Is Abalone, Bio designer and Fold it, the best software for Protein Structure Prediction of AIDS Virus?

Rabail Hafeez Qureshi and Dr.Bushra Noman
SZABIST,
Karachi, Pakistan.

Abstract: - *This research paper answers the question that whether Abalone version 1.8.1, Fold It Beta Version and Bio designer version 1.0 can help in predicting the protein structure of AIDS. This study also explores the various problems and issues in these software and how can we build better and more efficient protein structure prediction software. It also highlights the various methods, approaches and problems in other protein structure prediction software that have been mentioned in various research papers.*

Keywords: - *Protein Structure Prediction, HIV, comparative modelling, Abalone, Bio designer, Fold It Beta Version.*

INTRODUCTION

Since the beginning of 21st century, scientists have been trying to conquer the quest of understanding DNA. Biotechnology, nanotechnology, genetic engineering and DNA computing, all these new fields focus on just one tiny element that is the basic structure of all life forms. According to definition, Genetic Engineering is defined as the artificial manipulation of any living thing's genetic material [1]. It is any process or technique in which a human modifies the genetic code of any organism through artificial means [1]. Basically, this technology deals with the manipulation of genes or DNA Molecules. On the other hand, Nanotechnology is defined as the study of manipulating matter on atomic or molecular level.

DNA has a very complex structure, although its code was deciphered by scientists many years ago, it is still a very complicated structure to handle. There is still a lot of information encoded within DNA which hasn't been deciphered. A single strand of DNA molecule, if stretched, can go around the world several times [1]. And hence we need computers to help us understand more about DNA.

All the technologies mentioned above rely on understanding and manipulation of DNA. These technologies can be used in many areas such as medicine and drug design, genetic modification of food and animals, in making biomaterials, special electronics and even forensics [1].

The most remarkable aspect is that in the past two decades, the world has witnessed a convergence of various technologies. Biological sciences, physics, computing, psychology, engineering all have converged with one another to address problems and issues that have bothered humanity since a very long time. A common factor in all these convergences is DNA analysis as scientists are trying to answer questions and reveal the mysteries of the universe by studying the universe at its molecular level.

It has become increasingly important that new software algorithms and tools should be designed that can help scientists in understanding and studying the DNA and protein structure of any kind of matter. Today there are thousands of such tools and software available in the world, but it has become extremely tedious to search for the ones that are fast, efficient and accurate.

Software that helps analyse a protein or DNA sequence is popularly known as "protein structure prediction software" [2]. The reason why this topic was chosen for Independent Study is that particularly in Pakistan, the use of these software is not very common. There are many students and biomedical researchers and scientists who are unaware of these software and their applications.

This research paper aims at analysing two protein structure prediction software and an innovative protein folding game that tries to teach the dynamics of protein structure in a fun way.

PROBLEM DOMAIN

There are hundreds of protein prediction software available. It becomes hard to distinguish between all of them and to know that which one suits our needs the most. Furthermore it is very difficult to know that which software gives us accurate and quick results and also has an easy-to-use interface that can be easily understood by a biologist having no experience of working on these software.

The problem that has been focused on in this study is that which software is the best for the protein structure prediction of AIDS virus. AIDS is a devastating disease that has inflicted and taken millions of lives. It is an incurable disease till date. Scientists are working around the clock to find out a cure for this deadly disease. The protein structure prediction of AIDS virus is a very crucial and difficult task as AIDS has an extremely complicated protein structure [5]. Many scientists have tried to simulate the protein structure of AIDS so that they can use it to perform experiments and try to search for a cure. Simulating the protein structure of this virus can help scientists in saving a lot of time and money. They can vary temperature and other factors using the software, and study the effects of the variations.

The problem of predicting the three dimensional structure of any protein is considered as the "holy grail of molecular biology" and is even thought to be equal to deciphering the "second half of the genetic code"[7]. New tools, techniques and algorithms are very much needed in this field so that structures of complex proteins can be successfully predicted which can aid in drug discovery.

In this study, three software have been used in order to predict the protein structure of AIDS. Abalone version 1.8.1 and Bio designer version 1.0 are open source, free software available for download on the internet. Fold It Beta Version is a protein folding game which is again open source and freely available for download on the internet. Using the three software, we have analysed that how successfully can the protein structure of HIV AIDS be achieved on a simple Dell laptop (Inspiron 1525).

