	2
ATP1A1	P05023							T	1	1
ATP5A1	P25705							T	1	1
ATP5B	P06576							T	1	1
C1orf35	Q8IZH3							T	1	1
CAP1	Q6I9U6							T	1	1
CBX4	O00257	T							1	1
COX4I1	Q86WV2							T	1	1

CRYZ	Q08257				T			1	1
DAN26	Q99492	T						1	1
DDX17	Q59F66				T			1	1
DDX5	P17844				T			1	1
DLD	P09622				T			1	1
E1B	P03243				T			1	1
EEF1B2	Q6IBH9				T			1	1
EIF5A	Q7L7L3				T			1	1
ERBB3	P21860			T				1	1
ETFA	P13804				T			1	1
FBL	P22087			T				1	1
FLJ10891	Q9NV72	T						1	1
FLJ11012	Q9NV06			T				1	1
FLJ13964	Q9H835	T						1	1
FLJ14893	Q96JY7	T						1	1
FOXC2	Q99958	T						1	1
FUS	P35637				T			1	1
FZD8	Q9H461						T	1	1
GCN5L2	Q92830					T		1	1
GNB2L1	P63244				T			1	1
GTPBP4	Q9BZE4	T						1	1
HEL308	Q8TDG4						T	1	1
HIRA	P54198					T		1	1
HIST2H2BE	Q5QNX0				T			1	1
HNRPA2B1	P22626				T			1	1
HNRPA3	P51991					T		1	1
HNRPR	Q43390				T			1	1
HSPA1A	P08107			T				1	1
ILF2	Q12905				T			1	1
KCTD15	Q9BVI6			T				1	1

KIAA0310	Q9BV84	T				1	1
KIAA0553	Q9UKJ3	T				1	1
KIF3C	O14782				T	1	1
KRT10	P13645	T				1	1
LDHB	P07195		T			1	1
LRCH1	Q9Y2L9				T	1	1
MEF2C	Q06413				T	1	1
MRPL27	Q9P0M9		T			1	1
MTPN	P58546			T		1	1
NFKB1	P19838				T	1	1
NUP107	P57740	T				1	1
PDCD5	O14737			T		1	1
PFKM	P08237				T	1	1
PGAM3	Q8N0Y7			T		1	1
PHB	P35232			T		1	1
PLCL2	Q9UPR0				T	1	1
PLOD3	O60568	T				1	1
POLR3H	Q9Y535-2			T		1	1
PPM1G	O15355			T		1	1
RBM7	Q9Y580	T				1	1
RCN1	Q15293			T		1	1
RDBP	P18615				T	1	1
RPL18A	Q02543		T			1	1
RPL3	P39023	T				1	1
RUVBL1	Q9Y265			T		1	1
SDF2	Q99470		T			1	1
SEMA6A	Q9H2E6				T	1	1
SENP1	Q9P0U3	T				1	1
SFRS2	Q01130			T		1	1
SLC25A1	P53007		T			1	1

SMT3A	P55854	T					1	1
SMT3H1	P55854		T				1	1
SOD1	P00441				T		1	1
SOD2	P04179					T	1	1
SPTAN1	Q14917				T		1	1
STMN2	Q93045					T	1	1
TMEM25	Q86YD3					T	1	1
U2AF2	P26368				T		1	1
UBE1	P22314		T				1	1
WDR18	Q9BV38	T					1	1
WDR69	Q8N136					T	1	1
XPO7	Q9UIA9				T		1	1
ZIC2	O95409		T				1	1
ZNF384	Q7Z722		T				1	1
ZNF409	Q9UPU6				T		1	1
ZNF43	P17038		T				1	1
ZNF431	Q8TF32		T				1	1
ZNF84	P51523		T				1	1
ZNF9	P62633				T		1	1
AAAS	Q9NRG9	T					1	0
ABCF2	Q9UG63	T					1	0
ACSL3	O95573				T		1	0
ACTC	P04270		T				1	0
ACTG1	Q9BTD2		T				1	0
ACTN2	Q86TI8		T				1	0
AGPAT5	Q9NUQ2					T	1	0
AK2	Q5TIF8				T		1	0
ALDH18A1	P54886				T		1	0
ALDOC	Q6FH94				T		1	0
ALPL	P05186	T					1	0

