[S3-4] [4/20/2001(Fri) 11:15-11:40 / Hall A]

Lead Informatics Using Protein 3D-Structures

Seonggu Ro,* Dong-Kyu Shin, Hui-Jong Han, Young Ho Jeon, Eui-Jun Jeong, Kwang Yeon Hwang, Hye-Yeon Kim, Yong-Seok Heo, Tae Gyu Lee, Jin Hwan Kim

CrystalGenomics, Inc. 461-6, Jeonmin-dong, Yusong-gu, Daejeon, 305-390.

Drug discovery process includes target identification, target validation, hit generation, lead identification, lead optimization, in vivo efficacy, pre-clinical studies, clinical studies, and finally followed by drug approval. Recent achievement in human genome project introduced many new concepts as well as techniques into the drug discovery. These new research areas are bioinformatics, functional genomics, proteomics, pharmacogenomics as well as toxicogenomics. The lead informatics has also been emerged in order to manage the knowledge derived from the process of hit generation, lead identification and lead optimization. This system could accelerate the high-output generation of drug-candidates.

Most drug molecules interact with therapeutic target proteins in our body to execute their desirable or undesirable physiological effects. If we get the 3-dimensional structural information regarding the interaction between the target protein and the druggable molecule, the drug discovery process will be accelerated. As a starting point of our efforts in lead informatics, we established protein folding dictionary based on the 3D structures of a variety of disease related proteins. The subsequent steps such as virtual HTS, HTS-NMR and structural chemogenomics enabled us to generate drug candidates rapidly which could recognize active site of the protein.

The structural chemogenomics, which plays a major role in lead informatics, can be initiated by grouping the target proteins into their unique family according to the sequence homology, followed by structural analysis of the active sites as well as the interaction ones. Using virtual HTS, we then select the chemical moieties that can recognize specifically the target site of the each protein. After collecting compounds based on this information, we can build potential drug libraries for each therapeutic target family. Since the resulting libraries are classified based on the target family, folding states of the drug targeting sites and chemical classes of scaffolds, we only need to screen the specific chemical library for each target protein, not the whole library to discover lead compounds. Namely, structural chemogenomics can be described as the collection and grouping of all possible drug lead compounds directed to all drug targets (as expected, 3,000 ~ 4,000 proteins coded by the human genome).

This new technique becomes getting more attention in the area of high-throughput-screening (HTS) as described recently (Modern Drug Discovery 2001, Jan). Although the HTS techniques have been tremendously innovated for the last ten years, it has been reported that the increasing number of hits do not reflect the actual number of lead compounds, suggesting that new informatics system is required to improve the productivity of lead compounds.

In this presentation, we will discuss about the concept of lead informatics and the related techniques and share part of our experience..