This study not only shows as to how a protein structure prediction software can be used to produce 3D models of a virus so that it can be studied and experimented upon using simulation. This report also compares the 3 software with each other on the basis of system requirements, installation, time taken to produce results, help manuals and accuracy. Furthermore, the literature review highlights similar studies done by other researchers who have used different protein structure prediction software.

RESEARCH METHODOLOGY

The research methodology used in this study is a combination of analytical and empirical research along with methodological literature review. The PDB files and protein sequence of HIV AIDS were obtained from BLAST, which is an online tool.

After obtaining the required files and protein sequence, the three software were practically used to achieve results. Software manuals and online help was referred to, to understand the three software and how they work. Based on the results, the three software were compared to draw conclusion. The problems faced while using these software have been related to the ones that have mentioned in various research papers studied for the literature review.

Protein structure prediction of HIV AIDS was done on Abalone version 1.8.1 and Bio designer version 1.0 using two methods, which are:

1. Comparative homology modelling by using templates from PDB library
2. Abintio method by using primary sequence as an input.

Fold It Beta Version , which is a protein folding game, was analysed as to how effective it is in representing the protein structure prediction problem and how using various protein folding puzzles, the player can help in solving protein structures and eventually help in predicting protein structure of complicated viruses like HIV.

Based on the findings, suggestions have been made as to how we can make better protein structure prediction software.

What is Protein Structure Prediction?

Protein Structure prediction is defined as the prediction of the three dimensional structure of a protein from its Amino acid sequence [2]. This includes the prediction of primary, secondary and tertiary structure of the protein.

Protein structure prediction is a very important field in bioinformatics. It is being used in medicine to study different kinds of viruses and the effects of various enzymes and drugs on them [2]. Its use in medicine has led to and will in future lead to finding cures for diseases which have been termed as incurable since a very long time. Protein structure prediction is also being used in biotechnology to design enzymes and study various different proteins which are in turn used in fields like DNA computing and genetic engineering.

Protein structure prediction is a set of techniques which can help in predicting the structure of any kind of protein [2]. It has successfully predicted the structure of proteins like insulin and all the basic elements, but it still needs a lot of improvement so that it can predict the structures of more complex proteins like HIV and various forms of cancer.

In 1950, a scientist by the name of Ainfinsen, wrote a paper in which he described that the information needed to determine the protein structure of any kind of protein molecule is contained in the amino acid sequence [7]. Ainfensen's Thermodynamic Hypothesis said that the process of assembling proteins into their native structure is not by a biological process. It is rather a physical process which depends purely on the amino acid sequence [7]. This idea suggested the fact that three dimensional structure of any kind of protein can be predicted from its sequence only. Since this revolutionary idea was suggested, there have been many efforts where scientists have tried to solve the problem of protein structure prediction. Some researchers have tried to solve this problem from the aspect of biophysics, whereas others have used biological evolution and chemistry [7].

Experimental methods and techniques used in protein structure prediction such as high resolution electron microscopy, X-ray crystallography and nuclear magnetic resonance spectroscopy are extremely time consuming and expensive and sometimes not at all applicable [11]. Hence many scientists and researchers have been working since the 1960s to develop computational methods that can help in protein structure prediction [11]. Recently, a lot of improvement has been made but still there is a lot of room for improvement.

METHODS AND APPROACHES USED IN PROTEIN STRUCTURE PREDICTION

There are 4 main approaches being used in protein Structure prediction, which are:

1. Comparative or homology modelling
2. Fold recognition and threading methods
3. De novo or Ab initio methods
4. Hybrid or integrative methods

In homology or comparative modelling, the structure of a target sequence is determined by using the structures of proteins which belong to the same family of the target protein and have been determined using experimental methods. The protein structure obtained through experimental methods is used a template to model the target sequence or structure [11]. Homology modelling or comparative modelling is only possible if the required templates are available. The most important step in comparative modelling is to identify the most suitable template and to accurately align the sequence of that template with the target protein or sequence [11]. Comparative modelling techniques rely on three factors:

1. Availability of appropriate template structures [11].
2. Accuracy of alignment methods and their ability to accurately align template sequences with target sequences [11].
3. The structural and functional difference between the target and the template [11].

