SNOSite: Exploiting Maximal Dependence Decomposition to Identify Cysteine S-Nitrosylation with Substrate Site Specificity

Tzung-Yi Lee¹, Yi-Ju Chen², Tsung-Cheng Lu¹, Hsien-Da Huang³⁴, Yu-Ju Chen²

¹ Department of Computer Science and Engineering, Yuan Ze University, Chung-Li, Taiwan, ² Institute of Chemistry, Academia Sinica, Taipei, Taiwan, ³ Institute of Bioinformatics and Systems Biology, National Chiao Tung University, Hsin-Chu, Taiwan, ⁴ Department of Biological Science and Technology, Hsin-Chu, Taiwan

Abstract

S-nitrosylation, the covalent attachment of a nitric oxide to (NO) the sulfur atom of cysteine, is a selective and reversible protein post-translational modification (PTM) that regulates protein activity, localization, and stability. Despite its implication in the regulation of protein functions and cell signaling, the substrate specificity of cysteine S-nitrosylation remains unknown. Based on a total of 586 experimentally identified S-nitrosylation sites from SNAP/L-cysteine-stimulated mouse endothelial cells, this work presents an informatics investigation on S-nitrosylation sites including structural factors such as the flanking amino acids composition, the accessible surface area (ASA) and physicochemical properties, i.e. positive charge and side chain interaction parameter. Due to the difficulty to obtain the conserved motifs by conventional motif analysis, maximal dependence decomposition (MDD) has been applied to obtain statistically significant conserved motifs. Support vector machine (SVM) is applied to generate predictive model for each MDD-clustered motif. According to five-fold cross-validation, the MDD-clustered SVMs could achieve an accuracy of 0.902, and provides a promising performance in an independent test set. The effectiveness of the model was demonstrated on the correct identification of previously reported S-nitrosylation sites of Bos taurus dimethylarginine dimethylaminohydrolase 1 (DDAH1) and human hemoglobin subunit beta (HBB). Finally, the MDD-clustered model was adopted to construct an effective web-based tool, named SNOSite (http://csb.cse.yzu.edu.tw/SNOSite/), for identifying S-nitrosylation sites on the uncharacterized protein sequences.

Introduction

S-nitrosylation is a reversible post-translational modification (PTM) by covalent modification on the thiol group of cysteine (Cys) residues by nitric oxide (NO). Emerging evidences suggest that S-nitrosylation plays an important role in NO-related and redox pathway, especially in immune, cardiovascular, neuronal, and plant systems [1,2,3,4,5,6]. Moreover, different S-nitrosylation level and targets modulate the protein activity, localization, and stability [7,8,9] and further regulate the pathophysiological events, such neurodegenerative diseases and cancers [10,11,12]. Due to the labile nature and low abundance of endogenously S-nitrosylated proteins in vivo, the detail characteristics and mechanisms of S-nitrosylation still remain to be clarified. To our knowledge, the protein database of human, mouse, or rat possess only approximate 2% cysteine residues, however, not all cysteine residues on proteins can be S-nitrosylated by NO. Accumulating studies reveal that the cysteine residue, having low pKa or exposed thiol group on protein surface, is more accessible by NO modification [8,13].

With the increasing number of experimentally verified S-nitrosylation sites by proteomics advancement, several studies have revealed the S-nitrosylated cysteine residues may locate on acid-base motif, flanking with acidic (Aspartate or Glutamate) and basic (Arginine, Lysine, or Histidine) amino acids, or embed into the hydrophobic area [14,15,16,17,18,19]. Based on the structural analysis of S-nitrosylation on proteins, Marino et al. have revealed a modified acid-base motif, which is located more distantly to the cysteine and has its charged groups exposed [20]. However, whether other potential novel consensus S-nitrosylation motifs are present on proteins is not clear. The critical determinant of other structural component needs to be analyzed. Due to the labile nature of the S-NO bond and the low abundance of endogenously S-nitrosylated proteins in vivo, however, the unambiguous identification of S-nitrosylated proteins and S-nitrosylation sites remains challenging by commonly used proteomic technology [6,14,15,21,22,23,24]. From the structural point of view, thus, it is important to develop a method for the efficient and site-specific detection of protein S-nitrosylation, experimentally or computationally.

To date, approximately one thousand of proteins have been identified to related to S-nitrosylation in different biological systems [25], yet the specificity of S-nitrosylation sites are not completely understood. Using in silico prediction, GPS-SNO, has been proposed to computationally identify S-nitrosylation sites,
with a sensitivity of 53.57% and a specificity of 80.14% [26]. Recently, we have developed an S-alkylating biotin switch method and identified 586 S-nitrosylation sites corresponding to 384 S-nitrosylated proteins in SNAP/L-cysteine-stimulated mouse endothelial cells [19]. Using motif-X algorithm, 7 of 10 potential consensus motifs having local hydrophobicity at +2 position, containing acid-basic amino acids flanking with the central S-nitrosylating cysteine residues, were artificially extracted from ~30% S-nitrosylated peptides [19,27]. Considering that the majority of the S-nitrosylation sites did not match to the motif, other unknown structural factors must be taken into consideration. To further investigate potential S-nitrosylation motifs in primary amino acid sequence, the in silico characterization, i.e. amino acid composition, accessible surface area (ASA), and physicochemical properties, of protein S-nitrosylation sites is needed for distinguishing the S-nitrosylation sites from non-S-nitrosylation sites.

