CAMD in modern drug discovery

What does the future hold for CAMD in drug discovery research in the light of advances in genomics, HTS and combinatorial chemistry?

In recent years, genomics, high-throughput screening (HTS) and combinatorial chemistry have emerged as important platform technologies in modern drug discovery research. Each addresses particular needs: genomics provides important information on potential therapeutic targets, HTS technologies establish rapidly whether there are any active hits (or chemical leads) in existing compound collections (or libraries) and combinatorial chemistry promises to provide and expand such libraries both for lead discovery and optimization, rapidly and on a large scale.

For the past decade, computer-aided molecular design (CAMD) techniques have been viewed by some as a major technological platform in drug discovery. However, it is probably worthwhile to evaluate the relationships between the above relatively recent technologies and CAMD for the purpose of developing an efficient and effective discovery research paradigm. Does CAMD still have a role to play in drug discovery research, and if so, what is it to be?

The term drug discovery as used here refers to the discovery of small organic molecules intended to modulate a biological receptor or mechanism. The process of drug discovery research can be broadly described in terms of three stages:

- target identification,
- lead discovery of one or more compounds (that interact selectively with the relevant biological target, and
- optimization of the potency, safety and other physicochemical properties of the lead, to produce a development candidate that is ready to be taken into man.

Genomics research clearly plays a key role in the target identification stage. Bioinformatics approaches are providing the solution to the management of the vast amount of sequence data produced by genomics research, and are establishing a database system that will allow drug discovery scientists to ask important questions regarding the functional role of a biological molecule and its relationships with other biological molecules[1]. Bioinformatics has also contributed to our understanding of protein structure and the structural complementarity between a receptor and a drug that underpins their binding. The scientific importance of protein structure prediction is heightened by the fact that the number of known protein sequences (> 60,000) is far greater than the number of proteins (approximately 1,000) for which the 3D structure is known[2], and by the difficulties and uncertainties associated with the determination of protein structures by experimental means, such as X-ray crystallography and NMR. Although significant recent advances have been made in protein structure modelling, accurate prediction of protein structure remains largely an area of academic research.

At the lead discovery phase, HTS has become an integral and routine part of discovery research programmes where rapid and reliable assays are available for the chosen therapeutic targets, provided that the organization involved can afford the technology. Several important parameters are emerging in the use of HTS technologies. A significant advantage is speed; an advanced HTS system can carry out around 20,000 tests in a week. For a compound collection of a typical pharmaceutical company, the hit rate is generally in the region of 1%. If a compound collection has 200,000 compounds, a typical screening would generate around 2,000 active compounds (e.g. IC_{50} or K_i < 200 μM). Traditionally, such compounds have been accumulated by in-house synthetic effort and from natural product extracts over a period of years. Combinatorial library synthesis will continue to expand the compound collection.

If a suitable assay is available for HTS, it is almost certain that HTS will be the method of choice without any computational screening or selection. If the assay for the chosen target is not suitable for HTS, it is highly desirable and necessary to limit the number of compounds assayed, and chemical database searching techniques are therefore useful in this context. Important advances have been made in chemical database searching; for example, conformationally flexible searches whose ability to find active hits based on a good hypothesis is well known[3].

Clearly, how to make use of the information generated by HTS programmes is a challenge. Considerable effort is needed to decide which active hits should be followed up if there is no clear lead among the active hits. CAMD techniques such as molecular clustering, similarity and diversity analysis will be helpful.

There is also the problem of how to make use of activity data from the active hits to design further combinatorial libraries. 3D QSAR techniques[4] can be used to construct a pharmacophore

Jin Li, Proteus Molecular Design Ltd, Proteus House, Lyme Green Business Park, Macclesfield, Cheshire, UK SK11 0JL. tel: +44 1625 500555, fax: +44 1625 500666

  1359-6446/96/$15.00. PII: S1359-6446(96)20007-8

which, in principle, can be used to direct library design coupled with molecular diversity analysis[5,6]. Terms such as molecular similarity and diversity have been used widely in the context of combinatorial library design. It is important to consider such properties in terms of 3D structures. It is also helpful not to consider combinatorial library synthesis as a single monolithic concept. For example, libraries synthesized for screening against a wide range of therapeutic targets are very different from a focused library targeted against a specific biological receptor. The former requires diversity, whereas the latter requires that similarity to the lead structure is maintained while also achieving diversity. In the context of drug discovery, apart from their amenability to combinatorial synthesis, the compounds in both types of libraries must satisfy some general criteria, such as molecular weight constraints, chemical stability, conformational flexibility/inflexibility and lipophilicity.