Fold recognition and threading methods are used to model sequences have which have no similarity to proteins that have a known structure and a template.

De Novo or Ab initio methods predict the three dimensional structure of a protein using only its primary amino acid sequence and the laws of physics that govern protein folding [11]. These methods rely on the energy functions rather than PDB files which contain templates of known protein structures.

Hybrid or integrative methods combine experimental and computational techniques to predict protein structures. These relatively new methods are still being improved worked on. [11]

WHAT IS CASP?

CASP stands for Critical Assessment of techniques for protein structure prediction. It's a worldwide experiment that is conducted every 2 years since 1994 [4]. This experiment test all the protein structure prediction tools, methods and servers and produces an assessment of each [4]. The assessment contains performance results and show the level of efficiency and accuracy of each tool and method.

CASP also releases the results and rankings for the public on the internet so that biologists and researchers can know that which method or tool is the ranked the highest for protein structure prediction [4].

This experiment is carried out in a blind fashion way. It is completely automated and neither the organizers nor the evaluators have any beforehand information about the protein structure being predicted [4]. The experiment used structures that already been solved or structures that currently being solved using experimental techniques such as x-ray crystallography [4]. The experiment uses both comparative modelling and ab initio modelling.

LITERATURE REVIEW

Yang Zhang(2008) discusses in his paper that the advancements in the development of methodologies for protein structure prediction has come to a stable state [10]. Although there has been a lot of progress in this field due to the use of multiple structure templates, there is still a lot of room for improvement. Zhang then states that software and tools that use a hybrid of knowledge based and physics based system do a much better job at predicting the structure of any protein, but it still is a challenge to predict structures for proteins larger than 150 residues [10]. Yang proposes the idea that based on the 40, 000 structure available in PDB library, 4 million models can be predicted by simply using a combination of comparative modelling techniques and BLAST search [10]. He suggests that the current PDB library is almost complete and can aid in solving the protein prediction problem [10]. He further demonstrates in his paper that the current template structure identification techniques can be further improved by including additional structural information [10]. According to Zhang's paper, ROSETTA has worked really well for free or ab initio modelling [10].

Zhang along with Skolnick, in their joint paper, prove that the current PDB library is sufficient to solve the problem of protein structure prediction only if efficient algorithms to recognize folds are developed [15]. They used a structure alignment algorithm to obtain target template pairs and then build a homology model using TASSER [15]. The results they obtained showed that the PDB library is sufficient enough to solve the protein folding problem [15]. They also used a program called Prospector, and applied the same techniques [15]. The results showed that there is a need for developing powerful fold recognition algorithms. They also suggested in their paper, that there is also a need to develop better structure alignment algorithms which are not fast but also produce accurate results [15].

JaroslawMeller(2005) suggests in his paper that researchers and scientists should use nature as a guideline in designing protein structure simulation and modelling protocols [14]. Meller proves in his paper that Dynamic programming algorithms can be used to find best alignment of protein sequences. He also proves that linear programming techniques can be used to improve the performance of scoring functions for threading [14]. Meller suggests that the best way to solve the protein structure prediction problem is to combine techniques from data mining, machine learning, protein chemistry, protein structure and functional genomics [14].

Bujnicki and Fischer (2004) suggest in their paper that if scientists use a combination of various different models and methods, they can end up with better structures [19]. They suggest that we shouldn't rely on the results of just one server because most servers cannot differentiate between weak hits and wrong hits [19]. They did an analysis of five Meta servers which are:

1. PMOD
2. PCON
3. ALEPHOJURY
4. Rosetta
5. 3D SHOTGUN

Their results suggested that Meta server do a much better job at predicting structures than the simple primary servers [19]. They concluded that the higher the n in meta n, the better the results would be. From all the five Meta servers the duo analysed, they concluded that Rosetta is the best and produces efficient and better results than others [19].

Cymerman, Feder, PawŁowski, Kurowski and Bujnicki(2004) suggest a new strategy in their research paper [12]. They suggest that multiple methods should be used while searching databases and the results should be compared [12]. Combination of various pair wise alignment methods for searching databases can produce much accurate results [12].

Wallner and Elfson did a performance evaluation of 6 homology modelling programs. They used a common sequence alignment and compared the three dimensional structures obtained from the 6 homology software. Their results showed that protein sequences have more than 30% sequence identity can be predicted efficiently.