This work investigates site-specific characteristics for 586 experimentally verified S-nitrosylation sites [19] and applies maximal dependence decomposition (MDD) [28] to identify the potential substrate motifs of S-nitrosylation. With the application of MDD, a large group of aligned sequences can be moderated into subgroups that capture the most significant dependencies between positions. Support vector machine (SVM) is applied to generate the predictive model for each MDD-clustered subgroup. By further evaluation using five-fold cross-validation, the SVM models trained with MDD-clustered subgroups could improve predictive accuracy when compare to the model without the application of MDD clustering. Moreover, the experimental S-nitrosylation data from GPS-SNO (independent set) are used to test the effectiveness of the models that achieve the best accuracy in cross-validation. Finally, the models with MDD clustering method are adopted to implement an effective web-based tool, named SNOSite, for identifying cysteine S-nitrosylation sites. Two experimentally verified S-nitrosylated proteins, which were not included in training set, demonstrate the effectiveness of SNOSite. The in silico identification has potential for characterizing S-nitrosylation sites before experiments are performed.

Materials and Methods

Data preprocessing of training set and independent test set

With the high-throughput S-alkylating biotin switch method, a total of 586 S-nitrosylation sites corresponding to 384 S-nitrosylated proteins were experimentally identified in SNAP/L-cysteine-stimulated mouse endothelial cells for 30 minutes [19]. The experimental data on S-nitrosylated cysteines was determined using non-homologous training data. The positional weighted matrix (PWM) specified the relative frequency of amino acids that surround the S-nitrosylation sites, and was utilized in encoding the fragment sequences. A matrix of $m \times w$ elements was used to represent each residue of a training dataset, where $w$ stands for the window size and $m$ consists of 21 elements including 20 types of amino acids and one for terminal signal. In addition, WebLogo [32,33] is adopted to generate the graphical sequence logo for the relative frequency of the amino acids surrounding the S-nitrosylation sites.

Table 1. The statistics of experimentally verified S-nitrosylation sites in training set and independent test set.

| Data set               | Species     | Number of S-nitrosylated proteins | Number of S-nitrosylated cysteine | Number of non-S-nitrosylated cysteine |
|------------------------|-------------|-----------------------------------|-----------------------------------|---------------------------------------|
| Training set (Chen et al.) | Mouse       | 384                               | 586                               | 2,728                                 |
| Independent test set (GPS-SNO) | Multiple    | 327                               | 479                               | 2,501                                 |

Features investigation

Besides the composition of flanking amino acids (AA), the accessible surface area (ASA) and physicochemical properties around the S-nitrosylation sites were also investigated. Amino acids sequences with a S-nitrosylation site or cysteine in the center were extracted from positive and negative training sets, respectively, using a window of length $2w+1$ varying from four to ten. Different values of $w$ were used to determine the optimal window length. With reference the method of SulfoSite [31], the positional weighted matrix (PWM) of amino acids around the S-nitrosylated cysteines was determined using non-homologous training data. The positional weighted matrix (PWM) specified the relative frequency of amino acids that surround the S-nitrosylation sites, and was utilized in encoding the fragment sequences. A matrix of $m \times w$ elements was used to represent each residue of a training dataset, where $w$ stands for the window size and $m$ consists of 21 elements including 20 types of amino acids and one for terminal signal. In addition, WebLogo [32,33] is adopted to generate the graphical sequence logo for the relative frequency of the amino acids surrounding the S-nitrosylation sites.
corresponding amino acid at each position around the S-nitrosylation sites.

A side-chain of amino acid that undergoes post-translational modification prefers to be accessible on the surface of a protein [34]. Thus, the solvent-accessible surface area (ASA) was considered to evaluate the characteristics of S-nitrosylation sites. Since most of the experimental S-nitrosylated proteins do not have corresponding protein tertiary structures in PDB [35], an effective tool, RVP-Net [36,37], is applied to compute the ASA value from the protein sequence. RVP-net applied a neural network to predict the real ASA of residues based on information about their neighborhood, with a mean absolute error of 18.0–19.5%, defined as the absolute difference between the predicted and experimental values of relative ASA per residue [37]. The computed ASA is the percentage of the solvent-accessible area of each amino acid on the protein. The full-length protein sequences with experimentally identified S-nitrosylation sites are inputted to RVP-Net to compute the ASA value of all of the residues. The ASA values of amino acids around the S-nitrosylation sites are extracted and normalized to be between zero and one.

A previous work has utilized 31 informative physicochemical properties to identify protein ubiquitylation sites [38]. To investigate the characteristics of S-nitrosylation sites in comprehensive deal, 531 physicochemical properties, that were extracted from version 9.1 of AAindex [39], are evaluated the ability to distinguish the S-nitrosylation sites from the non-S-nitrosylation sites. AAindex [39] includes many published indices that specify the physicochemical properties of amino acids. Since each physicochemical property of the amino acids is specified by a set of 20 numerical values, the amino acids around the S-nitrosylation sites can be encoded according to the values associated with each physicochemical property. In order to identify the significant physicochemical properties, a measurement of F-score [40] has been applied to calculate a statistical value for each position surrounding S-nitrosylation sites. The F-score of the i th physicochemical feature is defined as:

\[
F \text{-score } (i) = \frac{(\overline{x}_i^{(+)} - \overline{x}_i)^2 + (\overline{x}_i^{(-)} - \overline{x}_i)^2}{\frac{1}{n^+ - 1} \sum_{k=1}^{n^+} (x_{ki}^{(+)} - \overline{x}_i^{(+)} )^2 + \frac{1}{n^- - 1} \sum_{k=1}^{n^-} (x_{ki}^{(-)} - \overline{x}_i^{(-)} )^2}
\]  

where \(\overline{x}_i^{(+)}\) and \(\overline{x}_i^{(-)}\) denote the average value of the \(i\)th feature in whole, positive, and negative data sets, respectively; \(n^+\) denotes the number of positive data set and \(n^-\) denotes the number of negative data set; \(x_{ki}^{(+)}\) denotes the \(i\)th feature of the \(k\)th positive instance, and \(x_{ki}^{(-)}\) denotes the \(i\)th feature of the \(k\)th negative instance [40].

