Identification of INSIG2 Associated with Obesity Disease

Student: Qiong Cheng[1]

Instructor: Dr. Irene T. Weber, Dr. Robert W. Harrison

Abstract

Obesity, as an increasingly serious public healthy disease, has been attracting more and more researchers. INSIG2, as an attractive candidate protein associated with obesity, has been studied recently. The structure prediction and verification have been significantly expected. So ***our project aim*** *is to identify its structure by employing some popular bioinformatics tools such as PDB, BLAST, ClusterW, RASMOL and AMMP.*

1. Introduction

Obesity has been attracting more and more researchers in molecule biology and genetics due to the known risk factor of the other chronic diseases and the enormous growth of the population with obesity across the world and all age groups. People with obesity have excessive body fat which greatly impairs their health and even leads to a series of diseases such as heart disease, diabetes, high blood pressure and some forms of cancer [NIH]. The disease may be inherited from generation to generation. Evidences in [CB01] [NIH], on one hand, indicate that genes make a difference to the disease. A significantly discovered gene is Leptin which is a landmark of research into the molecular basis of obesity control. The Leptin is a hormone which can be released into the blood stream and signal to the brain that the body is full with the food [NIH]. Evidences, on the other hand, show us that gene is not the only cause because environmental, psychological and other factors may together play an indispensable role [NIH].

[1] qcheng1@student.cs.gsu.edu

Obesity, as a metabolic disease, can only happen if the critical enzyme is not activated or the control mechanism of metabolic pathways is destroyed [NIH]. These years, researchers in biochemical and molecule biological areas made considerable improvement and defined a series of pathways which are related to obesity and regulating energy balance and adipogenesis. Figure 1 is one pathway, which is maturity onset diabetes of the young. The figure 1 indicates that INSIG (Insulin-induced protein) family plays an important part for obesity and its' prevention. As we know, insulin is known as fat storage. With the insulin secreting more, it will stimulate the body to store more fats. So it prompts the fat cell to have larger size rather than to increase the number of fat cells. The pathway control mechanism, in figure 1, will be destroyed once the members of the family are disabled so that one cause of obesity is triggered.

INSIG family contains a number of eukaryotic Insulin-induced proteins (INSIG-1 and INSIG-1) with the length of approximately 200 residues [BLAST]. INSIG-1 and INSIG-2 can be found in the endoplasmic reticulum and bind the sterol-sensing domain of SREBP cleavage-activating protein (SCAP). They prevent it from escorting SREBPs to the Golgi. Their combined action permits feedback regulation of cholesterol synthesis over a wide range of sterol concentrations [BLAST]. By blocking this movement, insig-2, like the previously described insig-1, prevents the proteolytic processing of SREBPs by Golgi enzymes, thereby blocking cholesterol synthesis. The sequences of human insig-1 and -2 are 59% identical. Both proteins are predicted to contain six transmembrane helices [DM02]. The proteins differ functionally in two aspects: (*i*) production of insig-1 in cultured mammalian cells requires nuclear SREBPs, but insig-2 does not; and (*ii*) at high levels of expression, insig-1 can block SCAP movement in the absence of exogenous sterols but insig-2 do not need [DM02].

INSIG2 (insulin induced protein 2) is a member of INSIG family [AN06]. It is an attractive candidate because the protein's product inhibits the synthesis of fatty acid and

cholesterol [AN06]. There are lots of variants on INSIG2. NP_057217 is one of them and discovered in homo sapiens. It has 225 residues. There were three papers about NP_057217 published in Natural Academic Science recently [YJM06] [YJMJJ06] [JYX06]. They studied the control mechanism of INSIG1 or INSIG2 but the structure of INSIG2 are still in research. This is my motivation to the project. In the following two sections, we present our analysis methods and the corresponding results.

2. Analysis Methods

We make full use of the following technologies and tools taught in Bioinformatics class and try to identify the structure of INSIG2 protein.

2.1 NCBI and PDB database searching

We use NCBI database to find the variant NP_057217 of INSIG2 of homo sapiens. The primary structure is *maegetespgpkkcgpyissvtsqsvnlmirgvvlffigvflalvlnllqiqrnvtlfppdviasifssa wwvppccgtasavigllypcidrhlgephkfkrewssvmrcvavfvginhasakvdfdnniqlsltlaalsiglwwtfdrsrsgfglgvgi aflatvvtqllvyngvyqytspdflyvrswlpciffaggitmgnigrqlamyeckviaekshqe*. But the other structures of the protein are still in construction. We use PDB database to find the known proteins and their structures.