Nayeem, Sitkoff and Krystek(2006) did an evaluation of 6 software, which are :

1. Insight
2. Prime
3. Profit
4. Look
5. ICM
6. Sybyl
7. CLUSTALW

They did comparative modelling and used proteins of sequences identities ranging from 19% to 76% [13]. They concluded from their results that for low sequences identities, the programs PROFIT and PRIME produce much better results [13].

Mihansan has done a comparison of tools and servers based on performance, accuracy of results, number of citations and ease of use [3]. He differentiates the servers and tools into 3 categories:

1. Standalone program

2. A server
3. Meta server.

Mihansan suggests that web servers have made a biologist's life much easier [7]. Biologists are free from the hassle of downloading and installing a tool or software that requires a lot of computational resources. Using a server means that all heavy computations are done somewhere else [7]. The biologist just has to input a protein sequence using a web browser and he will be emailed the results few days later. This method is extremely user friendly and requires less resource [7].

Mihansan(2010) then suggests that according to his results, meta servers are even better than simple servers [7]. Meta servers produce much accurate results as they obtain results from various different servers, combine the results and email the best results back to the biologist [7]. According to his paper, the Swiss Model is the most widely used server as it has the most number of citations [7]. Following Swiss Model is 3D- Jigsaw. According to Mihansan's results, Modeller is also good option as its open platform and can work on windows, Linux and Mac.

WHAT IS AIDS AND WHAT ARE THE PROBLEMS IN DISCOVERING AN ANTI AIDS DRUG?

AIDS stands for Acquired Immunodeficiency Syndrome, which is caused by the HIV virus. HIV stands for Human Immunodeficiency Virus [5]. AIDS will cause a human's immune system to collapse completely which results in the human acquiring life threatening infections. HIV AIDS is spread through the exchange of bodily fluids and use of contaminated syringes and needles.

Currently, around 40 million people all over the world suffer from this life threatening disease [5]. A person infected with HIV virus if left untreated can develop AIDS within 9 to 10 years [5]. Currently there is no cure available for this deadly disease. Scientists all over the world are trying to find a cure for this disease as it is spreading rapidly all over the world.

Protein structure of HIV consists of 3 main genes, namely, gag, pol and env [5].

It is very difficult to find a cure for AIDS or discover some kind of a vaccine that can work against it. The main reason is that the virus mutates very easily and rapidly into some many different forms and even if there was a vaccine against it, some parts of the virus will still remain resistant to it and the virus will keep on replicating and spreading.

Since it keeps mutating rapidly and has a complex structure, its protein structure prediction still hasn't been done accurately. Many scientists and researchers have tried different methods to achieve the protein structure of HIV so that they can design drugs.

Gerald and Myers (1996) show in their paper that every software has a specialty [16]. Some software is better at predicting beta sheets, while another can predict the helices accurately [16]. But there isn't any one method or software that can accurately predict all the helices, sheets, turns and the coils of the HIV protein structure [16]. And since the HIV protein structure is complicated and consists of above 120 residues, the problem becomes even worse [16]. They further suggest that this problem can be solved to some extent if structural information can be incorporated in the primary protein sequence of the virus.

VIF which stands for Virion Infectivity Factor is the agent necessary for the HIV to spread in the Human body [17]. The HIV itself cannot fight against the human body's anti-viral agents [17]. For it to spread throughout body successfully it needs the VIF [17]. VIF is the actual protein that fights the anti-viral agents of human body. Balaji, Kalpana and Shapshak produced a three dimensional structure of VIF using SWISS MODEL. VIF is a 190-240 amino acid protein [17]. They used comparative modelling technique and performed sequence alignment with the help of BLAST. They then submitted their results to the pdb library but those results are yet to be verified using experimental techniques [17].

In another paper, Andrianov and Anishchenko(2010) produced the protein structure of AIDS using a program called GROMACS [18]. They incorporated CYC B Peptide to study the effects. Their results showed that CYC B Peptide can be a possible candidate for anti AIDS drug discovery [18]. They however claim that their results still aren't accurate.

By using different techniques, methods, tools, servers and software, researchers all over the world are trying to produce an accurate protein structure prediction of HIV AIDS. They are facing a lot of problems. The HIV protein has a very complicated structure and can mutate easily. On the other hand, the computational protein structure prediction methods haven't been perfected as yet and cause bottlenecks which cannot be solved easily.