Data clustering by maximal dependence decomposition

The aim of this study is to investigate the motifs of S-nitrosylation sites based on the amino acid sequence. Due to the difficulty of detecting the conserved motifs for the sequence data with a larger size, this work applies maximal dependence decomposition (MDD) [28] to cluster all sequences of S-nitrosylation site into subgroups, which have obvious motifs. MDD is a methodology to group a set of aligned signal sequences to moderate a large group into subgroups that capture the most significant dependencies between positions [41]. In previous study [28], MDD is firstly proposed to group the splice sites during the identification process of splice site prediction. However, in this work, we group protein sequences instead of nucleotides. MDD adopts chi-square test \(\chi^2(A_i,A_j)\) to evaluate the dependence of amino acid occurrence between two positions \(A_i\) and \(A_j\) that surround the S-nitrosylated cysteines. In order to extract the motifs that have conserved biochemical property of amino acids when doing MDD, we categorize the twenty types of amino acids into five groups such as aliphatic, polar and uncharged, acid, basic, and aromatic groups, as the grouping given in Table S1 [Supplementary Materials]. Then, a contingency table of the amino acids occurrence between two positions is constructed, as presented in Fig. 1. The chi-square test is defined as:

\[
\chi^2(A_i,A_j) = \sum_{m=1}^{n} \sum_{n=1}^{m} \frac{(X_{mn} - E_{mn})^2}{E_{mn}}
\]  

where \(X_{mn}\) represents the number of sequences that have the amino acids of group \(m\) in position \(A_i\) and have the amino acids of group \(n\) in position \(A_j\) for each pair \((A_i, A_j)\) with \(i \neq j\). \(E_{mn}\) is calculated as \(X_{mn} \times \frac{X_{Cn}}{X_{C2}}\), where \(X_{mn} = X_{m1} + \ldots + X_{m5}\), \(X_C = X_{1n} + \ldots + X_{5n}\), and \(X\) denotes the total number of sequences. If a strong dependence are detected (defined as a \(X^2\) value is larger than 34.3, corresponding to a cutoff level of \(P = 0.005\) with 16 degrees of freedom) between two positions, then proceed as described by Burge and Karlin [26]. After the detection of maximal dependence of flanking positions, as the example illustrated in Fig. 1, position \(-2\) has the maximal dependence with the occurrence of basic amino acids. Subsequently, all data can be divided into two subgroups: one has the occurrence of basic amino acids in position \(-2\) and the other does not have the occurrence of basic amino acids in position \(-2\). The MDD clustering is a recursively process to divide the positive sets into tree-like subgroups. When applying MDD to cluster the sequences of a positive set, a parameter, i.e., the maximum-cluster-size, should be set. If the size of a subgroup is less the maximum-cluster-size, the subgroup will not be divided any more. The MDD process terminates until all the subgroup sizes are less than the value of maximum-cluster-size.

Model learning and evaluation

The support vector machine (SVM) is applied to generate computational models that incorporate the encoded amino acids, accessible surface area and physicochemical properties. Based on binary classification, the concept of SVM is to map the input samples into a higher dimensional space using a kernel function, and then to find a hyper-plane that discriminates between the two classes with maximal margin and minimal error. A public SVM library, LibSVM [42], is used to train the predictive model with positive and negative training sets, which are encoded with reference to various training features. The radial basis function (RBF) \(K(S_i,S_j) = \exp(-\gamma ||S_i - S_j||^2)\) is selected as the kernel function of SVM. Cross-validation is important to the application of the predictor [43]. Predictive performance of the constructed models is evaluated by performing \(k\)-fold cross validation. The training data is divided into \(k\) groups by splitting each dataset into \(k\) approximately equal sized subgroups. During cross-validation, each subgroup is regarded as the validation set in turn, and the remainder is regarded as the training set. Next, the following measures of predictive performance of the trained models are defined. Precision (Pr) = \(TP/(TP+FP)\), Sensitivity (Sn) = \(TP/(TP+FN)\), Specificity (Sp) = \(TN/(TN+FP)\), Accuracy (Acc) = \((TP+TN)/(TP+FP+TN+FN)\), Balanced Accuracy (BAcc) = \((Sn+Sp)/2\), and Matthews Correlation Coefficient.
\[
(MCC) = \frac{(TP \times TN) - (FN \times FP)}{\sqrt{(TP + FN) \times (TN + FP) \times (TP + FP) \times (TN + FN)}},
\]
where TP, TN, FP and FN represent the numbers of true positives, true negatives, false positives and false negatives, respectively.

Additionally, the parameters of the predictive models, window length, cost, and gamma value of the SVM models are optimized to maximize predictive accuracy. Finally, the window size and features that yield the highest accuracy are employed to construct predictive models for independent test.

