Protein Structure Modeling and Drug Discovery

By Teresa Overskei

Genomics & Bioinformatics
Due 12.13.02

The past several years have seen a shift from the more traditional "wet chemistry" drug discovery process to "*in silico*" processes, which take advantage of the data manipulation and molecular modeling prowess of modern computers. The resulting implications for drug discovery and development are unprecedented. While traditional, top-down drug development techniques took as long as 40 years to return a viable drug product from a basic understanding of the target protein's function, molecular computer-based techniques today allow a drug to be developed in only three years, as was the case with the chemotheraputic agent Herceptin®.[1]

Structural modeling and database screening in particular has had a similarly profound affect on alternative drug discovery methods. The human genome project has unveiled about 30,000 unique genes, any of which could be a potential new drug target. Furthermore, it is estimated that there may be between 10 and 100 slight variants of each gene, effectively complicating protein characterization.[2] Currently, very few of these known protein sequences are actually accompanied by a known 3D protein structure, but the gap is closing. Frequently tagged as "the major challenge of the post-genomic era,"[3] the translation of sequence data to structural knowledge has gained significant ground in the last several years. Considering recent improvements as only the tip of the iceberg, structural modeling and comparisons appear to be the next significant tool for more efficient and profitable drug discovery.

Computational Structure Modeling Methods

Since the first protein structure was determined in 1957, structure analysis has evolved from simple determination to modeling. X-ray crystallography was the primary means by which a physical molecular structure was determined experimentally, but this method is not widely applicable – only one in twenty proteins form useful crystals for structure studies, and some, especially membrane proteins and protein complexes, form no crystals at all.[4] Rather than force physical structure determination methods on a compound in an attempt to yield larger volumes of structural information, science has taken a more practical and rational approach through computer-based structure modeling integrated with physical data. Protein folding cannot be predicted from sequence alone,

however, as there are so many variables to take into account that most advanced computers are too slow to achieve this in a reasonable amount of time.

Homology modeling and threading technologies, which rely on already existing knowledge of 3D protein structures to infer structure from sequence, have been used in lieu of autonomous folding models. Threading methodologies are used in the limiting case when there is insufficient information for regular homology modeling to be possible. The inaccuracy in the inputs results in imprecision in the output: threading yields multiple structure predictions that supply a useful picture of how a protein may be structured only when taken into account together. Overall, though homology and threading techniques are useful for basic function elucidation, they are too approximate for drug design.

Another method for structure analysis, NMR, combines physical distance determinations with computational refinements. Unfortunately, this technique becomes less effective with larger target molecule sizes. Dynamic analyses of protein structure also represent the new front in computational methods. Rather than focusing solely on the 3D structure of a protein, it is increasingly common to model structure variation with time, or 4D structure modeling. This is achieved through simulation of the changes in molecular motions and properties of a protein over time.

Accuracy and Uses of Structure Modeling

Structure modeling methods are being continuously refined and now approach the accuracy levels of more traditional methods of structure determination. Homology methods are nearing X-ray crystallography resolution: in a comparison between the X-ray crystal protein structures and the homology-derived structures of San Diego-based Structural Bioinformatics Inc., it was found that about 96% were within 1.5Å or better, and 98% were within 2.0Å. Although the accuracy of the model declines as sequence similarity decreases, homology modeling methods are still effective at as low as 20% sequence similarity. At SBI-Advanced Technologies, a Denmark- based company, secondary structure prediction from sequence using homology modeling has an accuracy of about 80%.[5] Threading technologies, on the other hand, are less accurate – threading

predicted structures differ from the X-ray crystallography structure by 3.2Å – 17.6Å, whereas two crystal structures would typically differ from 1.5Å – 1.8Å.[6] A drawback of both methods is that they are most effective predicting the protein core structure are less accurate for surface structures, which is where most protein interactions and activity takes place.

Modeling is itself very flexible, and characteristics of the surroundings or neighboring compounds can be applied to the modeling process in order to predict structure in a particular situation. This flexibility is plainly absent from physical structure determination methods: X-ray and NMR structures alone represent the behavior of a protein in a specific environment and may not accurately reflect reality. Protein-protein and protein-ligand interactions such as docking can also be modeled, an exceptionally difficult task because proteins change shape in response to the presence of another molecule.