2.2 Blast Sequence Alignment searching

Every alignment tool provides their individual algorithm for searching the specified database, finding the efficient sequence motifs and calculating sequence similarity score with the tradeoff of efficiency and preciseness. NCBI Blast is a popular protein-to-protein alignment tool which supports the searching of several protein databases such as pdb, swissprot, nr, and env_nr. We input amino acid sequence of INSIG2 in BLAST and get 6 blast hits on pdb database. The scores of the six hits (1RHI-3, 1R08-3, 1BWV-A, 1IWA-A, 1RH5-A and 1RHZ-A) cover from

29.3 to 25.8, which are not high. The significant identities are also not high. The ones with acceptable score and significant identity are 1RHI and 1R08. 1RHI seems to be a little better than 1R08 from the result of BLAST.

2.3 LALIGN tool for pair wise alignment

This tool has no feature of searching database which is provided in BLAST but it supports more precise algorithm for the sequence alignment of protein and DNA. We aligned INSIG2 with the six proteins found in BLAST one by one. I found out the significant identities are a little different from the ones in BLAST. 1R08 has 27.6% but 1RHI has 25.5% with INSIG2. Both of them have 3 gaps. For 1R08, it has 105 residues; the biggest gap has 10 residues. We try to employ another alignment tool ClustalW to make a final decision since the results coming from BLAST and LALIGN are different.

2.4 ClustalW tool for multiple alignments

ClustalW is a general purpose alignment program of multiple sequences for DNA or proteins. Comparing with BLAST and LALIGN, it supports multiple sequence alignment and provides the view of evolutionary relationship tree from which we can judge the evolutionary distance between different proteins or DNAs. We input the sequences of 1R08, 1RHI and INSIG2 and get the score table and evolutionary tree as Figure 2. It indicates that 1R08 and 1RHI have quite similar evolution relationship with INSIG2. We finally determine to align the molecule structure of INSIG2 with the one of 1R08. In ClustalW, the gaps are not longer than the results from BLAST and LALIGN. There are 6 gaps over 242 residues after ignoring the gaps in endpoints but the longest gaps just has 4 residues long. So I used the alignment result from ClusterW as my target of structure homology. The alignment sequences for structure are as follows (The first sequence is 1R08's 3rd chain; the second sequence is INSIG2; * represents gap residue):

GLPTTTLPGS*GQFLTTDDRQSPSALPNYEPTPRIHIP*GKVHNLLEIIQVDTLIPMNNTHT*
KDEVNSYLIPLNANRQNEQVFGTNLFIGDGVFKTTLLGEIVQYYTHWSGSLRFSLMYTGPALSS
*AKLILAYTPPGARGPQDRREAMLGTHVVWDIGLQSTIVMTIPWTSGVQFR*YTDPDTYTSAGF*
LSCWYQTSLILPPETTGQVYLLSFISACPDFKLRLMKDTQTISQTVALTE

MAEGETESPGPKKCGPYISSVTSQSVNLMIRGVVLFFIGVFLALVLNLLQIQRNVTLFPPDVIAS
*IFSSAWWVPPCCG**TASAVIG*LLYP****CIDRHLGEPHKFKREWSSVMRCVAVFVGINHAS*
AKVDFDNNIQLSLTLAALSIGLWWTFDRSRSGFGLGVGIAFLATVVTQLLVYNGVYQYTSPDFL
****YVRSWLPCIFFAGGITMGNIGRQLAMYECKVIAEKSHQE********

2.5 AMMP molecular structure alignment

AMMP is Another Molecular Mechanics Program created by Dr. Robert W. Harrison. It supports the structure alignment of both small molecules and macromolecules including proteins, nucleic acids and other polymers [AMMP]. It implements the structure homology with a series of optimization models such as energy minimization. We employed homology mapping program (Newho) firstly to build a new pdb file from 1R08 pdb file using an alignment file; then we used previous AMMP's handling program PREAMMP which converts a Cartesian atomic description of the structure into the molecular geometry file which AMMP reads [AMMP]; finally, AMMP was used to build model and to do distance geometry and energy minimization. The aligned model was saved as a pdb file.

2.6 RasMol---Displaying Tool of PDB Structures

RasMol is a molecular graphics program aimed at the visualization of proteins, nucleic acids and small molecules [RASMOL]. We employ it to display the identified INSIG2 and to compare 1R08 with it.