**Results**

Positively charged and higher solvent accessible amino acids neighboring with the S-nitrosylated cysteines

This study focuses on the sequence-based analysis of substrate specificity for S-nitrosylation. To preliminarily evaluate the amino acid frequency neighboring the S-nitrosylated cysteine, the non-homologous S-nitrosylated cysteine is centered on position 0, and the flanking amino acids (−10 to +10) are graphically visualized as sequence logos. With the frequency plot of sequence logo representation given in Figure 2, no significant amino acids having high frequency is surrounding to the S-nitrosylation sites. In order to further explore the difference of amino acid composition between positive data and negative data, we applied a web-based tool TwoSampleLogo [44], that detects and displays statistically significant differences in position-specific symbol compositions between two sets of multiple sequence alignments. Figure 3 presents the position-specific difference of amino acid compositions between S-nitrosylation sites (586 sequences) and non-S-nitrosylation sites (2728 sequences). It reveals that the most pronounced feature of S-nitrosylation sites is the abundance of charged amino acids, especially the positively charged Lysine (K), Arginine (R), and Histidine (H), at positions −2, −1, +2, +3, and +9. Another featured characteristic is the depletion of neural amino acids, such as L, V, P, M, C, and S, locating centrally around position −7. The results revealed that the distant amino acids in sequence, which may be close to S-nitrosylation cysteines in three-dimensional structure, have notable difference between S-nitrosylation sites and non-S-nitrosylation sites. Another interesting feature is the absence of positively charged residues at position +1 that is immediately adjacent to the S-nitrosylation sites. For instance, as shown in the lower pane of Figure 3, the K and R are depleted at position +1. In comparison with the 21 motifs (Table S2 in Supplementary Materials) detected by motif-X [27], the positively charged amino acids were also absent at position +1 in previous study. Moreover, the positively charged K at position −9, −6, −5, +9 and R and H at position −7, −6, and +2 were also present (Table S2). The result not only consisted with the sequence analysis by motif-X but also indicated that the positive charged amino acids surrounding the S-nitrosylated cysteines may play an important role for S-nitrosylation.

Besides composition of amino acids, we further analyze the correlation of solvent accessible surface area (ASA) and S-
nitrosylation sites. Because most of the experimentally verified S-nitrosylated proteins do not have corresponding protein tertiary structures in Protein Data Bank (PDB) [35], RVP-Net [36,37], an ASA prediction tool that has been demonstrated to provide accurate ASA values similar to those observed in the protein tertiary structure [29], was applied to compute the ASA value of each residue in the protein sequence. Figure 4 presented the comparison of average percentage of ASA in the 21-mer window ($-10 \sim +10$) between S-nitrosylation and non-S-nitrosylation sites. This analysis showed that the cysteine residues have the lowest ASA on both S-nitrosylated or non-S-nitrosylated cysteines, suggesting low preference of solvent accessibility in S-nitrosylation sites. Moreover, the adjacent amino acids neighboring the centered S-nitrosylation sites have relatively higher preference of solvent-accessible surface area than that of non-S-nitrosylation sites, especially in the region of upstream sequences ($-2 \sim -2$). In particular, the positions $-5$, $-6$ and $-7$ have more obvious difference, which are also the locations for positive amino acids shown in Figure 3. Interestingly, the average percentage of ASA is particularly low at positions ($-1$ and $+3$) that are adjacent to the S-nitrosylation sites, suggesting, again, that the adjacent amino acids may regulate the S-nitrosylation on cysteine residues due to relative surface solvent accessibility.

Cross-validation of characteristics for flanking amino acids and S-nitrosylation sites

To determine what window lengths and features can perform best to identify the S-nitrosylation sites, the predictive models are trained with various window lengths and various features and are evaluated using cross-validation. Based on the position-specific difference of amino acid compositions between S-nitrosylation sites and non-S-nitrosylation sites (Fig. 3), 21-mer ($-10 \sim +10$) is selected as the window length in the following evaluation and implementation. Herein, four types of feature including amino acid (AA), amino acid composition (AAC), accessible surface area (ASA), and 21 motifs – are evaluated. The feature of amino acids is encoded using a 20-dimensional vector and a positional weighted matrix, named “AA_20D” and “AA_PWM”, respectively. The features of accessible surface area (ASA) and motifs are encoded using the ASA values and 21-dimensional binary vector (Figure S1 in Supplementary Materials), respectively. According to the predictive accuracy given in Table 2, of the models trained with individual features, that trained with amino acid composition (AAC) slightly outperforms that trained with amino acid (AA_20D or AA_PWM), accessible surface area (ASA), or 21 motifs. However, the model trained with only ASA has the lowest predictive accuracy, which is probably caused by the low ASA value of cysteines.

Additionally, the predictive power of the model trained with the hybrid combination of AA, AAC, ASA, and 21 motifs is also evaluated. Amino acid (AA_PWM) is regarded as the basic feature for training a model with other features. As described previously, the feature of AAC yields the best accuracy of over 0.60. Therefore, as shown in Table 2, the model trained with a combination of AA_PWM and AAC perform best. The predictive sensitivity, specificity, accuracy, and Matthews Correlation Coefficient (MCC) of the best model are 0.640, 0.681, 0.675, and 0.245, respectively. However, the predictive power of the model trained with the combination of all features (AA_PWM, AAC, ASA and 21 motifs) is not better than that trained with AAC alone, presumably due to the features of ASA or 21 motifs performing not well for identifying S-nitrosylation sites.

To further analyze the physicochemical property of S-nitrosylation sites and adjacent amino acids, a total of 531 physicochemical properties, extracted from version 9.1 of AAindex [39], were individually explored [38]. Figure 5 shows the top
twenty physicochemical properties ranked by the average value of F-score measurement in 21-mer window \((-10\rightarrow +10)\). This investigation reveals that the twenty physicochemical properties contain high F-score values at positions \(-7, -4, +1, +2, +3,\) and \(+9\), which have statistically significant difference between S-nitrosylation sites and non-S-nitrosylation sites. The predictive power of the twenty physicochemical properties was evaluated. According to the cross-validation performance of the models individually trained with each of the twenty physicochemical properties, the feature of positive charge perform better sensitivity, specificity, accuracy, balanced accuracy, and MCC than other physicochemical properties, which achieves an accuracy of about 0.60 (Table 3). The result was consisted with the position-specific difference of amino acid composition (Fig. 2), which contains positively charged amino acids neighboring to the S-nitrosylation sites. However, the model trained with the top ranking physicochemical property (side chain interaction parameter) is not performing as accurately as that trained with positive charge. Most of the twenty physicochemical properties reach an accuracy of about 0.55 (Table 3). Based on a concept of the forward feature selection, the ranked physicochemical properties can be sequentially added into the best model (\(\text{AA_PWM} + \text{AAC}\)) to evaluate whether the integration of physicochemical properties could improve the predictive performance. After evaluating the forward selection of top twenty physicochemical properties, the predictive power is not improved, when comparing to the model trained with \(\text{AA_PWM}\) and \(\text{AAC}\) (Figure S2 in Supplementary Materials).

