Chapter 1

Introduction

The drug discovery process

Drug discovery is a time consuming and expensive process. Estimates of time of currently bringing a new drug to market is between 10 and 17 years and cost on average $1.8 billion.¹ The drug discovery process is schematically presented in Figure 1. The first step is the investigation of the biochemical, cellular and pathophysiological mechanisms behind a certain disease. Understanding the cellular pathway is critical for identification of a molecular drug target. As soon as a target has been validated to have a potential impact for a special disease, the search for compounds that interact with the target starts. The medicinal chemistry part, screening, hit to lead and lead optimization, aims to identify and optimize compounds towards a potential drug. This process is described in more detail in the next paragraph. In the preclinical phase the compound is tested on animals to determine dosing and drug safety before it goes to human testing (clinical phase I/II/III). If the drug shows a beneficial effect in relevant patient groups without major side effects the drug can get the final approval and marketing can start.

Figure 1. Drug discovery pipeline. Biological part (target identification and validation) aims to select an appropriate disease related target. The medicinal chemistry part (screening, hit to lead and lead optimization) aims to develop potential drug compounds that interact with the selected target and show good pharmacokinetic and safety properties. Computational medicinal chemistry provides useful tools to guide medicinal chemists during this drug discovery procedure. In the development part (preclinic and clinic I/II/III) drug safety and effectiveness is tested. Finally registration authorities approve the drug for marketing.
**Medicinal chemistry:**

The initial screening process aims to identify hits that can be used as starting points for further optimization. The classical high throughput screening (HTS) is a very common approach for hit identification.\(^2\) In this process a big compound library is screened against a certain target. The screening library contains drug size compounds ideally including the drug candidate. As an alternative method for HTS fragment-based screening (FBS) has evolved as an attractive method for hit identification.\(^3_{-12}\) The heavy atom count of compounds in a fragment library is usually below 22 heavy atoms. The smaller size of the screening compounds usually leads to higher hit rates.\(^9\) For this reason FBS is now often applied as an alternative method. The fragment hits are subsequently ‘grown’ to drug candidates during the optimization process.

As soon as affinities of hits have been validated certain compound classes are selected for further optimization towards a drug candidate (hit to lead). Selection criteria are of course good affinity but also other properties apart from that are important. Molecules have to show good ADMET parameters. ADMET stands for absorption, distribution, metabolism, excretion and toxicology. During the lead optimization process both affinity and ADMET properties are optimized in an iterative cycle of compound synthesis and testing, which can be guided by computational medicinal chemistry (**Figure 2**).

![Figure 2. Lead optimization cycle](image-url)

---

\(^2\) [Reference](#)
\(^3\) [Reference](#)
\(^9\) [Reference](#)
Computational medicinal chemistry

The aim of computational medicinal chemistry is to support and guide the drug discovery process. Only 1 out of ~10,000 synthesized and tested compounds will reach the market. This represents an enormous investment in terms of time, and resources. Time and cost-effective decisions are highly appreciated before expensive synthesis and/or compound testing is started. For medicinal chemist simple guidelines are very useful for drug design. Additionally, computational methods for property predictions, modeling and virtual screening have been developed which can be included in different stages in the drug discovery process (Figure 1).

Guidelines for medicinal chemists

The awareness that drugs do not only need to show sufficient affinity but also need to have good physicochemical properties has lead to the development of simple guidelines like the Lipinski’s ‘Rule of 5’. This rule states that poor absorption or permeation of a molecule is more likely when there are more than 5 H-bond donors, 10 H-bond acceptors, the molecular weight (MW) is greater than 500 Da and the calculated octanol-water partition coefficient (ClogP) is greater than 5. It had been shown that MW and ClogP influence a lot of ADMET related parameters that are tested during drug optimization process. Too high MW for example is a reason for low absorption and low bioavailability and too high lipophilicity is the major reason for low solubility, off-target binding and toxicity. Out of these insights the ‘ADMET rule of thumbs’ was proposed, stating that a lot of ADMET properties will be more desirable if MW < 400 and ClogP < 4. Drug optimization therefore aims to increase affinity while controlling MW and lipophility to avoid so-called ‘molecular obesity’. In this sense several scores that combine affinity with MW and ClogP, like ligand efficiency and ligand-lipophilicy efficiency, are proposed. Such scores aim to be a better decision maker for hit selection and optimization than affinity on its own (see also Chapter 3).

