STRUCTURAL SYNTHESIS OF SMALL MOLECULE DRUG CANDIDATES

Metin Türkay
College of Engineering and Center for Computational Biology & Bioinformatics
Koç University, Istanbul, 34450, Turkey

Abstract

This paper introduces an optimization algorithm to synthesize small molecule drug candidates. A drug molecule acts on an active site of the target proteins in order to regulate the functioning mechanisms of the target; this regulation is achieved either by activation of inhibition. A necessary condition for the regulation by a drug molecule is the packing of the active site. The three-dimensional structure of the molecules can be determined using to experimental methods. The optimization algorithm synthesizes the drug molecule at the active site of the target protein using the basic assumptions of quantum chemistry. An important feature of the algorithm is the ability to synthesize a variety of drug candidates rather than a single molecule. The effectiveness of the algorithm is illustrated with an example.

Keywords

Drug design, structural biology, mixed-integer programming.

Introduction

Diseases in biological organisms are caused by failure of a protein (or a set of proteins) to perform its function in general. Biological and chemical activities of organisms can be regulated by introducing a drug to the biological system that will target the protein responsible for the particular disease of interest. The regulation of the biological system is established when a drug molecule interacts with the active site of the target protein. The objective in drug design is to find a suitable molecule that will bind to the active site of the protein strongly for the regulation of biological and chemical activities of biological systems.

The drug design problem has been addressed by experimental and computational methods. The experimental methods concentrate on high throughput screening methods; it is possible to test about 1 million chemical substances for interaction with a protein active site using robotics integrated systems and combinatorial libraries. The main difficulties with the experimental methods include: small number of available chemical substances for testing, high experiment costs, and the possibility that the chemical substance may be interacting with another active site of the target protein. Computational simulation methods test chemical structures stored in databases for binding to the active site of the protein. These methods are based on testing for activity on the target protein. The drawbacks of this method include: availability of only a very small number of chemical structures in databases for evaluation (databases contain only a very small fraction of vast number of possible molecular structures), databases often do not contain suitable molecules for the target protein, and the availability of the same database to every database user. Therefore, it is necessary to develop new strategies for drug discovery and design that will overcome the drawbacks of the traditional methods.

Drug Discovery Process

Drugs are chemical substances that are administered to people using a variety of delivery methods in order to treat or prevent diseases. The discovery of a single drug takes about 16 years and cost more than $802 on the average (PhRMA, 2003).

Most of the diseases in biological organisms are caused by failure of a protein (or a set of proteins) to perform its function. Biological and chemical activities of organisms can be regulated to prevent diseases by

introducing a drug to the biological system that will target the protein that is responsible of the particular disease of interest. The regulation of a biological system is established when a drug molecule interacts with the active site of the target protein. The drug discovery process is comprised of three main steps: target selection, lead identification and optimization and clinical trials. These main steps include different stages as shown in Figure 1.

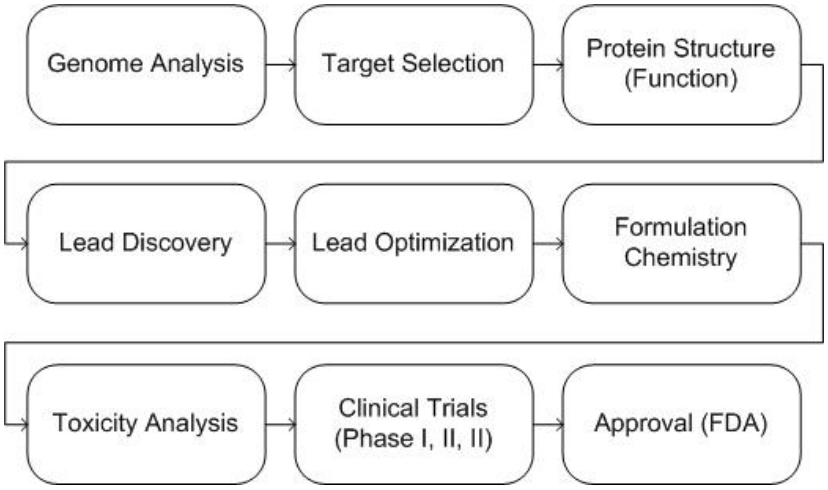

Fig. 1. Schematic representation of drug discovery process.

Target Selection

Major chemical and biological activities in organisms are initiated, catalyzed, carried out or terminated by biological entities like enzymes or hormones that are basically polypeptide chains. These biological entities have special sequences that determine their unique properties. The term "genome" means all set of genes of a biological organism. Genes are part of chromosomes that are located in the nucleus of the cell carrying heredity information that are responsible for coding at least one protein. Diseases in biological organisms are caused by the failure of a protein (or a set of proteins) to perform its (their) function in general. Therefore, the gene that codes the particular protein that plays an important role in the disease should be analyzed. When a new drug discovery process is initiated, the "target" gene or protein that plays a central role on the onset and progress of the disease of interest is identified. After the underlying problem resulting with the disease and all related symptoms is determined, the aim is to change, improve, inhibit or completely block concerned function of the target, depending on the way to prevent illness. The protein that is coded by the target gene can be analyzed to investigate appropriate chemicals that could interact with it to cure the disease. Analysis of a protein does not only mean discovery of its chemical structure or sequence of substructures in its chains but it means discovery of its functions, as well. Functions of proteins and their interactions with other substances in an organism are determined by three-dimensional structures and configurations. Proteins function through their active sites (also called binding site), which are the specific regions playing central role in proteins' interactions with other molecules. The drugs prevent the onset or progress of a particular disease by binding to the active site of target proteins. Therefore, understanding the functions and structures of proteins is indispensable in drug discovery process.