Exploring the potential S-nitrosylation motifs by MDD clustering

To improve the detection of the conserved motifs from large-scale S-nitrosylation data set, we further apply the maximal dependence decomposition (MDD) to cluster all 586 identified S-nitrosylated peptide sequences into 11 subgroups, capturing the most significant dependencies of amino acid composition between positions. Table S3 (Supplementary Materials) shows the number of positive data in each subgroup and their average performances of five-fold cross-validations. According to the chi-square test of the dependence of five amino acid groups in flanking positions (Table S1), 10 out of all MDD-clustered subgroups have the conserved motifs of positively charged amino acids (K, R and H) at a specific position. In particular, the first and fourth subgroups have the negatively charged amino acids (D and E) accompanied by positively charged amino acids on conserved motifs at two positions.

Table 2. The cross-validation performance of the models trained with various features.

| Training features | Sn   | Sp   | Pre  | Acc  | BAcc | MCC  |
|-------------------|------|------|------|------|------|------|
| Amino Acid (AA_20D) | 0.556 | 0.574 | 0.199 | 0.572 | 0.566 | 0.097 |
| Amino Acid (AA_PWM) | 0.585 | 0.586 | 0.212 | 0.587 | 0.586 | 0.127 |
| Amino Acid Composition (AAC) | 0.579 | 0.605 | 0.218 | 0.602 | 0.593 | 0.137 |
| Accessible Surface Area (ASA) | 0.540 | 0.553 | 0.187 | 0.552 | 0.547 | 0.069 |
| 21 Motifs | 0.556 | 0.563 | 0.195 | 0.562 | 0.560 | 0.088 |
| \(\text{AA_PWM}+\text{AAC}\) | 0.640 | 0.681 | 0.277 | 0.675 | 0.661 | 0.245 |
| \(\text{AA_PWM}+\text{ASA}\) | 0.561 | 0.583 | 0.204 | 0.580 | 0.573 | 0.108 |
| \(\text{AA_PWM}+21\text{Motifs}\) | 0.561 | 0.572 | 0.199 | 0.570 | 0.567 | 0.098 |
| \(\text{AA_PWM}+\text{AAC}+\text{ASA}\) | 0.578 | 0.603 | 0.217 | 0.599 | 0.591 | 0.134 |
| \(\text{AA_PWM}+\text{AAC}+21\text{Motifs}\) | 0.572 | 0.601 | 0.204 | 0.593 | 0.587 | 0.130 |
| \(\text{AA_PWM}+\text{AAC}+\text{ASA}+21\text{Motifs}\) | 0.588 | 0.589 | 0.214 | 0.589 | 0.589 | 0.131 |

Abbreviation: AA_20D, amino acids coding with 20-dimensional vector; AA_PWM, positional weighted matrix of flanking amino acids; ASA, accessible surface area; Pre, precision; Sn, sensitivity; Sp, specificity; Acc, accuracy; BAcc, balanced accuracy; MCC, Matthews Correlation Coefficient.

doi:10.1371/journal.pone.0021849.t002

Figure 4. Comparison of average percentage of ASA in the 21-mer window \((-10\rightarrow +10)\) between S-nitrosylation and non-S-nitrosylation sites.

doi:10.1371/journal.pone.0021849.g004
specific positions. Consistent with the previous study [16], trans-nitrosylation by an intermediary nitrosothiol (RSNO) is catalytically assisted by neighboring H and D that act as base and acid, respectively. However, the eleventh subgroup, that contains 68 S-nitrosylation sites, does not have a conserved motif.

Furthermore, all of 11 MDD-clustered subgroups are evaluated for their predictive power for identifying S-nitrosylation sites, based on five-fold cross-validation. The number of negative data in each subgroup is determined according to the ratio of non-S-nitrosylation sites (2728 sequences) to S-nitrosylation sites (586 sequences) in training data. To avoid the skew sampling of negative data, ten sets of negative data, which are randomly selected from all non-S-nitrosylation sites, are constructed for each subgroup. Thus, in each subgroup, the predictive model is trained with the combined features of positional weighted matrix (AA_PWM) and amino acid composition (AAC), and the five-fold cross-validation is implemented ten rounds. The average value of cross-validation performance in each subgroup is displayed in Table S3 (Supplementary Materials). It indicates that most of the 11 subgroups could achieve an average accuracy of about 0.900. Especially, the first, second, fourth and tenth subgroups perform with an average accuracy of about 0.950, but the eleventh subgroup has the worst performance with an accuracy of 0.845. In conclusion, the average accuracy of all 11 subgroups is 0.902 which increases 0.227 predictive accuracy comparing to the model trained without MDD clustering. This analysis indicates that the S-nitrosylated sequences in a large-scale data set can be alternatively clustered by MDD method, which significantly enhanced the signal of amino acids motif and improved the performance of the predictive model.

