J Comput Aided Mol Des (2007) 21:307–309
DOI 10.1007/s10822-007-9115-1

De novo design and synthetic accessibility

Johann Gasteiger

Published online: 28 June 2007

The design and development of a new drug is an extremely costly and time-consuming process. Therefore, in recent decades novel experimental and computational techniques have been introduced to make this process more efficient. Bioinformatics and chemoinformatics methods have become an integral part of the drug design progress. At the outset of the design of a new drug stands the quest for a novel structure that may carry the desired biological property. Both structure- and ligand-based methods have been developed to find new lead structures. Structure-bases methods need knowledge on the three-dimensional structure of the target protein. Then, either a *de novo* design method or a docking procedure can be applied. In *de novo* design, molecules are grown into the binding pocket of the protein so as to fill this pocket and provide sites for binding the ligand to the protein. In a docking experiment, three-dimensional structures are taken from a database and are fitted into the binding pocket so as, again, to fill this pocket and satisfy sites for binding.

Ligand-based methods do not need knowledge on the three-dimensional structure of a protein; in fact, they even do not need knowledge on what protein is actually involved for targeting. Rather, ligand-based methods try to find new lead structures by deriving knowledge from known ligands to this protein and use this knowledge to search in a database of structures, either by pharmacophore or by similarity searching.

De novo design methods can look back on a long history; the first systems were developed by Jeff Howe at Upjohns (Moon) and by Hans-Joachim Böhm at BASF (Ludi). Since then, several other systems have been designed.

One of the drawbacks of *de novo* design systems is that they may generate a large number of structures, some of them quite complicated and therefore difficult to synthesize by requiring long-step syntheses with carefully controlled reactions, in particular as stereochemistry is concerned.

It should also be mentioned that docking procedures and ligand-based methods might run into the same kind of problems if large databases containing virtual structures are used. As, eventually, these novel structures have to be synthesized and tested, a costly process, it becomes imperative to rank these structures according to their ease of synthesis, according to synthetic accessibility, in order to minimize the efforts in the laboratory. If a reasonable ranking is found that is also accepted by medicinal chemists trust into computational methods can be generated. Thus, chemoinformaticians and medicinal chemists will work closely together, a process that is so essential for success in drug design.

The American Chemical Society and the Chemical Structure Association have recognized the importance of *de novo* systems and the crucial importance of estimating synthetic accessibility and have asked me to organize a symposium on ''*de novo* Design and Synthetic Accessibility'' at the American Chemical Society Spring Meeting in Atlanta, GA, in March 2006.

Unfortunately, not all major players in these fields could be gained to participate in this symposium but the program could provide a good overview of the state of the art.

The program consisted of eight presentations:

– *Structure and reaction based evaluation of synthetic accessibility*. Krisztina Boda, Thomas Seidel, Achim Herwig, Oliver Sacher, Johann Gasteiger

J. Gasteiger (✉)
Computer-Chemie-Centrum, University Erlangen-Nuremberg, Erlangen, Germany
e-mail: gasteiger@chemie.uni-erlangen.de

- *Algorithms and cancer drugs: In silico design of S1008 ligands to block p53 binding*. John L. Whitlow, Yumin Li
- *Closing the loop: From high-throughput screening to synthesis of novel protein displacers*. N. Sukumar, Curt M. Breneman, Steven M. Cramer, James A. Moore, Kristin P. Bennett, Mark J. Embrechts, Min Li, Jia Liu, Long Han
- *De Novo design tools for the generation of synthetically accessible ligands*. A. Peter Johnson, Krisztina Boda, Shane Weaver, Aniko Valko, Vilmos Valko
- *FlexNovo: Structure-based searching in large fragment spaces.* Jörg Degen, Matthias Rarey
- *Concepts in receptor optimization: Targeting the peptide RGD.* Wei Chen, Chia en Chang, Michael K. Gilson
- *Generating and searching > 10E20 synthetically accessible structures*. Richard D. Cramer, Farhad Soltanshahi, Robert Jilek, Brian Campbell
- *ROBIA: Computational assessment of synthetic procedures.* Jonathan M. Goodman, Ingrid M. Socorro

The first talk was presented by **Krisztina Boda** (University of Erlangen-Nuremberg). She described a scoring method that rapidly evaluates synthetic accessibility of structures based on structural complexity, similarity to available starting materials and assessment of strategic bonds where a structure can be decomposed to smaller components. The scoring function was validated with estimations obtained from experienced medicinal chemist. The estimates of medicinal chemists varied remarkably and the estimates provided by the method presented here were within the range of these variations showing that they were of equal quality.

