How the artificial intelligence tool iSNO-PseAAC is working in predicting the cysteine S-nitrosylation sites in proteins

Kuo-Chen Chou*
Gordon Life Science Institute, Boston, Massachusetts 02478, United States of America

In 2013 a very powerful AI (artificial intelligence) tool has been established for identifying cysteine S-nitrosylation sites in proteins, which is one of the important post modifications in proteins [1].

To see how the web-server is working, please do the following.

Step 1. Open the web server at http://app.aporc.org/iSNO-PseAAC/ and you will see the top page of the predictor on your computer screen, as show in Figure 1. Click on the Read Me button to see a brief introduction about iSNO-PseAAC predictor and the caveat when using it.

Step 2. Either type or copy/paste the query protein sequences into the input box shown at the center of Figure 1. The input sequence should be in the FASTA format. A sequence in FASTA format consists of a single initial line beginning with a greater-than symbol (">") in the first column, followed by lines of sequence data. The words right after the ">" symbol in the single initial line are optional and only used for the purpose of identification and description. All lines should be no longer than 120 characters and usually do not exceed 80 characters. The sequence ends if another line starting with a ">" appears; this indicates the start of another sequence. Example sequences in FASTA format can be seen by clicking on the Example button right above the input box.

Step 3. Click on the Submit button to see the predicted result. For example, if you use the query protein sequences in the Example window as the input, after clicking the Submit button, you will see on your screen the predicted SNO site positions and the corresponding segments as formulated by Equation 1. All these results are fully consistent with the experimentally verified results. It takes about a few seconds for the above computation before the predicted results appear on the computer screen; the more number of query proteins and longer of each sequence, the more time it is usually needed.

Step 4. Click on the Citation button to find the relevant papers that document the detailed development and algorithm of iSNO-PseAAC.

Step 5. Click on the Data button to download the benchmark datasets used to train and test the iSNO-PseAAC predictor.

It is instructive to point out that the web-server predictor has been developed by strictly observing the guidelines of "Chou’s 5-steps rule" and hence have the following notable merits (see, e.g., [2-29] and three comprehensive review papers [30-32]): (1) crystal clear in logic development, (2) completely transparent in operation, (3) easily to repeat the reported results by other investigators, (4) with high potential in stimulating other sequence-analyzing methods, and (5) very convenient to be used by the majority of experimental scientists.

Moreover, it has not escaped our notice that during the development of iSNO-PseAAC web-server, the approach of general pseudo amino acid components [33] or PseAAC [34] had been utilized and hence its accuracy would be much higher than its counterparts, as concurred by many investigators [1-6,8-11,13,18,26,30-300].

It is anticipated that iSNO-PseAAC may become a very useful high throughput tool for conducting proteome analysis as well as drug development.

For the remarkable and awesome roles of the "5-steps rule" in driving proteome, genome analyses and drug development, see a series of recent papers [31,32,291,301-309] where the rule and its wide applications have been very impressively presented from various aspects or at different angles. References


Kuo-Chen C (2019) How the artificial intelligence tool iSNO-PseAAC is working in predicting the cysteine S-nitrosylation sites in proteins.