ABALONE VERSION 1.8.1

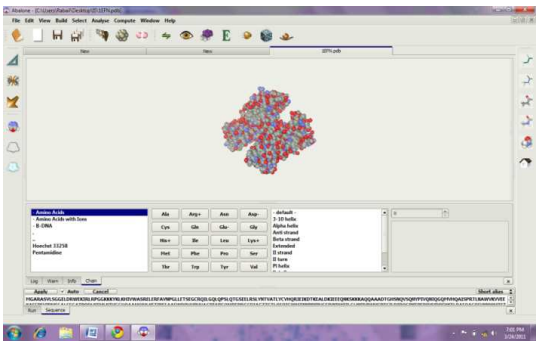
Abalone version 1.8.1 is bimolecular modelling software that offers comparative modelling and ab initio modelling of protein structures. It is standalone software and isn't connected to any web server. It's freely downloadable. It allows you to build a chain of amino acid sequence and also allows geometry editing. It also allows you to manipulate the model using molecular mechanics. You can do the following things using Abalone version 1.8.1:

- 3D modelling of proteins
- Simulation
- Analysis of temperature, pressure and dipole moment
- Compute intermolecular energy
- Molecular geometry optimization
- Combining models from several files
- Specify and check chemical models and force fields

Abalone version 1.8.1 can easily be installed on the windows platform. The comparative modelling of HIV AIDS was done first. The steps are as follows:

1. First input the HIV amino acid sequence obtained from BLAST. There is a blank space given at the end of the main screen for inputting the amino acid sequence.
2. The next step is to load the PDB structure templates that have been downloaded from the NCBI website.
3. Once the file is opened, you will see the protein structure of HIV on the screen

The steps are fairly simple and the user gets hold of it once he tries one or two times. Being from a non-biology background and someone who is a computer science graduate, after using this software I got to know a lot about molecular biology. At first it was quite hard to understand



the software and how everything works. A course in DNA computing helped me a lot, and once I knew the basics of molecular biology, I found my way through. Having said that, Abalone version 1.8.1 can easily do comparative modelling.

The next step was to test that can ab initio molecular modelling be done on Abalone version 1.8.1. The steps followed are given below:

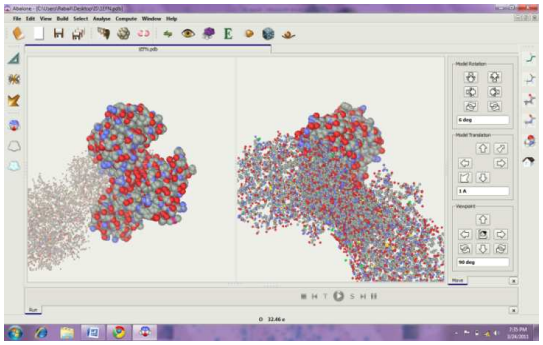
1. First step is to input the HIV protein sequence in the space given at the end of the main screen.

2. After inputting the amino acid sequence, the next step is to click on the apply button and wait for results. After waiting for about half an hour, the program stopped working and got hanged. The same method was tried 5 times and each time it was the same story. To check whether something wrong was being done, the same technique was tried with amino acid sequence of insulin. This time, the software responded within one minute and successfully produced the protein structure model of human insulin.

One of the major aims in this research was to try to incorporate a drug like paracetamol in the protein structure of HIV. After trying several methods, this was achieved using comparative modelling. The steps are as follows:

1. The first step is to input the amino acid sequence of HIV in the space given. After that the PDB structure template of HIV is loaded in the software.

2. The next step is to add the PDB structure template of paracetamol which was obtained from the NCBI website, using the add command. The result is shown below:



Paracetamol incorporated in HIV Protein Structure in Abalone version 1.8.1

BIO DESIGNER VERSION 1.0

Bio designer version 1.0 is a bimolecular modelling software that can be easily installed and run on personal computers. It can be downloaded for free and is a standalone program that can also do sequence alignment, a feature which is not present in Abalone version 1.8.1. It has the following features:

- Chain builder
- Aligning sequences
- Support for numerous file types
- Sequence processor
- Animation control
- Simulation

Comparative modelling and ab initio modelling were done on this software. The steps are as follows:

1. The first step is to input the HIV amino acid sequence in the sequence processor given below the screen, after inputting the sequence, open the structure template.