John L. Whitlow (East Carolina University) showed an application of a *de novo* design method to find a molecule with high binding affinity for the active binding site of S100B. The presented study includes screening a library of fragment molecules against the target active site and using the top-ranked fragment as a scaffold to design complete ligand molecules.

Nagamani Sukumar (Rensselaer Polytechnic Institute) showed that the identification of efficient displacers is a major challenge for protein displacement chromatography. An efficient procedure was presented that quickly predicts novel selective displacers using machine learning models. A small set of known selective displacers is used to train the model that is then utilized to identify novel selective displacers from available chemical catalogs.

A. Peter Johnson (University of Leeds) discussed the latest methods of his research group, which has a long tradition in developing *de novo* design methods. The problem of synthetic accessibility of *de novo* design structures is addressed by either incorporating synthetic constraints into the build-up process (virtual synthesis inside the receptor cavity) or post-screening the set of proposed structures by applying complexity or retrosynthetic analysis. The complexity analysis is based on statistical distributions of various cyclic and acyclic topologies and atom substitution patterns present in starting material catalogs and drug databases.

Jörg Degen (University of Hamburg) presented a new molecular design program, called FlexNovo. The program incrementally builds putative ligands within the receptor pocket incorporating pharmacophore constraints. Synthetic accessibility is ensured by utilizing building blocks that are derived from fragmenting known drug structures by common synthetic reactions. He showed several applications where the program was able to reproduce known inhibitors but he also highlighted the deficiencies of present scoring functions.

The afternoon session began with a presentation by **Wei Chen** (University of Maryland). He showed a study that uses an accurate and theoretically well-founded method of computing binding affinity. It was used to design novel receptors targeting biologically important peptides. The putative receptors constructed by fragment-based design software were presented and analysed.

Richard D. Cramer (Tripos) examined the limited novelty and content of commercially offered reactant databases and presented a very rapid similarity/QSAR-based ligand searching method. He showed that topomer space similarity can be used to forecast biological similarities. Topomer description of molecular structure is an effective ligand-based approach to predict and identify novel scaffolds within lead discovery libraries.

In the last talk of the symposium, **Jonathan Goodman** (University of Cambridge) emphasized the importance of accurately predicting reaction outcomes. The program ROBIA assists chemists in planning syntheses and in predicting favourable reaction pathways. The program uses a mechanistic approach to reaction prediction, generating possible intermediate structures from reactants to the products, combining general knowledge of organic chemistry with molecular modelling.

The talks presented and the lively question-answer sessions indicated that *de novo* design methods play a major role in drug discovery. The main deficiency of the method, that is the synthetic accessibility of the suggested structures, is presently addressed by various research groups and remarkable results have been achieved.

The presenters were asked to submit manuscripts for this special issue of the Journal of Computer-Aided Molecular Design on this topic. Unfortunately, not all presenters could be won to submit a manuscript. However, it is hoped that those contributions that were submitted and are contained in this issue give a good overview of the state of art

in the areas of *de novo* design and the estimation of synthetic accessibility.

It should become clear that the solution of these problems requires knowledge from quite different areas and asks for the collaboration of scientists with different backgrounds. In this context, the project Novobench, funded by the German Minister of Education and Research (BMBF) is worth mentioning. In this project, two academic groups from the University of Hamburg and the University of Erlangen-Nuremberg, two SMEs, BioSolveIT and Molecular Networks, as well as three pharmaceutical companies in Germany, E. Lilly, Hamburg, Altana, Konstanz and 4SC, Munich are collaborating in further developing a *de novo* design system (see contribution by Jörg Degen, et al.) and in developing methods for estimating synthetic accessibility (see contribution by Krisztina Boda, et al.) and evaluating the methods in an industrial environment. This project shows that the collaborative effort of several groups allows the solution of complicated problems.

I want to thank the American Chemical Society and the Chemical Structure Association for initiating and supporting this symposium. Furthermore, my thanks go to the presenters on this symposium and particularly to those that submitted manuscripts to this special issue. Last-not-least, I want to thank Terry Stouch and the editorial staff of the Journal of Computer-Aided Molecular Design, in particular Naomi Portnoy for initiating this special issue and helping in bringing it to success.