Lead Discovery and Optimization

The next step in the drug discovery process is to identify drug candidates that will interact with the active site of the target protein. In general, a large number of chemical substances that may be suitable drugs are tested on the target protein to observe the effect for interaction. The chemical substances that show any level of interaction are categorized as a hit. The drug candidates interact with the active site to have therapeutic effect by blocking, accelerating, decelerating, reversing or initiating reactions depending on the way to cure the disease. This phase ends up with a list of drug candidates interacting successfully with the target in laboratory conditions.

Drug candidates determined in lead discovery phase are further refined using experimental and computational methods. The result is a list of best performing substances to be active pharmaceutical ingredient for the drug. Various factors including maximum therapeutic efficiency, minimum side reactions with other substances or structures that can exist in the environment of the target or other regions of the organism, are considered in the selection process. Sometimes, the best performing drug candidate can

involve in reactions with other structures that can exhibit side effects and lead to even other diseases. Also, the drugs enter the blood after they are approved by liver; so it should be assured that the selected drugs do not severely affect or destroy this organ. These facts are analyzed through computations and experiments with relevant substances trying to imitate the real environment as much as possible. Result of this phase is not always the very best performing substance on the target but it is the most suitable drug candidates considering both therapeutic power and various other effects on all organs and systems of the body.

Clinical Trials

With proven success on laboratory animals with metabolic and hormonal systems as similar to humans' as possible, the drug candidates are tested on patients and volunteers in hospitals. It usually requires trials on thousands of people before exact design of the drug is completed and this particular phase is divided into three sub-phases each of which can only be initiated after approval. The clinical trials are conducted in three phases, one after successful completion of the other. If a drug passes all three phases successfully, an application to regulatory agencies is made for marketing the drug.

The structure of the target protein can be used to design drug that can be effective in curing or preventing a disease (Kuntz, 1992). Recent developments in experimental methods to determine the protein structure is reviewed and its role in drug design is discussed (Matsuzaki, 2004). The successful drug discovery examples are discussed (Henry, 1991) and the state-of-the-art in structure-based drug design is discussed in a recent review paper (Anderson, 2003). The structure-based drug design process start with a promising initial molecular structure whose efficiency is known either experimentally or predicted using computational methods. However, it is very important to get an initial molecular structure that could be effective in binding to the active site of the target protein. The objective of this paper is to present a synthesis approach to find promising candidates for structure-based drug design.

Synthesis Approach

The drug design problem has been addressed by experimental and computational methods. The experimental methods concentrate on high throughput screening methods; it is possible to test about 1 million chemical substances for interaction with a protein active site using robotics integrated systems and combinatorial libraries. The main difficulties with the experimental methods include: small number of available chemical substances for testing, high experiment costs, and the possibility that the chemical substance may be interacting with another active site of the target protein. Computational simulation methods test chemical structures stored in databases for binding to the active site of the protein. These methods are based on testing for activity on the target protein. The drawbacks of this method include: availability of only a very small number of chemical structures in databases for evaluation (databases contain only a very small fraction of vast number of possible molecular structures), databases often do not contain suitable molecules for the target protein, and the availability of the same database to every database user. Therefore, it is necessary to develop new strategies for drug discovery and design that will overcome the drawbacks of the traditional methods.

The synthesis of chemical structures that will bind to the selected active site of the target protein is addressed in this paper. The synthesis method assembles the drug molecule at the active site of the protein atom-by-atom to have a close fit. The method works on the following assumptions:

i. three dimensional structure of the active site for the target protein is known
ii. all atoms are perfect spheres with known diameter
iii. the distance between two atoms depend on the type of the bond between these atoms (single, double, or triple bond)
iv. the atoms can establish known covalent bond types (single, double, or triple bond)

The effectiveness of a drug molecule highly depends on how closely it binds to the active site of the target protein. Therefore, the method initializes the drug molecule by placing properly selected atoms at the functionally important parts of the active site in order to establish a strong binding. These atoms are then connected to each other considering the diameters of the atoms and the types of the bonds between these atoms given in Tables 1 and 2 (Tsai et. al., 1999 and Tsai et. al., 2001). Since the 3D geometry of the active site is known a geometric constraint satisfaction problem is solved with combinatorial assembly of the building blocks (atoms and bonds between them) of drug molecule. In addition, bioavailability is also included to have an upper bound on the size of the drug molecule that is being synthesized. The algorithm generates alternative drug molecules by changing atom and bond types. As a result, a number of molecular structures that will bind to the active site of the protein are synthesized. Following are the key advantages of the synthesis method:

i. fundamental assumptions of quantum chemistry are the only assumptions
ii. novel chemical structures that are not included in the databases and combinatorial libraries can be synthesized
iii. a number of chemical structures as drug candidates are generated

iv. the method is general and can be used to synthesize molecular structures that will bind to active sites of different proteins

Table 1 Bond types for different atoms.