Structure modeling is also extremely useful for prediction of protein function. Because structure dictates molecular behavior, it can be directly correlated to a compound's activity and function within living systems. This reasoning can be traced in reverse, so that when a compound with a particular function is sought (i.e. binding to a specific molecule with a known structure), the desired function can be related to structure, which is in turn predicted from sequence. This enables potential new drugs to be identified based on sequence alone.

Benefits to Drug Discovery

Out of the 30,000 human genes known, drugs on the market today target only about 500 different proteins.[7] Hence, the structural and functional understanding derived through modeling of this tremendous amount of untargeted proteins can be exploited for drug development purposes.

Structure-based computational screening is the most efficient method of sorting out potential drug targets or potential new drugs to interact with a given target. For example, when a high-quality structure of a target protein is available, it can be used to computationally screen protein and small molecule databases for a compound that binds

to the active site. A screen can pare down a list of thousands of potential targets to a more manageable number, and once these compounds with a high likelihood of binding have been singled out, optimization of the leads can take place. Such structure-driven screens have resulted in active hit rates of about 10%, in stark contrast with hit rates of about 0.01% with conventional high-throughput screening techniques.[8] Structural Bioinformatics uses computational technology in combination with X-ray crystallography to model a protein's structure and produce lead drug targets in as little as 60 days.[9]

When structural screens are used to nominate potential drug and drug target leads for further analysis, money and time is saved. Quicker and cheaper than conventional methods of lead selection, computational screening based on structure gives a company an advantage in drug discovery. Edward Maggio, CEO of Structural Bioinformatics, notes that the ability to screen for the most optimal drug targets means higher success rates, which in turn translates into reduced costs: "If you can go from a 15% success rate to a 25% rate, you'll save millions of dollars."[10]

An increasing number of partnerships and alliances are forming between pharmaceutical companies and structure modeling companies to take advantage of this edge in drug development, not to mention the financial incentives. In October 2002, for example, MDS Pharma Services Inc. of Montreal set up a strategic alliance with California-based Iconix Pharmaceuticals Inc. in order to use MDS's structure recognition technology to identify likely drug failures.[11]

Looking Ahead

The main factor holding back the efficient use of structural databases for both research purposes and drug discovery is a lack of database infrastructure and information transferability. Because there are few standards for database design and computing platforms, data cannot be shared and searched in the fastest and most comprehensive way. Many technology firms, having recognized this hole in the bioinformatics field, have sought to develop novel interfaces to fill the need for a common computational platform. This would allow organizations to pool their talents and accelerate the drug discovery process.

On the other hand, the efficient interplay between databases may be a moot point one day if computing power was increased such that stand-alone folding models would be practical. If computing advances continue to move at a rapid pace, with performance doubling every 18 months, then it is estimated that by 2010 the prediction of the folding patterns of simple proteins would be possible.[12] However, this time span is too long to ignore the advantages of integrating physical and modeled structures that are available today.

Bibliography

Augen, Jeffrey. "The Evolving Role of Information Technology in the Drug Discovery Process." Drug Discovery Today 1 March 2002.

Gwynne, Peter. "Probing for Patterns in the Data Deluge." Genomics and Proteomics 1 Nov. 2002.

Maggio, Edward T. and Ramnarayan, Kal. "Recent Developments in Computational Proteomics." Drug Discovery Today 1 Oct. 2001.

Munger, Frank. "In Biology, Proteins are Key; Predicting their Structure Could Speed New Drugs, Medical Breakthroughs." Knoxville News-Sentinel 21 Oct. 2002.

"Plexus Vaccine Inc and GlaxoSmithKline Biologicals Sign Structure-Based Antibody-Vaccine Research Agreement." PR Newswire 22 July 2002.

"Structural Bioinformatics Inc. and AtheroGenics Announce Drug Discovery Collaboration." Business Wire 14 Oct. 2002.

Endnotes

[1] Augen

[2] Maggio

[3] Maggio

[4] Maggio

[5] Maggio

[6] Maggio

[7] Maggio

[8] Augen

[9] "Structural Bioinformatics Inc. and ….."

[10] Gwynne

[11] Gwynne

[12] Augen