IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Big Data In Bioinformatics - Identifying Proteins

Umeda M.M.
Assistant Lecturer

Abstract: As we all know big data is used anywhere and everywhere, a big data is a large diverse set of structured, semi-structured and unstructured data. Here we are discussing about use of big data in bioinformatics. A project of bioinformatics to identifying protein disulfide bonds.

Disulfide bonds (SS) are post-translational modifications important for the proper folding and stabilization of any cellular proteins with the rapeuticuses, including antibodies and other biologics. With budding advances of biologics and biosimilar, there is a mounting need for a robust method for accurate identification of SS. Even though several mass spectrometry methods have emerged for this task, their practical use rests on the broad effectiveness of both sample preparation methods and bioinformatics tools. Here we present a new protocol tailored toward mapping SS; it uses readily available reagents, instruments, and software.

For sample preparation, a 4-h pepsin digestion at pH1.3 followed by an overnight trypsin digestion at pH6.5 can maximize the release of SS-containing peptides from non-reduced proteins, while minimizing SS scrambling. For LC/MS/MS analysis, SS-containing peptides can be efficiently fragmented with HCD in a Q Enactive Orbitrap mass spectrometer, preserving SS for subsequent identification. Our bioinformatics protocol describes how we tailored our freely downloadable and easy-to-use software, Spectrum Identification Machine for Cross-Linked Peptides (SIMXL), to minimize false identification and facilitate manual validation of SS-peptide mass spectra. To substantiate this optimized method, we've comprehensively identified 14 out of 17 known SS in BSA.

I. INTRODUCTION

Overview:

Disulfide bonds (SS) are formed between the sulfhydryl groups of vicinal cysteines and are among the most common post-translational modifications in proteins. They play important roles in folding proteins and stabilizing functional protein domains. Proper SS arrangements are important for maintaining protein functions and dysregulation of SS formations have been reported in diseases, including neurodegeneration, cancer, inflammation, and heart disease. Therefore, precise identification of SS is critical for understanding protein functions in cells and assuring accurate productions of therapeutic proteins.

The prevailing view is that disulfide bonds have been added during evolution to enhance the stability of proteins that function in a fluctuating cellular environment. However, recent evidence indicates that disulfide bonds can be more than inert structural motifs. The function of some secreted soluble proteins and cell-surface receptors is controlled by cleavage of one or more of their disulfide bonds; this cleavage is mediated by catalysts or facilitators that are specific for their substrate.

Problem Statement:

Difficult to find all combinations of Customer Bridge and evaluate protein sequence using tools and other manual technique.

Motivation:

A prediction technique using python packages can automate this. This will save lot of time for analyst. Use of big data makes it easier to different types data together. This will help in finding new proteins.

Methodology:

The input is taken from the fasta file this will give us a protein sequence from this protein sequence we create a peptide sequence then we parse y-ions and b-ions and calculate their molecular mass then finally we get spectral mass this spectral mass is compared with the theoretical molecular mass of the protein sequence which is the final output of this method. This process is carried out through python coding and other packages by which lead us to get the final output.

Conclusion:

By this system we can easily compare the theoretical molecular mass of the protein and experimental molecular mass of the protein. This system will help us to save a lot of time. This system gives the more efficient output. This system will lead to great high end medical researches using bioinformatics and big data.

References:

1. Hogg PJ, Disulfide bonds as switches for protein function, Trends biochem Sci 28(4)(2003) 210-4
2. Thornton JM, Disulfide bridges in globular proteins, J Mol Biol 151(2)(1981) 216-87