Evaluation of S-nitrosylation predictive models using independent test set

To evaluate effectiveness of the investigated features that achieve the best accuracy in cross-validation, an independent set is used to test the MDD-clustered models trained with positional weight matrix and amino acid composition. The independent test set is composed of the experimentally verified S-nitrosylation data of GPS-SNO [26] from multiple species, which contains a total of 479 positive data and 2501 negative data in 327 S-nitrosylated proteins. As shown in Table 4, the MDD-clustered models could perform with an accuracy of 0.627 in all independent test set. Due to the various motifs in multiple predictive models, for all data of independent testing, the estimated sensitivity (0.805) is higher than specificity (0.593). Furthermore, the predictive performance is estimated for various types of species. For E. coli, the MDD-clustered models have the highest accuracy (0.730) with the balanced sensitivity (0.702) and specificity (0.736). Reasonably, the MDD-clustered models that were trained with S-nitrosoproteome data set from SNAP/L-cysteine-stimulated mouse endothelial cells have a high sensitivity (0.830) for 106 mouse S-nitrosylation sites. The proposed method also has high sensitivity (0.819) for 105 rat S-nitrosylation sites. Overall, the independent testing demonstrates that the MDD-clustered models have higher estimated sensitivity comparing to specificity.
Implementation of web-based tool for identifying S-nitrosylation sites

With the time-consuming and laboratory-intensive experimental workflow, even though a protein can be S-nitrosylated, precise identification of the S-nitrosylation sites on the substrate is experimentally difficult. Therefore, an effective prediction tool should be developed to efficiently identify potential S-nitrosylation sites. Following evaluation by cross-validation and an independent test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-clustered models trained with positional weighted matrix of amino acids (AA_PWM) and amino acids composition test, the MDD-cla...
Substrate Motifs of S-Nitrosylation Sites

Case Study 1

| Swiss-Prot ID   | DDAH1_BOVIN                  |
|-----------------|------------------------------|
| Protein name    | N(G),N(G)-dimethylarginine dimethylaminohydrolase 1 |
| Synonyms        | DDAHI Dimethylargininase-1 |
| Gene name       | Name: DDAH1                  |

Nitrosylation Sites

| Position | Nitrosocysteine |
|----------|-----------------|
| 222      |                 |
| 274      |                 |

Paste a single sequence or several sequences in FASTA format into the field below:

>`DDAH1_BOVIN
MASLGHPATFGRATHYVVRALPESLAQQALRRRTKGeVDFARAERQHQLYGVGLGKLQLQVLPAI`

Submit a file (< 2MB) in FASTA format directly from your local disk: [Submit]

With Prediction Specificity: 95%

Result

| Protein Name | Locations | Nitrosylation Sites | The matched MDD-clustered Motifs |
|--------------|-----------|---------------------|-----------------------------------|
| DDAH1_BOVIN  | 178       | VVDALHLKSFCSMGPHLIAI|                                   |
| DDAH1_BOVIN  | 222       | LTVPDCTAANICYLNIPSKH|                                   |
| DDAH1_BOVIN  | 274       | ELKVDGLLTCSVLMNKVD |                                   |
| DDAH1_BOVIN  | 74        | QLPADDSLPCVFVEDAVVC|                                   |
the future investigation of substrate site specificity. Next, the second case study was performed on human hemoglobin subunit beta (HBB) which contains one S-nitrosylation site at position 94 [46]. The experimentally verified S-nitrosylation site at position 94 was correctly predicted by SNOSite (Figure S3 is Supplementary Materials). In addition, one more cysteine residue is predicted as a novel S-nitrosylation site. SNOSite can be accessed via a web interface, and is freely available to all interested users at http://csb.cse.yzu.edu.tw/SNOSite/.

Discussion

In this study, we reported a systematic informatics investigation on the S-nitrosylation substrate specificity from experimentally verified S-nitrosoproteome data [19]. The analysis of position-specific amino acids composition reveals that the most pronounced feature of S-nitrosylation sites is the abundance of positively charged amino acids at surrounding positions. This investigation also implicates that the distant amino acids in sequence (around position −7), which may be close to S-nitrosylation cysteine in three-dimensional structure, have notable difference between S-nitrosylation sites and non-S-nitrosylation sites. Additionally, the accessible surface area (ASA) and physicochemical properties are considered. Moreover, the S-nitrosylation sites have higher preference of solvent-accessible surface area, especially in the region of upstream sequences (−10−−2). Based on the F-score measurement of 531 physicochemical properties in 21-mer window (−10 to +10), twenty physicochemical properties are revealed that contain statistically significant difference at positions −7, −4, +1, +2, +5, and +9 of S-nitrosylation sites. According to the five-fold cross-validation, the model trained with the combined features of positional weighted matrix and amino acids composition gets the highest accuracy.

Due to the abundance of experimental data, this study focuses on investigating the motifs of S-nitrosylation sites based on the amino acid sequences. However, it is difficult to explore the conserved motifs from large-scale S-nitrosoproteome data set. Thus, this work applies maximal dependence decomposition (MDD) to cluster all sequences of S-nitrosylation site into 11 subgroups, which have obvious motifs. According to the chi-square test of the dependence in flanking positions, most of the MDD-clustered subgroups have the conserved motifs of positively charged amino acids (K, R and H) at a specific position. Particularly, two subgroups have the conserved motifs of positively charged and negatively charged amino acids at two specific positions. Although the newly identified motifs could not be experimentally verified, what has to be noticed is MDD clustering can help the biologist investigating the potential substrate motifs of S-nitrosylation sites. More noteworthy is the MDD-clustered motifs can be applied to improve the predictive power of computationally identifying S-nitrosylation sites with various substrate specificities. According to the evaluation of five-fold cross-validation, the models trained with 11 MDD-clustered motifs increases predictive accuracy of 0.227 comparing to the model trained without MDD clustering. This analysis indicates that the S-nitrosylated sequences with a larger size can be alternatively clustered by MDD method in order to enhance the signal of amino acids motif and improve the performance of the predictive model. The models trained with MDD-clustered subgroups in overall perform better than that without MDD clustering. Consequently, the models with MDD clustering method are applied to implement a novel web-based tool, named SNOSite, for identifying cysteine S-nitrosylation. Correct prediction on two experimentally verified S-nitrosylated proteins demonstrated the effectiveness of SNOSite.