3. Results of analysis

The target INSIG2 has the structure shown in the Figure 3 and 4. The ribbons display of the molecule structure is in Figure 4 and the molecule display is in Figure 3. The INSIG2

segment has 15 β–sheets and 7 α–helixes which are close to the data mentioned in [AN06]. As described in previous section 2.5, we employed a serial of AMMP tools to identify the molecule homology based on the aligned sequences. The result about INSIG2 molecule is 225.

For the identified INSIG2, we further compare it with the corresponding 236 residues of the third chain of 1R08 in RASMOL. Evidence, in Figure 5 and 6, indicates they look similar although they are different. Figure 5 shows the ribbons display of them and it indicates the similarity. Figure 6 proves that they are different since the green one is the structure of INSIG2 and the other colors are for the 3rd chain of 1R08. Furthermore, we tried to identify the biggest gap with 4 residues but RasMol shows us no atoms are found.

4. Conclusion from Research

We studied and employed the popular and excellent bioinformatics tools to identify INSIG2 which is an attractive candidate protein associated with obesity. As a computer science student, the experience is important for me. We are thinking of what we can do for bioinformatics and what point we should stand in. All the tools we learnt in class are useful and can be used for beginner and researchers; and we can work on wider applications of bioinformatics by practicing the theories we have learned from computer science area. AMMP, as an excellent tool developed by computer science professor Dr. Robert W. Harrison, used some combination optimization models and computer intelligence technologies such as genetic algorithm. We also believe that more classic algorithm and distributed and parallel algorithms can be used to improve the performance and preciseness for different projects. The most difficult thing for us is to learn how to exactly understand and define some research problem for the professional biologists, chemists and even physicist.

References

[AN06] Alan H., Norman P.G., Matthew B.M., Iris M. H., Arne P., Thomas I., "A Common Genetic Variant is Associated with Adult and Childhood Obesity", *Science, Vol 312*, 2006

[CB01] http://www.esi-topics.com/obesity/interviews/ClaudeBouchard.html

[KEGG] http://www.genome.ad.jp/kegg/metabolism.html

[NIH] NIH Publication No. 98-4083 (National Institutes of Health, Bethesda, MD, 1998).

[BLAST] http://www.ncbi.nlm.nih.gov/blast/

[LALIGN] http://www.ch.embnet.org/software/LALIGN_form.html

[CLUSTALW] http://www.ebi.ac.uk/clustalw/

[RASMOL] http://www.bernstein-plus-sons.com/software/rasmol/

[AMMP] http://asterix.cs.gsu.edu/ammp.html

[EXPASY] http://www.expasy.ch/ http://ca.expasy.org/tools/

[DM02] Daisuke Yabe, Michael S. Brown,* and Joseph L. Goldstein , "Insig-2, a second endoplasmic reticulum protein that binds SCAP and blocks export of sterol regulatory element-binding proteins", *Proc Natl Acad Sci. v99(20): 12753–12758,* 2002

[YJM06] Yi Gong, Joon No Lee, Michael S. Brown,* Joseph L. Goldstein,* and Jin Ye , "Juxtamembranous aspartic acid in Insig-1 and Insig-2 is required for cholesterol homeostasi", *Proc Natl Acad Sci.* v*103(16): 6154–6159,* 2006

[YJMJJ06] Yi Gong, Joon No Lee, Michael S. Brown,* Joseph L. Goldstein,* and Jin Ye , "Proteasomal degradation of ubiquitinated Insig proteins is determined by serine residues flanking ubiquitinated lysines", *Proc Natl Acad Sci.* v*103(16): 6154–6159,* 2006

[JYX06] Joon No Lee,* Yi Gong,* Xiangyu Zhang,† and Jin Ye*, "Proteasomal degradation of ubiquitinated Insig proteins is determined by serine residues flanking ubiquitinated lysines", *Proc Natl Acad Sci.* v 103(13): 4958–4963*,* 2006

