Drugs, structures, fragments : substructure-based approaches to GPCR drug discovery and design

Horst, E. van der

Summary

This thesis is all about cheminformatics, and its impact on drug discovery.

In **chapter 1** subjects discussed in this thesis were introduced. Some thoughts on drugs and drug discovery *per se* and how developments in informatics and computer science offer new opportunities, were brought forward. Aspects of chemical space, both real and virtual, were addressed. We also introduced the realm of compounds and targets, with specific emphasis on the most important drug targets of today, the so-called G protein-coupled receptors (GPCRs).

In **chapter 2** we reviewed a number of computational strategies to dissect molecules into sets of constituting atoms, leading to fragments of different nature. The reason for doing these, often computationally intensive, operations is found in the wealth of information that can be gleaned from such analyses. Virtual and real-world compound libraries can be mined for their diversity and/or similarity. In addition, the 'synthetic habits' of medicinal chemists can be explored. Furthermore, occurrence and co-occurrence of fragments may suggest new directions into chemical space. Fragments that appear linked to side effects can also be identified. This may help the medicinal chemist in designing safer or more selective lead compounds. Conversely, desired activities can be linked to fragments, and such information may be a decisive factor in a successful medicinal chemistry program. With both the large number of HTS campaigns being performed and the resulting data increasingly being made available in the public domain, it is anticipated that steadily more dedicated datasets will become available for fragment mining. Rule- and knowledge-based design efforts will certainly benefit from this.

In **chapter 3**, we conducted frequent substructure mining to identify structural features that discriminate between ligands that do bind to G protein-coupled receptors (GPCRs) and those that do not. In most cases, special chemical

representations resulted in the most significant substructures. Substructures found to be characteristic for the background control set reflected reactions that may have been used to construct this library. Alkane amine substructures were identified as most important for GPCR ligands, e.g. the butylamine substructure, often linked to an aromatic system. Hierarchical analysis of targeted GPCRs revealed well-known motifs and new substructural features, e.g., the imidazole-like substructure common for the histamine binding receptor ligands.

In **chapter 4**, we compared a sequence-based classification of receptors to a ligand-based classification of the same group of receptors. At the same time we evaluated the potential to use sequence relatedness as a predictor for ligand interactions thus aiding the quest for ligands of orphan receptors. We presented a classification of GPCRs that is purely based on their ligands, complementing sequence-based phylogenetic classifications of these receptors. Targets were hierarchically classified into phylogenetic trees, for both sequence space and ligand (substructure) space. The overall organization of the sequence-based tree and substructure-based tree was similar. The similarities and differences with traditional sequence-based classifications were investigated: our ligand-based classification uncovered relationships among GPCRs that are not apparent from the sequence-based classification. This may shed light on potential cross-reactivity of GPCR ligands and will aid the design of new ligands with the desired activity profiles. In addition, we linked the ligand-based classification with a ligand-focused sequence-based classification described in literature and proved the potential of this method for de-orphanization of GPCRs.

In **chapter 5**, a virtual ligand-based screening approach was designed and evaluated for the discovery of new A_{2A} adenosine receptor (AR) ligands. Several screening models were constructed by deriving the distinguishing structural features from selected sets of A_{2A} AR antagonists, so-called frequent substructure mining (see also **chapter 3**). The best model in statistical terms was subsequently applied to large-scale virtual screening of a commercial vendor library. This resulted in the selection of 36 candidates for acquisition and testing. Of the selected candidates, eight compounds

significantly inhibited radioligand binding at A_{2A} AR at 10 µM, corresponding to a "hit rate" of 22%. This hit rate is quite comparable to recent target-based virtual screening studies, while both approaches yield new, non-overlapping sets of ligands.

In **chapter 6**, a novel multi-objective evolutionary (MOE) *de novo* design method was developed and applied in this work to the discovery of new antagonists for the human A_1 AR. This method consists of several iterative cycles of structure generation, evaluation and selection. We applied an evolutionary algorithm (the so-called Molecule Commander) to generate candidate A_1 AR antagonists, which were evaluated against multiple criteria and objectives consisting of high (predicted) affinity and selectivity for the receptor, together with good ADMET properties. A pharmacophore model for the A_1 AR was created to serve as an objective function for evolution. We finally obtained a huge collection of 3.946 unique compounds from which we derived chemical scaffolds. Six of these were selected for actual synthesis and subsequently tested for activity towards all adenosine receptor subtypes, two of which were active in the (sub)micromolar range. To further investigate our evolutionary design method, we performed systematic modifications on one of these two scaffolds. We observed that an increased affinity with appreciable selectivity for A_1 AR over the other adenosine receptor subtypes was achieved through substitution of the scaffold.

In **chapter 7** we arrived at the general conclusions of my research and the future perspectives I foresee. Cheminformatics was the leading principle in this thesis. It led to the retrieval of active molecules from databases, it provided hints for de-orphanization procedures, and was pivotal in the automated design of novel chemical entities, overall corroborating its value for drug discovery. For the future a more 'open' approach to drug discovery seems mandatory, as the old paradigms do not seem to deliver a sufficient number of new drugs for current medical needs.