

Chapter 19

In Silico Pharmaco–Gene– Informatic Identification of Insulin–Like Proteins in Plants

Koona Saradha Jyothi
Andhra University, India

G. R. Sridhar
Endocrine and Diabetes Centre, India

Kudipudi Srinivas
V R Siddhartha Engineering College, India

B. Subba Rao
Andhra University, India

Allam Apparao
Jawaharlal Nehru Technological University (JNTU), India

ABSTRACT

*This chapter presents an extension of the authors' earlier work, where they showed that nucleotide/amino acid sequences related to insulin occurred in the plant kingdom. It was believed that plants did not have, nor did they need insulin, a protein hormone considered to be restricted to the animal kingdom. In the current study, the human insulin sequence was initially obtained from UniProt/SwissProt (accession no. P01308). Plant genome sequences were obtained from NCBI PubMed (*Bauhinia purpurea* [Gi|229412], *Vigna unguiculata* [P83770], and *Canavalia ensiformis* [Gi|7438602]. Scores were obtained from ProtFun 2.2 [<http://www.cbs.dtu.dk/services/ProtFun/>]. At the next stage, functions of insulin and glucokinase (insulin like proteins in plants) were predicted by the Protein Function Prediction database (<http://dragon.bio.purdue.edu/pfp/index.html>), followed by functional site prediction from the ELM database (<http://elm.eu.org/>). ProtFun predicted the following functions: human insulin (Cell envelope), Jack bean (Energy metabolism), *Bauhinia purpurea* (Translation). The amino acid Glycine at 32 positions was most highly conserved. Present predictions advocate the use of these sequences (QHLCGS motif) as targets for probing the other plants with lesser homology. In summary our in silico studies have suggested that *Bauhinia purpurea* (Purple orchid tree-BP), *Vigna unguiculata* (Cow pea-CP) and *Canavalia ensiformis* (Jack bean-JB) have conserved the important regions of the human insulin protein.*

DOI: 10.4018/978-1-4666-0309-7.ch019

INTRODUCTION

Informatics has been firmly integrated into biology and drug discovery. Technological advances in both biology and computation improved the way we can 'survey and interrogate biological phenomenon,' which helps to give a clear insights between sequence and function (Weng & Guigo, 2008). This has been applied to a broad sweep of biological phenomena, including recognition to develop national policies for human health (Smith, et al., 2005), ethnobotany genomics, where informatics adds value to both traditional knowledge and scientific biologic knowledge (Newmaster & Ragupathy, 2010), and to conservation biology (Primmer, 2009). It helps in making sense of voluminous and complex data that is being rapidly generated (Blundell, et al., 2006).

In the field of drug discovery, informatics plays a key role in drug development, by identifying 'biological pathways and processes' that can lead to more efficient drug target identification (Krajilevic, et al., 2004).

Structure-based enzyme engineering can help in development of novel hybrid enzymes, which can be targets for next generation drug screening (Ruan, et al., 2009). Other technologies use simulation annealing via bioinformatic approach for protein-ligand flexible docking (Tayaka, et al., 2008). Further advances utilize prediction of drug-target interaction networks (Yamanishi, et al., 2010) and use of cross-genome sequence comparison for generation of new molecule models (Sonmez, et al., 2009).

Sequence based homology studies play an important role in evolutionary tracing as well as grouping of the sequences with common related functions. Thus, it can be presumed that two or more biological species, systems, or molecules that display homology likely share a common evolutionary ancestor. Sequence homology searches are typically performed with a query

DNA or protein sequence to identify known genes or gene products that share significant similarity and hence might inform on the ancestry, heritage, and possible functions of the query gene. One of the driving forces behind bioinformatics is the search for similarities between different biomolecules. Homology methods are the most powerful and are based on the detection of significant extended sequence similar to a protein of known structure, or of a sequence pattern characteristic of a protein family.

At present, various methods are available to analyze biological sequence information. However, with the advent of proteomic era, there is growing demand for analysis of huge amount of biological sequence information and it has become necessary to have programs that would provide speedy analysis (Shil, et al., 2006). It is claimed that the Ishan sequence homology package developed by these workers uses various tools like Fasta, ClustalW, and has made the analysis much faster with reduced manual intervention. The sequence alignment in these methods is a way of arranging the primary sequence (DNA, RNA, or protein) to identify regions of similarity that may be a consequence of functional, structural or evolutionary relationships among the sequences.

Abdulhafezselim (2009) has shown that the bioinformatic analysis of osteoactivin sequences in rats related to osteopetrosis is based on prediction of novel functions using Pfam, Smart, and Elm. The sequence analysis and homology modeling by Costanzi *et al.* (2004) on the various proteins related to P2y nucleotide receptors were able to classify the sequences with related functions. Very recently, Chitale *et al.* (2009), have shown that the conventional PSI-Blast and the Protein Function Prediction (PFP) algorithm is highly effective for predicting functions for unknown proteins.

The development of computational tools has changed the pace and direction of pharmacological discovery, not only in terms of classifying and

16 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/silico-pharmaco-gene-informatic-identification/64080

Related Content

Using a Genetic Algorithm and Markov Clustering on Protein–Protein Interaction Graphs

Charalampos Moschopoulos, Grigorios Beligiannis, Spiridon Likothanassis and Sophia Kossida (2013). *Bioinformatics: Concepts, Methodologies, Tools, and Applications* (pp. 805-816).

www.irma-international.org/chapter/using-genetic-algorithm-markov-clustering/76096

Scaling Unsupervised Risk Stratification to Massive Clinical Datasets

Zeeshan Syed and Ilan Rubinfeld (2011). *International Journal of Knowledge Discovery in Bioinformatics* (pp. 45-59).

www.irma-international.org/article/scaling-unsupervised-risk-stratification-massive/52770

Spatial Structures of Fibrillar Proteins

Gennadiy Vladimirovich Zhizhin (2022). *International Journal of Applied Research in Bioinformatics* (pp. 1-14).

www.irma-international.org/article/spatial-structures-fibrillar-proteins/290346

Data Stewards, Curators, and Experts: Library Data Engagement at Samuel J. Wood Library at Weil Cornell Medicine

Peter R. Oxley, Sarah Ben Maamar and Terrie Wheeler (2024). *Research Anthology on Bioinformatics, Genomics, and Computational Biology* (pp. 566-583).

www.irma-international.org/chapter/data-stewards-curators-experts/342544

Bolstered Resubstitution Error

(2011). *Feature Selection and Ensemble Methods for Bioinformatics: Algorithmic Classification and Implementations* (pp. 383-405).

www.irma-international.org/chapter/bolstered-resubstitution-error/53914