ANGLE BASED PROTEIN TERTIARY STRUCTURE PREDICTION USING BEES OPTIMIZATION ALGORITHM

ZAKARIA NOOR ALDEEN MAHMOOD AL-QATTAN

UNIVERSITI UTARA MALAYSIA 2010

ANGLE BASED PROTEIN TERTIARY STRUCTURE PREDICTION USING BEES OPTIMIZATION ALGORITHM

A project submitted to Dean of Postgraduate Studies and Research in partial

Fulfillment of the requirement for the degree

Master of Science of Intelligent Systems

Universiti Utara Malaysia

By

Zakaria Noor Aldeen Mahmood Al-qattan

KOLEJ SASTERA DAN SAINS
(College of Arts and Sciences)
Universiti Utara Malaysia

PERAKUAN KERJA KERTAS PROJEK
(Certificate of Project Paper)

Saya, yang bertandatangan, memperakukan bahawa
(I, the undersigned, certifies that)

ZAKARIA NOOR ALDEEN MAHMOOD ALQATTAN
(802403)

calon untuk Ijazah
(candidate for the degree of) **MSc. (Intelligent System)**

telah mengemukakan kertas projek yang bertajuk
(has presented his/her project of the following title)

ANGLE BASED PROTEIN TERTIARY STRUCTURE PREDICTION USING BEES OPTIMIZATION ALGORITHM

seperti yang tercatat di muka surat tajuk dan kulit kertas projek
(as it appears on the title page and front cover of project)

bahawa kertas projek tersebut boleh diterima dari segi bentuk serta kandungan dan meliputi bidang ilmu dengan memuaskan.
(that this project is in acceptable form and content, and that a satisfactory knowledge of the field is covered by the project).

Nama Penyelia
(Name of Supervisor) : **DR. MASSUDI MAHMUDDIN**

Tandatangan
(Signature) : _____ Tarikh (Date) : 12/10/10

Nama Penilai
(Name of Evaluator) : **MDM. NUR AZZAH ABU BAKAR**

Tandatangan
(Signature) : _____ Tarikh (Date) : 17/10/10

PERMISSION TO USE

In presenting this project in partial fulfillment of the requirements for a postgraduate degree from the Universiti Utara Malaysia, I agree that the University Library may make it freely available for inspection. I further agree that permission for copying of this project in any manner in whole or in part, for scholarly purposes may be granted by my supervisor(s) or in their absence by the Dean of Postgraduate Studies and Research. It is understood that any copying or publication or use of this project or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to Universiti Utara Malaysia for any scholarly use which may be made of any material from my project.

Requests for permission to copy or to make other use of materials in this project, in whole or in part, should be addressed to

Dean of Postgraduate Studies and Research
College of Arts and Sciences
Universiti Utara Malaysia
06010 UUM Sintok
Kedah Darul Aman
Malaysia

ABSTRACT

The expand development in the area of scientific researches, especially on biological field, leads to the emergence and discovery of many new chemical compounds. Proteins are one of the hot topics and main concern among the biological subjects for the researchers; due to its complexity, diversity and it participate in each biological structure. In order to perform their function they tend to fold into their tertiary structure. There are two main ways to determine their structure, one is by laboratory experiment which is very expensive and time subsuming, and the second is by computation. In computation way the operation is known as optimization problem and the optimum solution is to find the conformation with the lowest free energy. In this project, angles based control with Bees Optimization search algorithm were adopted to search with guidance the protein conformational space in order to find the optimum solution. The experiment was conducted on short sequence protein (Met-enkephaline) which has been used in previous researches. The prototype system was built using Visual C# 2008 to fulfill the protein 3D structure prediction requirements. WEKA program application was used for main chain angles (Phi and Psi) data classification. The experiment shows a good accuracy in term of prediction and lowest free energy identification.

ACKNOWLEDGEMENTS

In the name of Allah, the Most Beneficent, the Most Merciful,

"وَاتَّقُوا يَوْمًا تُرْجَعُونَ فِيهِ إِلَى اللَّهِ ثُمَّ تُوَفَّىٰ كُلُّ نَفْسٍ مَّا كَسَبَتْ وَهُمْ لَا يُظْلَمُونَ"

Means "And be afraid of the Day when you shall be brought back to Allah. Then every person shall be paid what he earned, and they shall not be dealt with unjustly." Holly Quran, Al-Baqara (281).