Figures and Tables

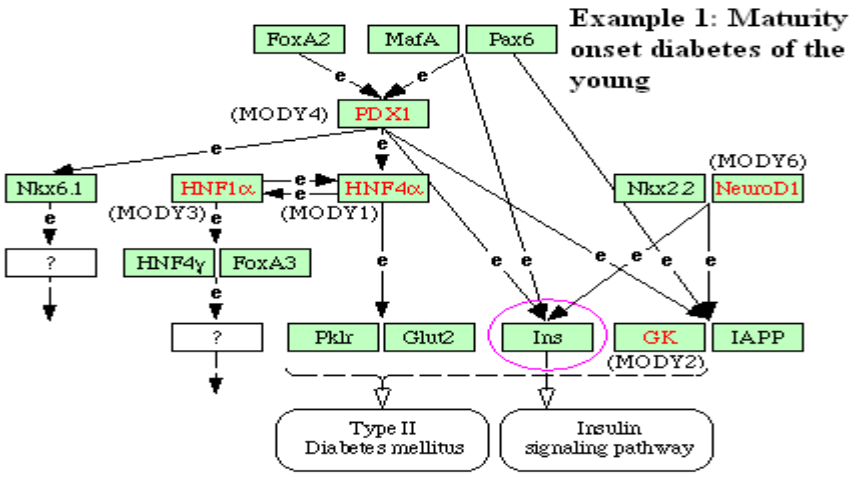

Figure 1: INSIG Family's ROLE

The red circles emphasize the crucial role in the metabolic disease.

SeqA	Name	Len(aa)	SeqB	Name	Len(aa)	Score
1	1R08_3|PDBID|CHAIN|SEQUENCE	236	2	INSIG2_1|NCBI|CHAIN|SEQUENCE	225	15
1	1R08_3|PDBID|CHAIN|SEQUENCE	236	3	1RHI_3|PDBID|CHAIN|SEQUENCE	236	89
2	INSIG2_1|NCBI|CHAIN|SEQUENCE	225	3	1RHI_3|PDBID|CHAIN|SEQUENCE	236	9

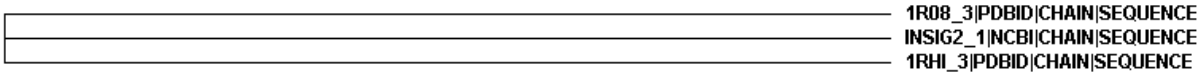

Figure 2. Scores table and cluster tree among 1R08, 1RHI and INSIG2 in ClusteW tool

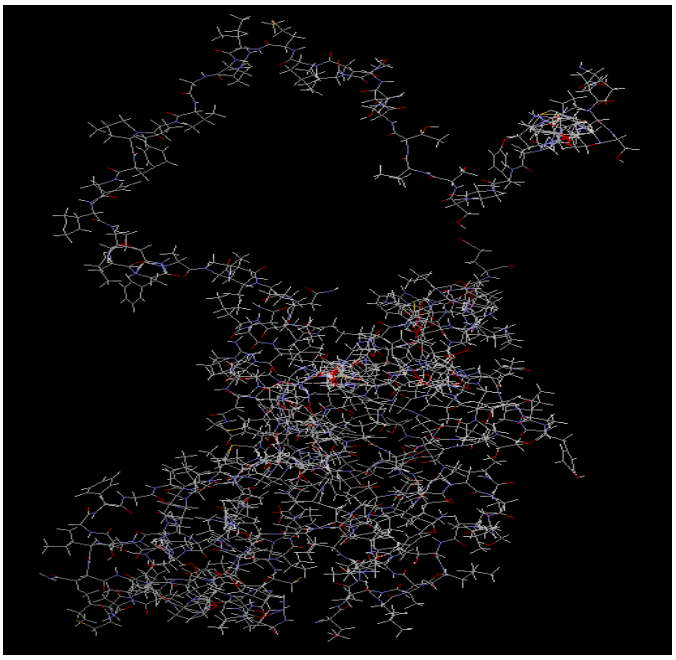

Figure 3. The Identified Molecular Structure of INSIG2 in RASMOL Tool

Figure 4. The Identified Molecular Structure of INSIG2 – Ribbons Display in RASMOL Tool