ANKRD41	A0JLW0				T	1	0
BTBD1	Q9H0C5				T	1	0
BTF3	P20290			T		1	0
CCT2	P78371			T		1	0
CCT4	Q53QP9			T		1	0
CIRH1A	Q969X6	T				1	0
CSPG4	Q6UVK1	T				1	0
CSTB	P04080			T		1	0
CYCS	P99999			T		1	0
D10S102	Q65ZQ3			T		1	0
DCPS	Q96C86				T	1	0
DKFZp761N202	Q8N395		T			1	0
DLX2	Q07687		T			1	0
DUSP11	O75319	T				1	0
UBC9	P63279			T		1	0
FLJ23109	Q9H5S6		T			1	0
FLJ36350	Q8N211		T			1	0
FLJ39023	Q8N8Q4		T			1	0
FLJ40479	Q8N7Q3		T			1	0
FLJ46644	Q6ZR52		T			1	0
FUBP1	Q96AE4				T	1	0
GTF2A1	P52655		T			1	0
H2AFJ	Q9NV63		T			1	0
H2afz	P17317		T			1	0
HBA1	P01922	T				1	0
HIST2H2AA	P20670		T			1	0
HIST3H3	Q6FGU4				T	1	0
HNRPD1	O14979				T	1	0
HNRPH1	P31943				T	1	0
HNRPH3	P31942				T	1	0

HNRPK	P61978				T		1	0
K-ALPHA-1	Q8WU19		T				1	0
KCNC3	Q14003					T	1	0
KLF7	O75840					T	1	0
KPNB1	Q14974				T		1	0
LOC65250	NP_075561					T	1	0
MCTS1	Q9ULC4					T	1	0
NUP43	Q8NFH3	T					1	0
OAT	P04181				T		1	0
PAFAH1B3	Q15102	T					1	0
PCBP2	Q15366				T		1	0
PFDN2	Q9UHV9				T		1	0
PHF6	Q8IWS0		T				1	0
PLS3	P13797				T		1	0
POLR1C	O15160-2				T		1	0
POLR2E	P19388				T		1	0
POLR3B	Q9NW08				T		1	0
POLR3C	Q9BUI4				T		1	0
POTE2	Q6S8J3		T				1	0
PPL18	Q07020				T		1	0
PRDM4	Q9UKN5		T				1	0
PRDX2	P32119				T		1	0
PTMA	Q8TBK9				T		1	0
PYCS	Q5T567		T				1	0
RAB11B	Q15907				T		1	0
RAB1B	Q9H0U4				T		1	0
RP5-858B16.1	Q5W1B5		T				1	0
RPS12	P25398				T		1	0
RPS27A	P62988		T				1	0
SF3B4	Q15427		T				1	0

SFN	Q96DH0		T			1	0	
SLC39A7	Q92504		T			1	0	
SND1	Q7KZF4				T	1	0	
SNRP70	P08621				T	1	0	
TAGLN2	Q6FGI1				T	1	0	
TBL3	Q96A78	T				1	0	
TIE1	P35590					T	1	0
TUBA3	Q71U36		T			1	0	
TUBA6	Q9BQE3		T			1	0	
TUBB2	P07437		T			1	0	
UBE2L3	P68036				T	1	0	
UBLE1A	Q9UBE0			T		1	0	
UBLE1B	Q9UBT2			T		1	0	
UQCRC1	P31930				T	1	0	
VDAC1	Q5FVE7				T	1	0	
YWHAQ	P27348				T	1	0	
YWHAZ	Q86V33		T			1	0	
ZNF100	Q8IYN0		T			1	0	
ZNF121	P58317		T			1	0	
ZNF141	Q15928		T			1	0	
ZNF15L1	P17019		T			1	0	
ZNF207	O43670		T			1	0	
ZNF208	O43345		T			1	0	
ZNF257	Q8NE34		T			1	0	
ZNF479	Q96JC4		T			1	0	
ZNF492	Q9P255		T			1	0	
ZNF512	Q96ME7		T			1	0	
ZNF539	Q86XL7		T			1	0	
ZNF588	Q9UII5		T			1	0	
ZNF629	Q9UEG4		T			1	0	