Where the 3D structure of a biological target is known, new structure-based drug design techniques have proven useful in producing active compounds[7], and are potentially very useful in providing chemical insight for combinatorial library design[8] and in suggesting specific compounds to synthesize. The challenge lies in achieving an effective combination of structural constraints, physicochemical properties and synthetic accessibility in the chemical designs produced. Most drugs interact with their receptors in a specific structural arrangement, hence use of the available 3D structural information of a target to direct the lead discovery must be preferred over the brute-force, random synthesis approach.

Structure-based ligand design techniques have improved significantly in recent years[9]. They are usually referred to as *de novo* design programs in the literature. Such programs usually operate in three stages:

1) The binding site information of a target receptor is used to define the likely characteristics of the ligands in terms of hydrogen bonding patterns, hydrophobic binding regions and charge centres based on molecular recognition rules derived from analysing X-ray crystallography databases.
2) Databases of 3D structures of small and medium-sized organic molecules or fragments are searched in order to find potential ligands that can satisfy the binding characteristics defined previously. This process often involves extensive conformational analysis and the construction of novel structures by connecting predefined fragments together in chemically sensible ways. Many structures are usually generated at the end of this process.
3) The chemical structures generated in the previous stage are evaluated according to a scoring function, which takes account factors such as interaction energy, conformational flexibility and chiral centres. Human ingenuity will play a great part in the evaluation stage. It is conceivable that scoring functions will be improved as more experimental data become available and could be extended to include other important drug properties such as detection of toxicophores. Coupled with combinatorial synthetic principles, these CAMD techniques will be very productive in lead discovery.

With all of these techniques currently applied to lead discovery, it is not difficult to see that the bottleneck will shift markedly towards the lead optimization stage. It is well known that considerable uncertainty and difficulty exist in designing *in vivo* properties such as bioavailability, safety and metabolic stability. However, experimental effort to develop simple screening tests will generate large volumes of systematic data for a wide range of compounds. Once these data become available in a consistent and reliable format, computational techniques such as QSAR can be used to analyse them, and the resulting knowledge and rules learnt can be fed into the designs at an earlier stage.

It is becoming clearer that drug discovery must be approached in a multidisciplinary way incorporating biomedical research, bioinformatics, HTS, structural biology, combinatorial chemistry, medicinal chemistry, computational chemistry and pharmaceutical sciences[10]. CAMD's role in drug discovery should be viewed with realistic expectations. Within that context, techniques developed in CAMD in past decades, such as chemical database searching, molecular property calculation (log*P* and solubility), 3D QSAR and pseudoreceptor construction, structure-based ligand design and associated binding affinity calculation, can make important contributions to lead discovery and the design of combinatorial libraries. When 3D structural information about a therapeutic target and its substrate or ligand is available, CAMD's role can be very significant in both lead discovery and lead optimization. It is also important for CAMD research scientists to realize the opportunities offered by combinatorial chemistry and HTS to make more data available more rapidly than before, and to respond with high throughput computational studies made possible by advances in software and hardware. Greater effort should be put into the collection and organization of *in vivo* data for a range of chemical compounds and their analysis by CAMD techniques so that a better understanding of the relationships between chemical compound structure and properties of bioavailability, metabolism and toxicity can be gained at early stages of discovery research.

Jin Li

REFERENCES

1 Karp, P.D. (1995) *J. Comp. Biol.* 2, 573–586
2 Eisenhaber, F., Persson, B. and Argos, P. (1995) *Crit. Rev. Biochem. Mol. Biol.* 30, 1–94
3 Ashton, M.J., Jaye, M.C. and Mason, J.S. (1996) *Drug Discovery Today* 1, 71–78
4 Green, S.M. and Marshall, G.R. (1995) *Trends Pharmacol. Sci.* 16, 285–291
5 Martin, E. *et al.* (1995) *J. Med. Chem.* 38, 1431–1436
6 Sheridan, R.P. and Kearsley, S.K. (1995) *J. Chem. Inf. Comput. Sci.* 35, 310–320
7 Bohacek, S.R., McMartin, C. and Guida, W.C. (1996) *Med. Res. Rev.* 16, 3–50
8 Combs, A.P. *et al.* (1996) *J. Am. Chem. Soc.* 118, 287–288
9 Clark, D.E., Murray, C.W. and Li, J. in *Reviews in Computational Chemistry* (Vol. 11) (Lipkowitz, K.B. and Boyd, D.B., eds) (in press)
10 Kubinyi, H. (1995) *Pharmazie* 50, 647–662