

Efficiency In Detection of Porcine and Human Diabetes Insulin Using Smith-Waterman and Needleman-Wunsch Algorithms

Muhammad¹, Iskandar and Akzam Saidin

Faculty of Computing, Universiti Teknologi Malaysia (UTM), Malaysia

Muhammad94sukeri@gmail.com¹, iskandar@utm.my²

Abstract. The aim of this research is to determine the efficiency of algorithm that can produce a significance result to differentiate the human and porcine diabetes insulin. The thesis is related to Halal research, in order to propose a new method by replacing wet laboratory using computational approaches, thus, the output from the program plays an important role. To achieve this aim, the algorithms of pairwise sequences alignments using both global and local alignments, namely Needleman-Wunsch and Smith-Waterman algorithms are implemented in the program and will be compared to determine the homologues sequences of porcine proinsulin in term of identity percentage, gap percentage and value of score alignments. The algorithms that provide high identity and less gap will be selected as the best algorithm in differentiate between human and porcine insulin. The thesis shows that Smith-Waterman provides a better result than Needleman-Wunsch algorithms.

Keywords: Halal insulin, Sequence alignment, Needleman-Wunsch, Smith-Waterman.

1 Introduction

Islam is not only a mere religion, it has certain rule that need to be applied and practice by the Muslim. This practice includes all of the aspect of living including by observing dietary law which come from Islamic teaching. Muslim are very cautions in buying or consuming every product that meet it requirement to ensure the Halal-ness. Some of the product in the market are clearly about the status of Halal in the product, but some of its show less information regarding of their product, this issues are really concerning towards the Muslim users as they cannot identify the product is whether Halal to be consume (Halim et.al, 2014).

Halal certification involve a process of detection using wet laboratory to determine the composition of the product to ensure the quality and halalness. This thesis is to proposed a new method using dry laboratory by using a computational method or mathematical analyses through computer-generated model. Since Halal detection is a crucial part, most of the product are conducted through wet laboratory using a PCR and immunoassay. Accoding to profound kestrel laboratoy, rapid immunoassay tests are convenient and inexpensive, however the trade-off is the sensitivity and accuracy. Culture plates are accurate and reliable but also require longer process time. Thus, PCR and qPCR is by far the best option in terms of specificity, sensitivity and speed.

The main purpose of the project is to proposed a porcine detection method by using a computational approach in identify the sequence of insulin that may inflicted with the composition from the animal insulin made from porcine. This research might be able to

enhances a further research in Halal area to set upon a new threshold of porcine and human insulin as a guide to differentiate between human and porcine insulin. The method used in this paper is by doing a pairwise sequence alignments by compare with sequence of precursor insulin collected from NCBI (National Center for Biotechnology Information) to a reference sample of porcine proinsulin obtains from research paper by finding the similarity of the of human and porcine diabetes insulin based on the criteria of identity percentage, gap percentage and score alignment.

Based on the review during MIHAS 2016, for every company that applying for the Halal certification must go through a certain process before receives the status of their product. This process involved is by doing an experimental using a wet laboratory to determine whether their product is purely Halal and the procedure is a very long process and could take months to obtains the results. By using a wet laboratory focusing on computational approach, this might become as another alternative to reduce the cost and time processing.

The objective of this paper is to analyse data of precursor insulin of human and porcine in form of protein sequence, apply Smith-Waterman and Needleman-Wunsch algorithms in pairwise sequence alignments based on precursor insulin and porcine proinsulin and evaluate the result of both selected algorithms and compared it based on percentage identity, percentage gap and score alignments to determine which of the algorithms produce a significance result in differentiate the human and porcine insulin.

The significance of this research able improve the knowledge in computational method in order to detect porcine in diabetes insulin. The important of using computational approaches in detecting porcine in diabetes insulin might be an another alternative to replace existing method that will help to reduce time and cost in doing the experimental process.

2 Related Work

Islam is not a religion but it is also include the ways of life for a person with applying certain rules and manners for every Muslim. The term of halal is based on Arabic word that bring the definition of permitted or the lawful. The haram is the word that opposite the Halal, that bring the definition of unlawful and prohibited For Muslim, the term of halal and haram are the basically term used that applied in their daily life(Eliasi et al,2002).

Insulin contains 3 categories made from animal insulin,human insulin and analogue insulin. Animal insulin are extracted from the pancreases of pig to produce porcine insulin or bovine insulin and it provide almost the same effect when applied to human body(Erika et al,2013).For human synthetic insulin is made using laboratory by recombinant DNA technology. The insulin product mimics the structure of DNA insulin based on human pancreases secretion(Aggarwal et.al,2012) while analogue insulin is a sub group of human insulin. This type of insulin is a synthetically modified insulin according to the structure of chemical properties of human insulin through genetic engineering process in order to achieve a compound with different chemical properties as human insulin. The chemical structure to increase rapid acting or more uniformly acting form of insulin in absorption, metabolism, excretion and distribution characteristics(Lily et.al,2011).