Bond Type	Bond Type
-C≡C-	≡C-NH_2
=C=C=	≡C-NH_3^+
≡C-C≡	=C=O
=C=N-	≡C-OH
≡C-N=	≡C-S-
=C=NH-	≡C-SH

Table 2. Van der Waals Radii for different atoms.

Atom	Radius(A)	Atom	Radius (A)
=C=	1.61	-NH_2	1.64
=CH-	1.76	-NH_3^+	1.64
=CH_2	1.88	=O	1.42
-CH_3	1.88	-OH	1.46
=N-	1.64	-S-	1.77
=NH	1.64	-SH	1.77

Example

The proposed method is illustrated on a protein that plays a central role in pancreatitis (Whitcomb et. al., 1996). The protein data bank (PDB) contains number entries that give the three-dimensional structure of the protein that is responsible for pancreatitis; among the given structures 1TPA is selected in this study. The active site of 1TPA is located between resides 189-195 as shown in Fig. 2 and synthesis approach is applied to find structurally attractive drug candidates. A promising drug candidate is synthesized at the active site of the target protein, 1TPA, as shown in Fig. 3. The synthesized drug candidate that is marked with red for easy differentiation in the active site, has a molecular weight less than 500.

Conclusions

In this paper a novel synthesis approach for drug candidates is presented. The approach uses the three-dimensional structure of the active site of the target protein that is responsible for onset and progress of the disease of interest. The synthesis approach makes the fundamental assumptions of quantum chemistry and synthesizes the drug molecules using the available atoms, all possible bonds that these atoms make and the Van der Waals radii between these atoms depending in the type of the bond. The efficiency of the synthesis approach is illustrated with an example.

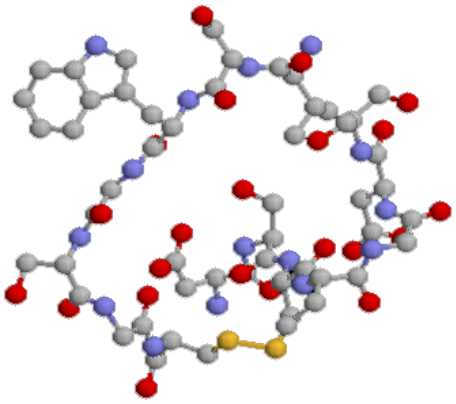

Fig. 2. The active site of protein 1TPA.

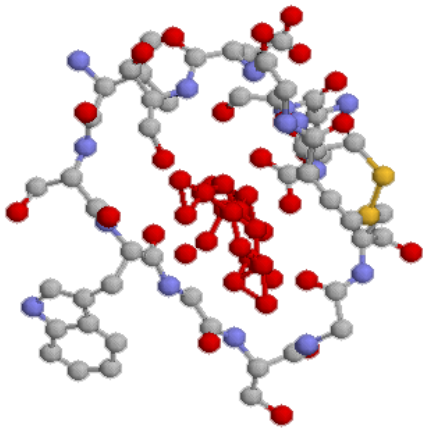

Fig. 3. The synthesized drug candidate inside the active site of protein 1TPA.

References

Anderson, A. C. (2003). The Process of Structure-Based Drug Design, *Chemistry & Biology*, **10**, 787-797.

Henry, C. M. (2001). Sructure-Based Drug Design, *Chem & Eng. News*, **79** (23), 69-74.

Kuntz, I. D. (1992). Structure-Based Strategies for Drug Design and Discovery, *Science*, **257**, 1078-1082.

Matsuzaki, T, (2004). Protein Crystallography: A "Must" Technology for Drug Design, Intl. Symp. On Portable Syncrotron Light Sources and Advanced Applications, Kusatsu, Japan.

PhRMA, (2003). Pharmaceutical Research and Manufacturers Association of America Annual Report 2003-2004, Washington D.C.

Tsai, J., R. Taylor, C. Chothia, M. Gerstein (1999). The packing density in proteins: standard radii and volumes, *J. Mol. Biol.*, **290**, 253-266.

Tsai, J., N. Voss, M. Gerstein (2001). Determining the minimum number of types necessary to represent the sizes of protein atoms, *Bioinformatics*, **17**, 949-956.

Whitcomb, D.C., R.A. Preston, C.E. Aston, M.J. Sossenheimer, C.D. Ulrich, S.P. Martin, L.K. Gates Jr., S.T. Amann, P.P. Toskes, R. Liddle, K. McGrath, G. Uomo, J.C. Post, G.D. Ehrlich (1996). Hereditary pancreatitis is caused by a mutation in the cationic trypsinogen gene, *Nature Genet.,* **14**, 141-145.