ZNF679	Q8IYX0	T	1	0
ZNF91	Q05481	T	1	0

Table S6 - Comparisons of SUMOsp 2.0 with SUMOplot, SUMOsp 1.0 and SUMOpre. *a.* The training data set contained 279 positive sites and 7,004 negative sites; *b.* The new data set contained 53 positive sites and 1,073 negative sites; *c.* We fixed the *Sn* value of SUMOsp 2.0 to be identical with SUMOsp 1.0 under high threshold, and compared their *Sp* values; *d.* We fixed the *Sn* value of SUMOsp 2.0 to be identical with SUMOpre under the cut-off of 0.3, and compared their *Sp* values.

Predictor	Threshold	Training data set ^a				New data set ^b			
		<i>Ac</i>	<i>Sn</i>	<i>Sp</i>	<i>MCC</i>	<i>Ac</i>	<i>Sn</i>	<i>Sp</i>	<i>MCC</i>
SUMOsp 2.0	High	96.09%	80.65%	96.70%	0.6128	96.45%	71.70%	97.67%	0.6392
SUMOplot	High	92.60%	80.29%	93.09%	0.4754	92.90%	67.92%	94.13%	0.4641
SUMOsp 2.0	Low	85.72%	92.47%	85.45%	0.3933	86.32%	75.47%	86.86%	0.3594
SUMOplot	Low	80.32%	90.32%	79.93%	0.3216	79.13%	83.02%	78.94%	0.3073
SUMOsp 1.0	Medium	79.84%	88.53%	79.50%	0.3098	80.28%	77.36%	80.43%	0.2941
SUMOsp 2.0^c		95.46%	83.15%	95.94%	0.5917	95.65%	67.92%	97.02%	0.5774
SUMOsp 1.0	High	92.34%	83.15%	92.70%	0.4811	92.72%	73.58%	93.66%	0.4857
SUMOsp 2.0^d		97.64%	64.16%	98.97%	0.6643	97.16%	54.72%	99.25%	0.6412
SUMOpre	0.3	97.83%	64.16%	99.17%	0.6851	97.25%	56.60%	99.25%	0.6551
SUMOsp 2.0		95.52%	82.80%	96.03%	0.5934	95.56%	66.04%	97.02%	0.5645
SUMOpre	0.2	92.90%	82.44%	93.32%	0.4940	93.43%	69.81%	94.59%	0.4908

Table S7 - For *S. cerevisiae*, all potentially sumoylated substrates detected in at least four large-scale experiments were listed. Also, all known sumoylated proteins detected in at least one experiment were shown. Among 31 biochemically identified sites in 16 substrates, SUMOsp 2.0 has correctly predicted 24 (~77%) of them as positive hits. *a.* UniProt, the UniProt accession number of the protein; *b.* *Exp.*, the number of experiments in which the protein was detected. *c.* Predicted sites, the biochemically verified sites were marked in red. *d.* The real sumoylation sites were missed by SUMOsp 2.0.

Name	UniProt ^a	Exp. ^b	Predicted sites ^c	Missed ^d
Shs1	Q07657	6	K426,K437	
Cdc11	P32458	6	K412	
Tup1	P16649	6	K270	
Cdc3	P32457	5	K4,K30,K63,K287,K443,K465	K11
Tfg1	P41895	5	K91,K497,K574,K658,K733	
Cet1	O13297	5	K136,K150,K196	
Pol30	P15873	5	K5,K127,K196	K164
Aos1	Q06624	5	K4	
Rsc58	Q07979	5	K328	K322
Rpc53	P25441	4	K51,K115,K216,K236,K322,K325	
Rpa190	P10964	4	K756,K1234,K1310,K1334,K1473	
Stb3	Q12427	4	K17,K162,K384,K512	
Toa1	P32773	4	K89,K149,K153,K205	
Cdc48	P25694	4	K496,K726,K740	
Hsc82	P15108	4	K69,K238,K257	
Rsc8	P43609	4	K41,K207,K221	
Ymr111c	Q04461	4	K123,K130,K231	
Bop3	P53958	4	K53,K199	
Eft2	P32324	4	K589,K829	
Pgk1	P00560	4	K142,K416	
Ssa1	P10591	4	K106,K523	
Uba2	P52488	4	K73,K229	
Cdc19	P00549	4	K240	
Rps3	P05750	4	K212	
Sod1	P00445	4	K19	K69
Tif2	P10081	4	K89	
Adh1	P00330	4		
Fba1	P14540	4		
Rhr2	P41277	4		
Ubc9	P50623	4		