2. Once you click on open, you will get the protein structure of HIV on the screen.

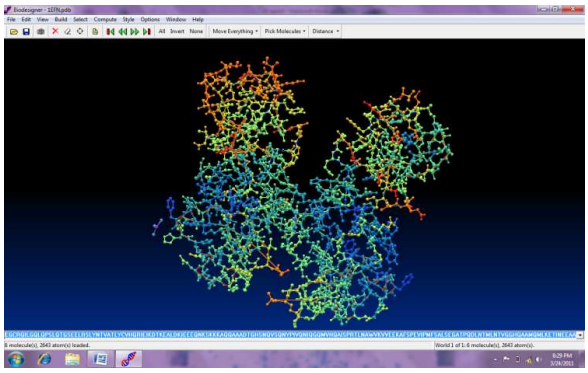


Fig. 3 HIV Protein Structure in Bio designer version 1.0

After doing comparative modelling, ab initio method was tried. The first step was the same as Abalone version 1.8.1, inputting the amino acid sequence in the sequence processor. Like Abalone version 1.8.1, Bio designer version 1.0 also stopped working and got hanged in the middle. When the same thing was tried with insulin, the program was able to successfully produce the protein structure model of human insulin.

The next step was incorporating paracetamol into the HIV model. The steps are as follows:

1. Input the HIV amino acid sequence and open the structure template to produce the protein structure of HIV.
2. After that, open the structure template of paracetamol and incorporate it with HIV.

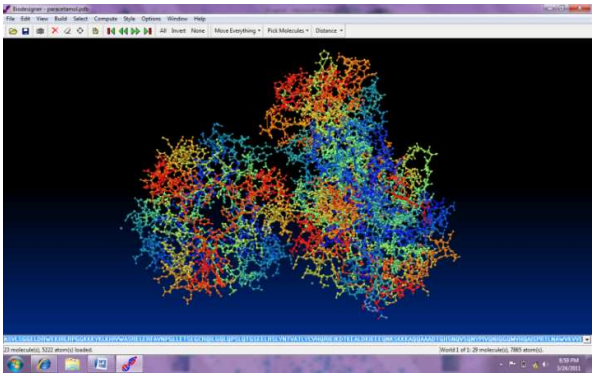


Fig.4 Paracetamol incorporated in HIV protein structure in Bio designer version 1.0

FOLD IT BETA VERSION

Fold It Beta Version is a multiplayer, experimental computer game based on protein folding. It was designed by David Baker and his team of researchers at the University of

Washington. Its first version was released in 2008. This game is part of the Games-with-purpose initiative as it helps in solving problems that the computer cannot solve or has difficulty in solving [20]. This game is based on popular protein structure prediction software called Rosetta. David Baker has used the same algorithm that was used in Rosetta.

Fold It Beta Version attempts to apply human brain's problem solving skills to the problem of protein structure prediction [20]. Around 60,000 people around the world are playing this game and trying to solve the protein structure prediction problem. Fold It Beta Version has series of puzzles through which players try to solve protein structures [20]. Since the players can be people who have absolutely no idea about microbiology, there is tutorial about the basics of protein folding and how to solve the puzzles given.

As each player solves the puzzle, he is taken on to the next level and given complicated puzzles to solve. The game can also be connected to the server, and players can share puzzles, chat with each other and even create and join a group to share experiences and results.

David Baker and his team, state in their research paper that through this game they have tried to make protein folding fun and interactive [20]. They have incorporated the social networking element in this game and added attractive visuals to catch the players' attention. They have made the game as addictive as possible so that players who don't even know a single thing about protein folding can solve puzzles and maybe one come up with a cure for diseases like HIV AIDS and cancer [20].

I downloaded this game from the foldit.org website and installed it on my computer.

It gives you the option of either playing offline or connecting to the Fold It Beta Version server so that you can have an updated set of puzzles and also get to interact with people. It also gives you the option of playing intro puzzles, science puzzles and contests. As a beginner, I played some intro puzzles. From a total of 29 puzzles; I have only been able to solve 3 successfully. The ones I have solved are known as backbone packing.