Furthermore, the experimental S-nitrosylation data of GPS- SNO [26] from multiple species is regarded as independent sets and is used to test the effectiveness of the models that achieve the best accuracy in cross-validation. Independent testing indicates that the model trained with MDD-clustered motifs could perform robustly for the test data from human, mouse, rat, and E. coli species. Although the proposed method can perform accurately and robustly according to independent tests, some issues should still be addressed in future work. Firstly, the structural preferences of S-nitrosylation sites should be investigated in greater detail - especially for the data whose flanking amino acids are not conserved. In addition to the solvent accessible surface area, secondary structure, the B-factor, intrinsic disordered region, protein linker region, and other factors should be examined at experimental S-nitrosylation sites which are located in the protein regions with PDB entries. Secondly, the biological function of S-nitrosylated proteins needs to be studied. The analysis of Gene Ontology [47] or the network of protein-protein interaction may provide a clue for inferring the function of S-nitrosylated proteins. Finally, the independent testing indicates that the predictive model could not perform well in part of test data that is not homologous to the training data. The acquisition of additional experimentally verified S-nitrosylation data is needed to re-calibrate more accurate MDD-clustered motifs. The proposed method can be improved by considering the motifs that are intrinsically included in the test data.

Supporting Information

Figure S1 The coding scheme of 21 motifs for learning SVM classifier. A binary vector with 21 dimensions is used to denote what kind of motifs does a sequence has. (TIF)

Figure S2 The predictive accuracy of the best model (AA_PW-AAC) trained with forward selection of top twenty physico-chemical properties. (TIF)

Figure S3 A case study of human hemoglobin subunit beta (HBB) which contains one S-nitrosylation site at position 94. (TIF)

Table S1 The grouping of amino acids used in MDD clustering. (DOC)

Table S2 The 21 motifs of S-nitrosylation sites (586 sequences) identified by motif-x program with the parameters of motif occurrences and statistical significance are more than 10 and less than 0.01, respectively. (DOC)

Table S3 The 11 MDD-clustered subgroups and their average performances of five-fold cross-validations. (DOC)

Author Contributions

Conceived and designed the experiments: T-YL H-DH Y-JC. Performed the experiments: Y-JC. Analyzed the data: T-YL T-CL. Wrote the paper: T-YL Y-JC Y-JC. Responsible for the design, computational modeling, and web interface development: T-CL.
References