Top2	P06786	3	K428,K716,K1136,K1220,K1246,K1277	
Gcn5	Q03330	2	K25,K76,K358,K380	
Sir4	P11978	2	K237,K654,K1037	K1128
Nfi1	Q12216	2	K438,K446	
Reb1	P21538	2	K718,K807	
Cbf2	P32504	1	K556,K651	K779
Tec1	P18412	1	K54,K99	
Rpl25	P04456	1		K61

Table S8 - For *H. sapiens*, all potentially sumoylated substrates detected in at least two large-scale experiments were listed. Also, all known sumoylated proteins detected in at least one experiment were shown. Among 33 biochemically verified sites in 16 substrates, SUMOsp 2.0 could correctly predict 26 (~79%) of them as positive hits. The biochemically detected sites were marked in red.

Name	UniProt	Exp.	Predicted sites	Missed
PTBP1	Q9BUQ0	4		
HNRPM	P52272	3	K17,K37,K120,K685,K698	
TOP2B	Q02880	3	K33,K34,K248,K1271,K1304	
SF3A2	Q15428	3	K103,K213	
HNRPF	P52597	3	K72	
HNRPL	P14866	3	K233	
SMT3H2	P61956	3	K11	
RPL15	P61313	3		
RPL27A	P46776	3		
SUMO1	P63165	3		
CHD4	Q14839	2	K72,K123,K618,K711,K838,K1212,K1228,K1304,K1565,K1572,K1620,K1663,K1664,K1669,K1766	
SAFB2	Q14151	2	K65,K94,K230,K284,K293,K385,K395,K484,K501,K517,K524,K528,K531,K541,K602	
TOP1	P11387	2	K103,K117,K153,K196,K197,K252,K321,K328,K354,K436,K654,K669,K700	
PARP1	P09874	2	K203,K233,K249,K486,K512,K761,K838	
SART1	O43290	2	K94,K141,K411,K709,K742	
GTF2I	P78347	2	K221,K240,K456,K991	
NOP5	Q9Y2X3	2	K390,K415,K467,K497	
TMPO	P42167	2	K17,K195,K198,K401	
RBM33	Q9NSV0	2	K215,K219,K237	
ZNF281	Q9Y2X9	2	K101,K128,K225	
ATXN2L	Q8WWM7	2	K573,K641	
KIAA1321	Q9P2M5	2	K74,K281	
PML	P29590	2	K160,K490	K65
RANGAP1	P46060	2	K481,K524	
SFPQ	P23246	2	K338,K349	
TKT	P29401	2	K352,K523	
TRIM28	Q13263	2	K554,K676	K575,K750,K779,K804
BAT1	Q13838	2	K53	
CHERP	O00302	2	K660	

FASN	P49327	2	K786	
HNRPA1	P09651	2	K183	
KCTD1	Q719H9	2	K252	
NONO	Q15233	2	K126	
SNRP116	Q15029	2	K448	
ACTB	P60709	2		
GPI	P06744	2		
KHSRP	Q92945	2		
RAN	P62826	2		
RPL8	P25120	2		
RPS18	P25232	2		
SDF2L1	Q9HCN8	2		
SF1	Q15637	2		
YY1	P25490	2		K288
SYMPK	Q92797	1	K361,K477,K483,K601,K1239	
DHX9	Q08211	1	K54,K76,K120,K365	
GATA2	P23769	1	K222,K389	
NPM1	P06748	1	K54,K263	
SREBF2	Q12772	1	K420,K464	
CBX4	O00257	1	K492	
MEF2C	Q06413	1	K391	
SMT3A	P55854	1	K11	
SOD1	P00441	1	K76	
UBC9	P63279	1		K153