Fig. 5 Protein Folding puzzle of level 4

The above screen shot is of the 4th puzzle I have been trying to solve. By simply following the hints given, anyone can solve the puzzle and gain score. As we keep solving the puzzles, the level of difficulty increases. I was able to solve the first 3 puzzles very quickly, but got stuck on the 4th puzzle as the level of difficulty increased.

COMPARISON AND EVALUATION

TABLE II
COMPARISON OF SOFTWARE

	Abalone version 1.8.1	**Biodesigner version 1.0**	**Fold It Beta Version**
Ab Initio Modelling	×	×	×
Comparative Homology Modelling	✓	✓	×
Incorporating drugs into protein structures	✓	✓	×
Quality of graphics	Average	Excellent	Good
Animation	Average	Excellent	Good
User Friendly	✓	✓	✓
Tutorials and online help	✓	✓	✓

All the three software are freely downloadable and work well on the personal computer. All three are easy to use but these are some of the points which I think are worth mentioning:

- Abalone version 1.8.1 and Bio designer version 1.0, both are very good software for comparative homology modelling. I was able to successfully to comparative homology modelling of HIV, human insulin and the drug paracetamol.
- As far as the ab initio method is concerned, both the software failed at ab initio modelling of HIV. On the other hand, both the software can successfully do ab initio modelling of Human insulin. This shows that both the software cannot do ab initio modelling of any protein sequence that has more than 150 residues.
- A very good thing about Abalone version 1.8.1 and Bio designer version 1.0 is that both the software allow the user to integrate two models together to study that exactly what happens when two proteins combine. I successfully incorporated the protein sequence of paracetamol and insulin in to the protein structure of HIV. Being from a computer science background, of course I couldn't understand the biology behind it, but I can certainly train biologists as to how these software can used to achieve that.
- Abalone version 1.8.1 offers more features than Bio designer version 1.0 in terms of protein and molecule manipulation. It has features that can help a user in simulating temperature and pressure conditions. It also offers geometry optimization and also computing intermolecular energy.
- On the other hand, Bio designer version 1.0 offers excellent graphics and animation options. The user can choose from a number of different graphical views and animation options.
- Fold It Beta Version is indeed an interactive game that can get you hooked to it. I enjoyed playing it and was able to solve 3 puzzles, but I believe that they can improve the tutorial a bit more so that players can know what exactly they are doing. Although they have done a fantastic job in

making the tutorial as easy to understand as possible, but I think there is a room for improvement.

SUGGESTIONS

Based on my results and literature, I believe that following steps should be taken to solve the protein structure prediction problem of AIDS and other proteins:

- All software, tools and methods should incorporate Monte Carlo methods to improve the quality and accuracy of the results they produce. Although there are some software like Abalone version 1.8.1, that have incorporated Monte Carlo methods in the program, there is a need for other to follow in the same footsteps.
- Linear programming should be used to improve sequence alignment methods. If sequence alignment methods are improved, the protein structure prediction problem can be solved using the current PDB library.
- Standalone software should be connected to Meta servers so that results from different servers can be compared and analysed to produce the most accurate protein structure prediction.
- Protein structure prediction of complex proteins like HIV, SIV, VIF and cancer should be done using super computers like Blue Gene. Most tools and methods fail to predict the protein structure once the sequence goes above 150 residues. Using super computers or even parallel processing can solve this problem to much extent.
- Biotechnology should be made part of the computer science curriculum to educate computer science students about molecular biology and the problems faced by biologists and doctors all over the world.
- No standards or guidelines are available. Scientists can only rely on CASP, which itself focuses only on tools that are server based. CASP doesn't do experiments on standalone programs, and hence they are missing from the rating lists.

CONCLUSION

To conclude, Abalone version 1.8.1 and Bio designer version 1.0 can do comparative homology modelling of HIV AIDS and also incorporate the drug paracetamol in to it. But both the software cannot do ab initio modelling of amino acid sequences with residues above 150, but they can do successfully ab initio modelling of amino acid sequences with less than 150 residues. Fold It Beta Version is an excellent game that can help educate people all over the world about the protein structure prediction problem and give them a chance to find cure for diseases like AIDS and cancer. It is a simple game which can explore the depths to which a human brain can solve complex problems. Whether it can help in protein structure prediction of HIV AIDS, that is a question which will be answered in the coming few years when more and more people start playing this game.