The implementation of Smith-Waterman algorithms is based on the method known as dynamic programming, which the algorithm tends to align with any of the length at any of the location in the sequences by determining whether an optimal alignment can be found (Ayuade et al., 2007) while Needleman-Wunsch is based on global optimization that allows the subject sequence alignment to span equally to the entire length of the query sequences, this method attempts to determine every residue sequences and align it accordingly (Brudno et al., 2003).

3 Methodology

The section is to present the methodology involved in the research. Figure 1 illustrates the steps involving the research that include 4 phases. The first phase discusses the problem identification and specification related to the research. The second phase is about data definition and collection that define the data experiments related to the studies of insulin. The third phase is the implementation of both local and global alignments, namely Smith-Waterman and Needleman-Wunsch algorithms. The last phase is the analysis of the result, this phase is to analyse the result based on the sequence similarity of data insulin and porcine proinsulin for both of the results from the algorithms.

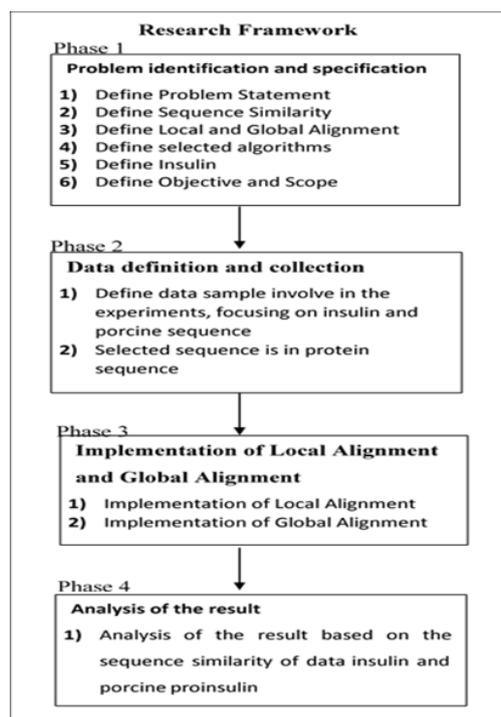


Figure 1: Overview of research methodology

4 Experimental Result

In this research, the experimental data are in protein sequences and consists of two categories, data sample and reference sample. The data sample are precursor insulin consists of human, porcine and analogues insulin while reference sample are the proinsulin of porcine (Ronald Chance,1968). The data from data sample and reference sample will undergo pairwise sequence alignments using Smith-Waterman and Needleman-Wunsch algorithms/program to determine the similarity of both sequences with assigned gap penalty of 10, gap extension of 0.5 and with the implementation of BLOSUM62. The output from both of the program will be compared and evaluated based on the optimal alignments, identity percentage, gap percentage and the score alignments. This output from the program may enhance further research in determining the threshold to differentiate between human and porcine insulin. For each of the data sample will undergo pairwise sequence alignments with reference alignments using Smith-Waterman and Needleman-Wunsch algorithms.

Table 1: Data Sample and Reference Sample

No	Data Sample	Type
1	>gi 124617 sp P01308.1 INS_HUMAN RecName: Full=Insulin; Contains: RecName: Full=Insulin B chain; Contains: RecName: Full=Insulin A chain; Flags: Precursor MALWMRLLPLLALLALWGPDPAAAFV NQHLCGSHLVEALYLVCGERGFFYTPK TRREAEDLQVGQVELGGGPGAGSLQPL ALEGSLQKRGIVEQCCTSICSLYQLENY CN	Human Insulin
2	>gi 12643972 sp P01315.2 INS_PIG RecName: Full=Insulin; Contains: RecName: Full=Insulin B chain; Contains: RecName: Full=Insulin A chain; Flags: Precursor MALWMRLLPLLALLALWGPDPAAAFV NQHLCGSHLVEALYLVCGERGFFYTPK TRREAE VGQVELGGGPGAGSLQPLALEGSLQKR GIVEQCCTSICSLYQLENYCN	Human Insulin2
3	>gi 172073148 ref NP_001103242.1 insulin precursor [Sus scrofa] MALWTRLLPLLALLALWAPAPAQAFVN QHLCGSHLVEALYLVCGERGFFYTP KARREAENPQAGAVELGGGLGGLQAL ALEGPPQKRGIVEQCCTSICSLYQLENY CN	Porcine Insulin
4	>Synthetic AnalogueHuman MALWMRLLPLLALLALWGPDPAAAFV NQHLCGSHLVEALYLVCGERGFFYTPK TRREAE DLQVGQVELGGGPGAGSLQPLALEGSL QKRGIVEQCCTSICSLYQLENYCNL	Analogue Insulin
5	>Porcine Proinsulin RREAQNPQAGAVELGGGLGGLGALALE GPPQKR	Porcine Proinsulin

5 Discussion

Based on the result from the program implemented from the Smith-Waterman and Needleman-Wunsch algorithms are illustrated in Figure below according to the score alignment, identity percentage and gap percentage. According to the program, Smith-Waterman shows a significance result when compared to Needleman-Wunsch algorithms in term of identity percentage and gap percentage but the score alignment remain the same for both of the program.

