

Attenuating attrition

By C. Simone Fishburn, Senior Editor

High drug attrition rates from poor safety have spawned numerous efforts to use *in silico* methods to improve molecule design, but none of the algorithms created so far has emerged as a true game changer. Now, **AstraZeneca plc** and **Roche** have concluded that better prediction requires more data, and they are pooling their information via an intermediary cheminformatics company, **MedChemica Ltd.**, to produce a new set of design rules.

In 1997, Christopher Lipinski created the first widely recognized set of rules to guide the optimization of physicochemical properties, such as solubility and lipophilicity, and facilitate the generation of drug-like compounds.

Since then, computational chemists have created numerous algorithms to aid molecule design, most of which have improved upon Lipinski's rule of five. Fewer advances have been made that help medicinal chemists optimize biological properties such as toxicity and ADME.

MedChemica, created by three former AstraZeneca scientists, believes it can jump-start progress by analyzing how changing a molecule's structure affects its behavior in biological assays.

Although other algorithms try to relate structure to biological function, most of the analyses look at modifications across a wide array of diverse structures. MedChemica's approach is to look at modifications in a set of similar structures and see how minor differences affect the compounds' biological activity.

Al Dossetter, managing director of MedChemica, said the advantage of the company's platform is the WizePairZ algorithm that looks at pairs of fragments that are similar in structure but differ by a chemical group, such as a change from chlorine to fluorine or the addition of a methyl group.

This platform, he told *SciBX*, captures the chemical environment of the fragment change. For example, it incorporates the fact that the effect of changing chlorine to fluorine on a molecule will depend on the surrounding structure. The result is a rule that is context dependent.

The MedChemica approach applies to small molecules and uses only partial chemical structures, thus keeping compound identities out of the picture.

Because the platform does not reveal compound identities, AstraZeneca and Roche can share knowledge without disclosing proprietary information.

By collaborating to produce a new set of design rules, the pharmas are aiming to reduce lead optimization time in drug discovery, during which compounds from screening hits are modified to improve their pharmacokinetics and reduce their potential for toxicity.

The lead optimization process generally requires the synthesis of 500–1,000 molecules that differ by minor chemical modifications and undergo testing in a battery of preclinical assays. With the aid of an improved set of rules, they hope to reduce the number of compounds that need to be synthesized to reach the optimal clinical candidate.

Although each pharma has rich compound libraries with millions of molecules and large amounts of experimental data, AstraZeneca and Roche believe that detecting significant trends requires even greater statistical power, which will come with consolidating their databases and increasing the number of matched pairs.

Dossetter said smaller databases only allow researchers to extract one to five matched pairs, which have a low fidelity of prediction. Ten matched pairs are sufficient to draw a prediction, but reliability increases significantly with 20 matched pairs.

The MedChemica database contains 1.2 million datapoints, each of which represents a single molecule fragment in a single assay. It includes 31 different assays, although more are likely to be added in the future, and not all molecules have been tested in all assays.

Prior to partnering with MedChemica, AstraZeneca and Roche each had large libraries of compounds with corresponding experimental data. MedChemica is compiling those data into a single integrated database.

The proportion of the full MedChemica database contributed by each partner was not disclosed. The Roche data includes molecules and results from studies at its **Genentech Inc.** unit.

Compatibility of the datasets from Roche and AstraZeneca was key to forming the collaboration. The two companies are in discussions with other big pharmas that also may join the partnership and would need to provide complementary datasets.

Mike Snowden, head of discovery sciences innovative medicines at AstraZeneca, said that the compatibility requirements for joining are having a database with a diverse compound background, large numbers of molecule pairs and biological data that are produced by similar techniques to those of Roche and AstraZeneca.

"By pooling datasets, we believe predictions of the platform will get better and better," he told *SciBX*.

Each partner will get a copy of the resulting database, dubbed the Grand Rule Database. MedChemica also will offer molecule optimization consulting services to nonpartners.

Snowden is not concerned about losing a competitive advantage or leveling the playing field. Although the rules can benefit all medicinal chemists, he said success in drug development depends on having a good molecule at the starting point.

The principal limitation of the collaborative effort may be that the assay data are entirely based on *in vitro* and cell-based experiments. As yet, no *in vivo* data either from animals or clinical trials have been included. Since *in vitro* data often do not translate directly to results in the clinic, the rules MedChemica derives may shorten the time to selecting a clinical candidate but may not alter its chances of success in human trials.

"By pooling datasets, we believe predictions of the platform will get better and better."

—Mike Snowden,
AstraZeneca plc

Snowden acknowledged that the tool may have limitations but said the goal is to learn how to predict the compounds that should not be made.

The shared database should be up and running by year end.

Fishburn, C.S. *SciBX* 6(26); doi:10.1038/scibx.2013.647

Published online July 11, 2013

COMPANIES AND INSTITUTIONS MENTIONED

AstraZeneca plc (LSE:AZN; NYSE:AZN), London, U.K.

Genentech Inc., South San Francisco, Calif.

MedChemica Ltd., Newcastle-under-Lyme, U.K.

Roche (SIX:ROG; OTCQX:RHHBY), Basel, Switzerland