There is still a long way to go. Ab initio modelling is still an impossible task for many software and tools. Meta servers have contributed a lot to the field of biotechnology, and they continue to perform well. But we still need to design more efficient algorithms that produce accurate results and build better protein prediction software so that one day scientists can even have a cure for not only HIV AIDS but also the key to achieve eternal youth and maybe even find the God particle.

REFERENCES

[1] Wikipedia.(2011).*Genetic Engineering* [Online]. Available: http://en.wikipedia.org/wiki/Genetic_engineering

[2] Wikipedia.(2011).*Protein Structure Prediction* [Online]. Available:http://en.wikipedia.org/wiki/Protein_structure_prediction

[3] Wikipedia.(2011).*List of Protein Structure Prediction* [Online].Available:http://en.wikipedia.org/wiki/List_of_protein_structure_prediction_software

[4] Wikipedia.(2011).*CASP*[Online]. Available: http://en.wikipedia.org/wiki/CASP

[5]Wikipedia.(2011).*HIV*[Online].Available:http://en.wikipedia.org/wiki/HIV

[6]NBCI.(2011).Available:http://blast.ncbi.nlm.nih.gov/Blast.cgi

[7] M. MIHĂŞAN. *BASIC PROTEIN STRUCTURE PREDICTION FOR THE BIOLOGIST: A REVIEW*. Arch. Biol. Sci., Belgrade, 62 (4), 857-871, 2010

[8] John Moult. *Rigorous performance evaluation in protein structure modeling and implications for computational biology*. Philosophical Transactions of the Royal Society. Published online 1 February 2006. Available: http://www.jstor.org/pss/20209645

[9] D. Cozzetto1, and A. Tramontano. *Advances and Pitfalls in Protein Structure Prediction.* Current Protein and Peptide Science, 9, 567-577.2008

[10] Yang Zhang. *Progress and challenges in protein structure prediction*. Current Opinion in Structural Biology , 18:342–348. 2008

[11] T. Schwede, A. Sali, N. Eswar and M. C. Peitsch. (2008). *COMPUTATIONAL STRUCTURAL BIOLOGY - Methods and Applications*[Online]. Available: http://www.worldscibooks.com/etextbook/6659/6659_chap01.pdf

[12] I.Cymerman, M. Feder, M. PawŁowski, M.A. Kurowski, J.M. Bujnicki. *Computational Methods for Protein Structure Prediction and Fold Recognition.* Nucleic Acids and Molecular Biology,Vol. 15.2004

[13] Akbar Nayeem, DoreeSitkoff, and Stanley Krystek, Jr. *A comparative study of available software for high-accuracy homology modeling: From sequence alignments to structural models.* Protein Sci.15(4): 808–824.2006

[14] JarosławMeller.(2005) *Advances in Protein Structure Prediction: Algorithms and Applications[Online]*.Available: http://folding.cchmc.org/publications/habilitacja_cz1.pdf

[15] Yang Zhang and Jeffrey Skolnick. *The protein structure prediction problem could be solved using the current PDB library.*ProcNatlAcadSci U S A. 2005 January 25; 102(4): 1029–1034. 2005

[16]] Gerald Myers and Andrew Farmer.(1996). *HIV Alignments, Database Searches, and Structure Prediction[Online]*. Available: http://hcv.lanl.gov/content/sequence/HIV/COMPENDIUM/1996/PART-III/6.pdf

[17] SeetharaamanBalaji ,RangaswamyKalpana and Paul Shapshak. *Paradigm development: Comparative and predictive 3D modeling of HIV-1 Virion Infectivity Factor (vif)*. Bioinformation 1(8): 290-309 (2006).

[18] Alexander M. Andrianov, Ivan V. Anishchenko. *Computer-assisted anti-AIDS drug development: cyclophilin B against the HIV-1 subtype A V3 loop.* Health Vol.2, No.7, 661-671 (2010).

[19] J.M. Bujnicki, D. Fischer. *'Meta'Approaches to Protein Structure Prediction.* Nucleic Acids and Molecular Biology,Vol. 15. 2004

[20] Seth Cooper, FirasKhatib, AdrienTreuille, Janos Barbero, Jeehyung Lee, Michael Beenen, Andrew Leaver-Fay, David Baker, ZoranPopovic&Foldit players. *Predicting protein structures with a multiplayer online game*. Nature, Vol. 466. Pages: 756-760. 2010.