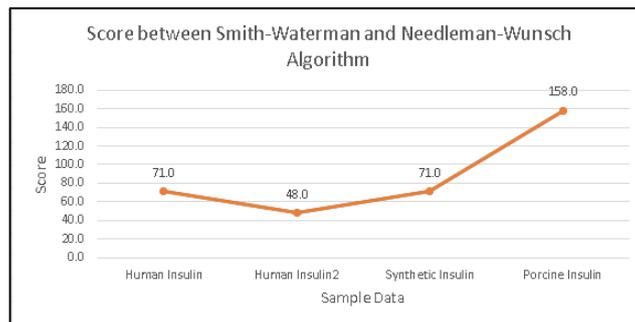


Figure 2: Score Alignment

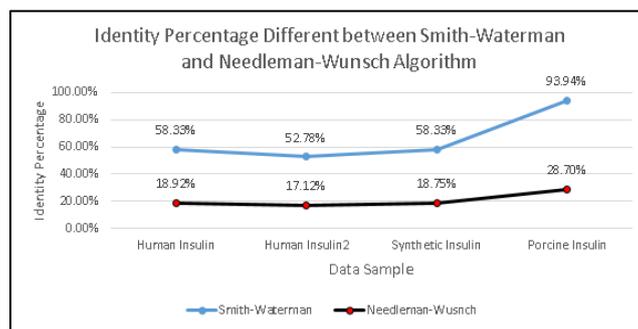


Figure 3: Identity Percentage Different

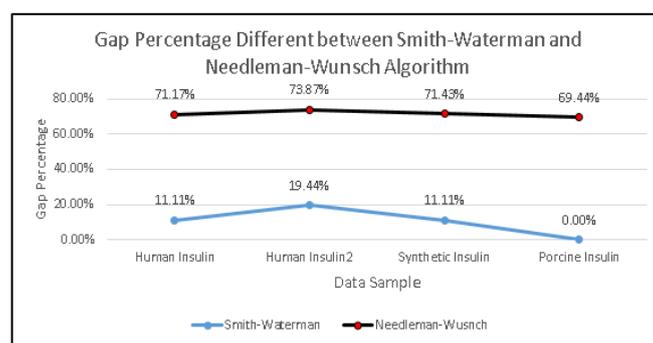


Figure 4: Gap Percentage Difference

6 Conclusion

In the research, the aim of the project is to determine which of the algorithms in pairwise sequence alignment provide the most significant result in detecting sequence similarity to the porcine proinsulin. The research is intended to solve the problem related to the wet laboratory that are usually expensive and time consuming by proposing a new method using computational approaches that can be implemented in the manufacturing level of Halal detection for diabetes insulin.

Based on the result, both of the programs manage to differentiate between human and porcine insulin but Smith-Waterman shows a significant result over the Needleman-Wunsch algorithm. Therefore, for future works, the output from the program can be a resource to enhance research in Halal areas and can be used to determine a threshold allowed between human and porcine sequence percentages to be Halal. Smith-Waterman tends to produce a better result over the Needleman-Wunsch algorithms as the reference sample sequence are shorter than the data sample and produce a better output in terms of the sequence similarity (Polyanovsky et al., 2011).

Acknowledgements. I would like to express my gratitude toward my supervisor and co-supervisor, Dr Iskandar and Akzam Saidin for their continuous support and guidance throughout the whole project and also immensely grateful to the previous researcher for their absolute works.

References

25. Pearson, William R. "Searching protein sequence libraries: comparison of the sensitivity and selectivity of the Smith-Waterman and FASTA algorithms." *Genomics* 11.3 (1991): 635-650.
26. Ayguade, Eduard, Juan J. Navarro, and Dani Jimenez-Gonzalez. "Smith-Waterman Algorithm." (2007).
27. Brudno, Michael, et al. "Glocal alignment: finding rearrangements during alignment." *Bioinformatics* 19.suppl 1 (2003): i54-i62.
28. Aggarwal, Saurabh R. "What's fueling the biotech engine [mdash] 2011 to 2012." *Nature biotechnology* 30.12 (2012): 1191-1197.
29. An, Jie, et al. "Hepatic expression of malonyl-CoA decarboxylase reverses muscle, liver and whole-animal insulin resistance." *Nature medicine* 10.3 (2004): 268-274.
30. Polyanovsky, Valery O., Mikhail A. Roytberg, and Vladimir G. Tumanyan. "Comparative analysis of the quality of a global algorithm and a local algorithm for alignment of two sequences." *Algorithms for molecular biology* 6.1 (2011): 25.
31. Chance, Ronald E., Robert M. Ellis, and William W. Bromer. "Porcine proinsulin: characterization and amino acid sequence." *Science* 161.3837 (1968): 165-167.