All thanks to ALLAH who helped me to finish this project after a long time of continuous work. Allah created the ideas and gave the assistance. All thanks and gratitude to Allah who gave me the strength and the ability to complete my study for Master degree. Thanks Allah Almighty for giving me brothers and dear friends I've met during my study in Malaysia. All thanks to them for standing next to me and for their prayers for me. All thanks and appreciation to my ideal example person after Prophet Muhammad (pbuh), my dear brother Mohammad Abu Musa, who stood beside me and had a great influence in creating the idea of the project and also during its implementation. Thanks to my father who supported me financially, prayer and counsel, and may Allah's mercy and forgiveness be upon my mother who I miss her too much. Thanks to my father's wife who encouraged me to travel for study. Best wishes to my kind brother Mohanned may Allah bless his marriage. Thanks to my future wife who gave me the motivation all the time during my study and I wish, I can find her soon.

Finally, thanks and appreciations to all my lecturers at Universiti Utara Malaysia. And a very special thanks to Dr. Massudi bin Mahmuddin who gave me friendship beside supervision, and who was very supportive and cooperative all the time.

TABLE OF CONTENT

CHAPTER TWO: LITERATURE REVIEW

CHAPTER TREE: RESEARCH METHODOLOGY

CHAPTER FOUR: PROTOTYPE DEVELOPMENT

CHAPTER FIVE: RESULTS AND FINDINGS

CHAPTER SIX: CONCLUSION

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

Å	Angstrom
ΔG	Binding energy
ΔG0	Adjustable parameter for binding energy
ΔGele	Energy term describes electrostatic interaction
ΔGvdW	Energy term describes van der Waals interaction
Aij, Bij	Coefficients for van der Waals interaction energy
rij	The distance between the ligand atoms, i, and protein atoms, j
ε	Depth of the energy well
reqm	The equilibrium distance between two atoms where the interatomic interaction energy equals to the depth of the energy well.
qi, qj	Partial atomic point charges for ligand atoms, i, and protein atoms,
ε(rij)	Distance-depended dielectric constant
Phi(ϕ),Psi(ψ), Omega (ω),	the main chain torsion angles

LIST OF ABBREVIATIONS

3D	Three-dimensional
kcal	Kilo calorie
RMSD	Root mean square of deviation
VDW	van der Waals
PDB	Protein data bank
NMR	Nuclear magnetic resonance
BCO	Bees colony optimization algorithm
ABC `	Artificial bees colony
GA	Genetic algorithm

LIST OF APPENDICES

CHAPTER ONE

INTRODUCTION

1.0 Background

Protein folding prediction is one of the interesting concepts in the area of Bioinformatics, as importance of the proteins to the humanity. Proteins involve in performing specific functions in the human's body, their function and performance properties depend upon the three dimensional (tertiary) structure they have (Branden & Tooz, 1999; Fidanova & Lirkov, 2008; Bahamish et al., 2008, 2009; Chen & Li, 2010; Fonseca et al., 2010). Each protein constructed from a sequence of amino acids connected to each other in a long chain. There are 20 different amino acids that combine together to build the protein's sequence, each of which may declared more than once in the same sequence (Branden & Tooz, 1999). According to Findanova and Lirkov (2008), each protein may have range between 20 to 40000 amino acids and most of the proteins have around hundred of amino acids. They also mentioned that the order of the amino acids in the protein is the most important influance factors to the uniqueness of fold of the protein.

There are three main approaches for protein prediction problem; the first is called Comparative Modeling method, which follows the fact of proteins is related depending on the similarity in there sequence. This similarity is measured in order to identify the identical residues in the protein sequence based on proteins optimal structural superposition which is

The contents of the thesis is for internal user only

REFERENCES

Al-qattan, M. N. A.-d. M. (2009). *Combinatorial Design Of Virtual Sialic Acid Analogues Against Influenza A Hemagglutinin Using Structure And Fragment Based Pproaches.* (Unpublished master's thesis). Universiti Sains Malaysia, Penang, Malaysia.

Bahamish, H. A. A., Abdullah, R., & Salam, R. A. (2008). Protein conformational search using bees algorithm. In*: AICMS 08: Second Asia International Conference on Modeling and Simulation,* IEEE Computer, Kuala Lumpur,911–916.