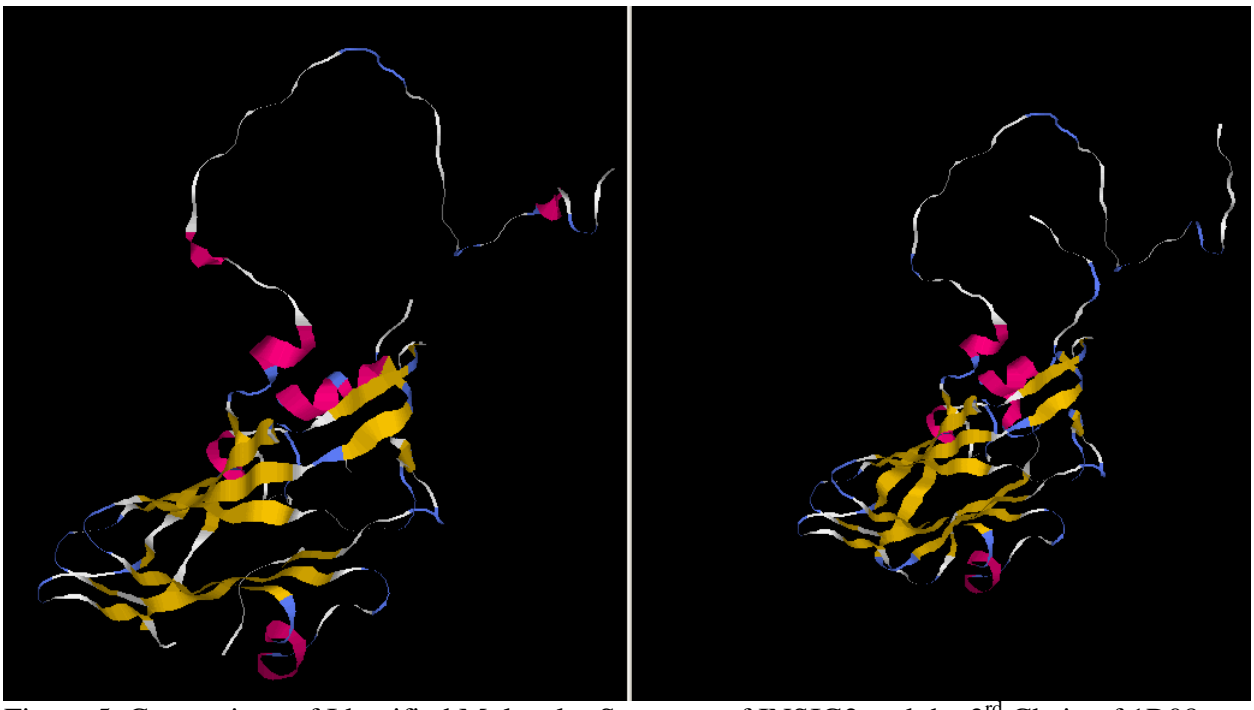

Figure 5. Comparison of Identified Molecular Structure of INSIG2 and the 3rd Chain of 1R08 ---- Ribbons Display in RASMOL Tool
The left figure is INSIG2; the right figure is 3rd Chain of 1R08.

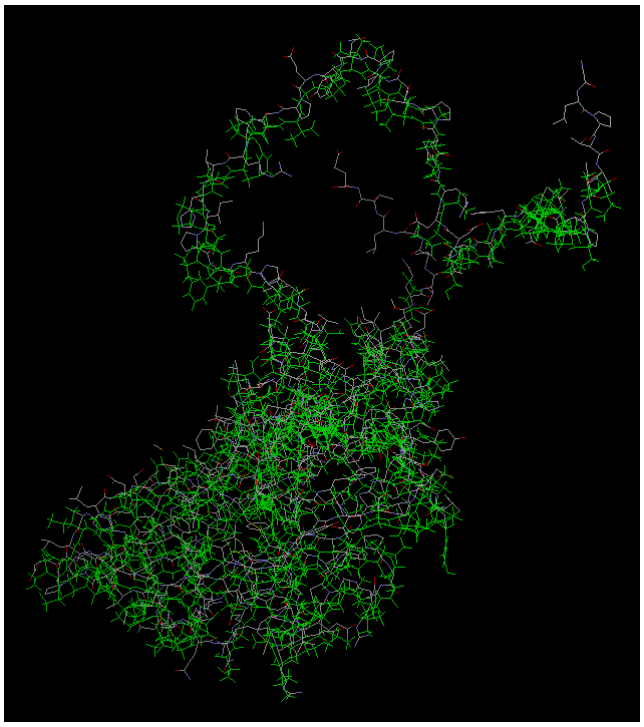

Figure 6. Big Contrast of Identified Molecular Structure of INSIG2 and the 3rd Chain of 1R08 --- in RASMOL Tool

The green molecule is the INSIG2. The other color chain is the 3rd chain of 1R08. The figure indicates there exist the differences between them.