Abstract

At present, the combination of high drug development costs and external pressure to lower consumer prices is forcing the pharmaceutical industry to innovate in ways unlike ever before. One of the main drivers of increased productivity in research and development recently has been the application of computational methods to the drug discovery process. While this investment has generated promising insights in many cases, there is still much progress to be made.

There currently exists a dichotomy in the types of algorithms employed which are roughly defined by the extent to which they compromise predictive accuracy for computational efficiency, and vice versa. Many computational drug discovery algorithms exist which yield commendable predictive power but are typically associated with overwhelming resource costs. High-throughput methods are also available, but often suffer from disappointing and inconsistent performance.

In the world of kinase inhibitor design, which often takes advantage of such computational tools, small molecules tend to have myriad side effects. These are usually caused by off-target binding, especially with other kinases (given the large size of the enzyme family and overall structural conservation), and so inhibitors with tunable selectivity are generally desirable. This issue is compounded when considering therapeutically relevant targets like Abelson Protein Tyrosine Kinase (Abl) and Lymphocyte Specific Protein Tyrosine Kinase (Lck) which have opposing effects in many cancers.

This work attempts to solve both of these problems by creating a methodology which incorporates high-throughput computational drug discovery methods, modern machine learning techniques, and novel protein-ligand binding descriptors based on backbone hydrogen bond (dehydron) wrapping, chosen because of their potential in differentiating between kinases. Using this approach, a procedure was developed to quickly screen focused chemical libraries (in order to narrow the domain of applicability and keep medicinal chemistry at the forefront of development) for detection of selective kinase inhibitors. In particular, five pharmacologically relevant kinases were investigated to provide a proof of concept, including those listed above.

Ultimately, this work shows that dehydron wrapping indeed has predictive value, though it's likely hindered by common and current issues derived from noisy training data, imperfect feature selection algorithms, and simplifying assumptions made by high-throughput algorithms used for structural determination. It also shows that the procedure's predictive value varies depending on the target, leading to the conclusion that the utility of dehydron wrapping for drug design is not necessarily universal, as originally thought. However, for those targets which are amenable to the concept, there are two major benefits: relatively few features are required to produce modest results; and those structural features chosen are easily interpretable and can thereby improve the overall design process by pointing out regions to optimize within any given lead. Of the five kinases explored, Src and Lck are shown in this work to fit particularly well with the general hypothesis; given their importance in treating cancer and evading off-target related side effects, the developed methodology now has the potential to play a major role in the development of drug candidates which specifically inhibit and avoid these kinases.