Bahamish, H. A. A., Abdullah, R., & Salam, R. A. (2009, 25-29 May 2009). Protein Tertiary Structure Prediction Using Artificial Bee Colony Algorithm. *Proceedings of the Third Asia International Conference on Modelling & Simulation,* 2009. AMS '09, Bali.

Berjanskii, M., Neal, S., & Wishart, D. (2006). PREDITOR: a web server for predicting protein torsion angle restraints. *Nucleic Acids Research, 34*, W63.

Bernasconi, A., & Segre, A. (2000). Ab initio methods for protein structure prediction: A new technique based on Ramachandran Plots. *European Research Consortium for Informatics and Mathematics News, 43.*

Branden, C., & Tooze, J. (1999). *Introduction to protein structure.* New York: Garland Publisher.

Carugo, O., & Pongor, S. (2001). A normalized root-mean-spuare distance for comparing protein three-dimensional structures. *Protein Science, 10*(7), 1470-1473.

Chen, P., & Li, J. (2010). Prediction of protein long-range contacts using an ensemble of genetic algorithm classifiers with sequence profile centers. *BMC Structural Biology* 10(Suppl 1),S2.

Cui, Y., Chen, R. S., & Wong, W. H. (1998). Protein folding simulation with genetic algorithm and supersecondary structure constraints. *Proteins: Structure, Function, and Bioinformatics, 31*(3), 247-257.

Das, S., Abraham, A., & Konar, A. (2008). Swarm intelligence algorithms in bioinformatics. *Computational Intelligence in Bioinformatics, 94*, 113-147.

Fidanova, S., & Lirkov, I. (2008, 20-22 Oct. 2008). Ant colony system approach for protein folding. *International Multi conference on Computer Science and Information Technology*, 2008. IMCSIT 2008, 887-891.

Fonseca, R., Paluszewski, M., & Winter, P.(2010). Protein Structure Prediction Using Bee Colony Optimization Metaheuristic. *Journal of Mathematical Modelling and Algorithms, 9*(2), 181-194.

Goodsell, D., & Olson, A. (1990). Automated docking of substrates to proteins by simulated annealing. *Proteins: Structure, Function, and Bioinformatics, 8*(3), 195-202.

Karaboga, D., & Akay, B. (2009). A survey: algorithms simulating bee swarm intelligence. *Artificial Intelligence Review, 31*(1), 61-85.

Karaboga, D., & Basturk, B. (2007). A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm. *Journal of Global Optimization, 39*(3), 459-471.

McConnell, S. (1996). *Rapid Development: Taming Wild Software Schedules*. Washington: Microsoft Press.

Meng, E., Shoichet, B., & Kuntz, I. (1992). Automated docking with grid-based energy evaluation. *Journal of Computational Chemistry, 13*(4), 505-524.

Murray, G. (2005, 2005-09-07). Rotation About an Arbitrary Axis in 3 Dimensions. Retrieved 23 september, 2010. http://inside.mines.edu/~gmurray/ArbitraryAxisRotation/.

Neal, S., Berjanskii, M., Zhang, H., & Wishart, D. S. (2006). Accurate prediction of protein torsion angles using chemical shifts and sequence homology. *Magnetic Resonance in Chemistry, 44*(S1), S158-S167.

Petrescu, A. J., Calmettes, P., Durand, D., Receveur, V., & Smith, J. C. (2000). Change in backbone torsion angle distribution on protein folding. *Protein Science, 9*(6), 1129-1136.

Pham, D., Ghanbarzadeh, A., Koc, E., Otri, S., Rahim, S., & Zaidi, M. (2006). The Bees Algorithm – A Novel Tool for Complex Optimisation Problems. *Cardiff CF24 3AA(Intelligent Production Machines and Systems)*, 454-459.

Ramachandran, G., & Sasisekharan, V. (1968). Conformation of polypeptides and proteins. *Adv. Protein Chem, 23*(283), 438.

Siciliano, B., Sciavicco, L., Villani, L., & Oriolo, G. (2008). *Robotics: modelling, planning and control*: Verlag: Springer.

Teodorovic, D., Lucic, P., Markovic, G., & Orco, M. D. (2006). *Bee Colony Optimization: Principles and Applications. Proceedings of the 8th Seminar on Neural Network Applications in Electrical Engineering*, 151-156, Belgrade, Serbia & Montenegro.

Unger, R. (2004). The genetic algorithm approach to protein structure prediction. *Applications of Evolutionary Computation in Chemistry*, 2697-2699.