1. Bogdan C (2001) Nitric oxide and the immune response. Nat Immun 2: 907–916.
2. Karpuzoglu E, Ahmed SA (2006) estrogen regulation of nitric oxide and inducible nitric oxide synthase iNOS in immune cells: Implications for immunity, autoimmune diseases, and apoptosis. Nitric Oxide 15: 177–186.
3. Lindermayr C, Saalbach G, Durner J (2005) Proteomic identification of S-nitrosylated proteins in Arabidopsis thaliana. Plant Physiol 137: 921–930.
4. Gao N, Gao B, Dang D, Wang X, Song J, et al. (2009) S-nitrosylated proteins in respiratory biology. Am J Respir Crit Care Med 173: 1186–1193.
5. Stamler JS, Lamas S, Fang FC (2001) Nitrosylation, the prototypic redox-based signaling mechanism. FEMS Microbiol Lett 174: 257–265.
6. Jaffrey SR, Eoyang-Bromage H, Ferris CD, Tempst P, Snyder SH (2001) Protein S-nitrosylation: a physiological signal for neuronal nitric oxide. Nat Cell Biol 3: 193–197.
7. Nakamura T, Lipton S (2009) Cell death: protein misfolding and neurodegenerative diseases. Apoptosis 14: 455–468.
8. Hess DT, Matsuno A, Kim SO, Marshall HE, Stamler JS (2009) Protein S-nitrosylation: purview and parameters. Nat Rev Mol Cell Biol 10: 156–160.
9. Lam YW, Yuan Y, Isaac J, Bais CS, Miller J, et al. (2010) Comprehensive Identification and Modified-Site Mapping of S-Nitrosylated Targets in Prostate Epithelial Cells. PLoS One 5: e9075.
10. Yao D, Gu Z, Nakamura T, Shi Z-Q, Ma Y, et al. (2004) Nitrosative stress linked to sporadic Parkinson’s disease. S-nitrosylation of parkin regulates its E3 ubiquitin ligase activity. Proc Nat Acad Sci U S A 101: 10810–10814.
11. Cho D-H, Nakamura T, Fang J, Cieplak P, Godzik A, et al. (2009) S-Nitrosoglutathione targets. Proteomics 9: 4366–4375.
12. Hess DT, Matsuno A, Kim SO, Marshall HE, Stamler JS (2005) Protein S-nitrosylation: purview and parameters. Nat Rev Mol Cell Biol 6: 130–146.
13. Derakhshan B, Hao G, Gross SS (2007) Balancing reactivity against selectivity: the evolution of protein S-nitrosylation as an effector of cell signaling by nitric oxide. Cardiogenes 75: 210–219.
14. Hao G, Derakhshan B, Shi L, Camargo F, Gross SS (2006) SNOsID, a proteomic method for identification of cysteine S-nitrosylation sites in complex protein mixtures. Proc Natl Acad Sci U S A 103: 1012–1017.
15. Greco TM, Hodara R, Parapadakis I, Heijnen HJ, Demeney MK, et al. (2006) Identification of S-nitrosylation motifs by site-specific mapping of the S-nitrosocysteine proteome in human vascular smooth muscle cells. Proc Natl Acad Sci U S A 103: 7429–7435.
16. Lane P, Hao G, Gross SS (2001) S-nitrosylation is emerging as a specific and fundamental posttranslational protein modification: head-to-head comparison with O-phosphorylation. Sci STKE 2001: re1.
17. Stamler JS, Toune EJ, Lipton SA, Sacher N (1997) SNO signals: translocation, regulation, and a conserved motif. Neuron 18: 691–696.
18. Greco TM, Hodara R, Parapadakis I, Heijnen HJ, Demeney MK, et al. (2006) Identification of S-nitrosylation motifs by site-specific mapping of the S-nitrosocysteine proteome in human vascular smooth muscle cells. Proc Natl Acad Sci U S A 103: 7429–7435.
19. Chen YJ, Ku W-C, Lin P-Y, Chou H-C, Khoo K-H, et al. (2010) S-Alkylating sites. J Biol Chem 285: 3410–3416.
20. Jia L, Bonaventura C, Bonaventura J, Stamler JS (1996) S-nitrosohaemoglobin: a dynamic activity of blood involved in vascular control. Nature 380: 221–226.
21. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, et al. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 25: 23–29.
22. Karpuzoglu E, Ahmed SA (2006) Estrogen regulation of nitric oxide and respiratory biology. Am J Respir Crit Care Med 173: 1186–1193.
23. Romero-Puertas MC, Campostrini N, Matte A, Righetti PG, Perazzolli M, et al. (2009) Proteomic analysis and site-directed mutagenesis. J Proteome Res 8: 239–243.
24. Ahmad S, Gromiha MM, Sarai A (2003) RVP-net: online prediction of real valued accessible surface area of proteins from single sequences. Bioinformatics 19: 1849–1851.
25. Fürst A, Jakobsche LM, Rauh R, Riede P (2006) Two Sample Logo: a graphical representation of the differences between two sets of sequence alignments. Bioinformatics 22: 1536–1537.
26. Knipp M, Braun O, Geirng PM, Sack R, Vasak M (2003) Zn(II)-free dimethyldiguaninase-1 (DDAH-1) is inhibited upon specific Cys-S-nitrosylation. J Biol Chem 278: 3410–3416.
27. Tung CW, Ho SY (2008) Computational identification of ubiquitylation sites from protein sequences. BMC Bioinformatics 9: 510.
28. Kusasala S, Pokarowski P, Pokorski M, Kolinski A, Katayama T, et al. (2008) Aanide: amino acid index database, progress report 2008. Nucleic Acids Res 36: D202–203.
29. Liu C-J, Chen Y-W (2003) Combining SVMs with various feature selection strategies. NIPS 2003 feature selection challenge. pp 1–16.
30. Lee TY, Lin QZ, Hsieh SJ, Bretana NA, Lu CT (2011) Exploiting maximal dependence decomposition to identify conserved motifs from a group of aligned sequence signatures. Bioinformatics 27: 247–250.
31. Chen Y, Ischiropoulos H (2006) Identification of S-nitrosylation motifs by site-specific mapping of the S-nitrosocysteine proteome in mesangial cells. Mol Cell Proteomics 2: 156–163.
32. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, et al. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 25: 23–29.
33. Vacic V, Jakobsche LM, Riede P, Rauh R (2006) Two Sample Logo: a graphical representation of the differences between two sets of sequence alignments. Bioinformatics 22: 1536–1537.
34. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, et al. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 25: 23–29.
35. Karpuzoglu E, Ahmed SA (2006) Estrogen regulation of nitric oxide and respiratory biology. Am J Respir Crit Care Med 173: 1186–1193.
36. Ahmad S, Gromiha MM, Sarai A (2003) RVP-net: online prediction of real valued accessible surface area of proteins from single sequences. Bioinformatics 19: 1849–1851.
37. Fauchère JL, Charton M, Kier LB, Verloop A, Pliska V (1988) Amino acid side chain parameters for correlation studies in biology and pharmacology. Int J Pept Protein Res 36: D202–205.
38. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, et al. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 25: 23–29.
39. Karpuzoglu E, Ahmed SA (2006) Estrogen regulation of nitric oxide and respiratory biology. Am J Respir Crit Care Med 173: 1186–1193.
40. Fauchère JL, Charton M, Kier LB, Verloop A, Pliska V (1988) Amino acid side chain parameters for correlation studies in biology and pharmacology. Int J Pept Protein Res 36: D202–205.
41. Karpuzoglu E, Ahmed SA (2006) Estrogen regulation of nitric oxide and respiratory biology. Am J Respir Crit Care Med 173: 1186–1193.
42. Fauchère JL, Charton M, Kier LB, Verloop A, Pliska V (1988) Amino acid side chain parameters for correlation studies in biology and pharmacology. Int J Pept Protein Res 36: D202–205.
43. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, et al. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 25: 23–29.
44. Fauchère JL, Charton M, Kier LB, Verloop A, Pliska V (1988) Amino acid side chain parameters for correlation studies in biology and pharmacology. Int J Pept Protein Res 36: D202–205.
45. Karpuzoglu E, Ahmed SA (2006) Estrogen regulation of nitric oxide and respiratory biology. Am J Respir Crit Care Med 173: 1186–1193.
46. Fauchère JL, Charton M, Kier LB, Verloop A, Pliska V (1988) Amino acid side chain parameters for correlation studies in biology and pharmacology. Int J Pept Protein Res 36: D202–205.