Property prediction

Prediction of properties is cheaper and faster than most of the experimental assay methods. Therefore, large databases of compounds are often tested in silico before they – or better, subsets of them – are submitted to in vitro testing. Properties of interest are affinity, ADMET parameters and physicochemical properties like ClogP, solubility, melting point and pKa. Several computational methods for prediction of properties of a given
molecular structure are described and include linear models where molecular descriptors are directly related to the property of interest, or machine learning techniques like support vector machines or artificial neuronal networks. Machine learning techniques are normally more accurate but only provide a black-box model for the prediction on the property of interest.\textsuperscript{23} Having accurate property prediction in hand might be useful for decision making but it might be even more useful for medicinal chemist to understand the underlying design principles to achieve a certain property (like low Clog, low melting point, high solubility,…). In this sense a quite recently described method, matched molecular pairs analysis, is a very appropriate tool to directly relate a change in the structure to an effect on a certain property.\textsuperscript{24-31} A demonstration how matched molecular pairs analysis can be applied to estimate the effects of a change in descriptors on the melting point is presented in Chapter 4.

For the target of interest also affinity can be predicted. This can be done purely ligand-based by relating structural changes to a change in affinity (Free Wilson Approach).\textsuperscript{32} A more rational design can be enabled if also the 3D structure of the target is available which allows an estimation of important protein ligand interactions.

**Virtual screening**

Virtual screening is the computational search for new drug candidates with desired biological activities in large computer databases of small molecules that do not even have to exist physically.\textsuperscript{33} It is a fast and cost effective process that has evolved as an integral part in the drug discovery process. It is often applied either to reduce the size of a screening library or to completely replace high throughput screening, which is frequently the case in academia. In case the target 3D structure is available docking and scoring can be applied to rank the compounds. Ligand-based techniques (see Chapter 5) provide alternative and complementary approaches particularly when the target structure is not available but active ligand molecules are at hand.\textsuperscript{34} They are based on the nearest neighbour principle, that similar compounds to known active compounds might be also active.\textsuperscript{35} Several success stories in structure- as well as ligand-based virtual screening are reported showing the effectiveness of these methods to identify novel hits.\textsuperscript{36-39}

**Molecular modeling:**

Knowing the binding mode of active compounds is essential for rational drug design. In the case an experimental 3D structure of the target is not available the structure can be modeled utilizing the experimental structure of a closely related target. Molecular docking is the preferred method to investigate binding poses and protein ligand interactions.
Scoring functions give a rough evaluation of the correctness of the binding mode. However, other approaches like interaction fingerprint (IFP) that determines the protein-ligand interaction similarity to an experimentally supported binding mode have also been shown to be successful in pose selection.\textsuperscript{37,38} To evaluate the validity of a binding pose, knowledge of ligand structure-activity relationship (SAR) and protein SAR is advantageous.\textsuperscript{40-42} Site directed mutagenesis studies are useful to reveal important interaction sites in the binding pocket. Consideration of the dynamic behaviour of essential protein-ligand interactions using MD-simulations can be used as a further validation of protein-ligand binding poses and to explain observed SAR (see \textbf{Chapter 6}).\textsuperscript{40,42} In case of flexible ligands, energy calculations of different ligand conformations might be useful to get insight in the actual binding conformation (see \textbf{Chapter 7}).

\textbf{Pharmaceutical targets}

\textbf{Criteria for target selection}

Choosing a target to elicit a therapeutic effect is among the thorniest problems in drug discovery. Molecular and cell biology is providing new insight into molecular pathways. Genetic approaches like gene expression profiling and gene knockout screening can link disease phenotypes to special genes and proteins. A potential target needs to combine ‘drugability’ with disease modifying properties. Other criteria like expression, purification, assay development, crystallization properties and structural information can be decisive factors for target selection.

\textbf{G-protein coupled receptors (GPCRs) – most targeted drug target class}

From a comprehensive analysis in 2006 of existing drugs and their targets a consensus number of 324 drug targets for all classes of approved therapeutic drugs was proposed.\textsuperscript{43} Rhodopsine-like (or class A) G-protein coupled receptors (GPCRs) are the target where most of the drugs act on (26.8\% of approved drugs).

GPCRs are integral membrane proteins consisting of seven $\alpha$-helices that resemble a barrel within the membrane (\textbf{Figure 3}) The transmembrane (TM) $\alpha$-helices are connected by three extracellular loops (ELs) and three intracellular loops. Ligands enter the binding pocket from the extracellular side. Upon agonist binding, GPCRs get activated by conformational rearrangements of the TM $\alpha$-helices. On the intracellular side G-proteins and also $\beta$-arrestins\textsuperscript{44} are the two primary signal transducers that trigger a signal cascade in the cytoplasm.\textsuperscript{45}
The success of drug discovery in the field of GPCR in the past 4 decades was significant although GPCRs, like other membrane proteins, are difficult to crystallize. In 2000, the first crystal structure of a mammalian GPCR, that of bovine rhodopsin was solved.\textsuperscript{47} Success stories of recently solved GPCR structures (see Table 1) are expected to enhance research and drug discovery in the GPCR field.\textsuperscript{48}

**Table 1. x-ray structures of GPCRs**

<table>
<thead>
<tr>
<th>receptor</th>
<th>species</th>
<th>year of first structure</th>
<th>REF</th>
</tr>
</thead>
<tbody>
<tr>
<td>rhodopsin</td>
<td>bovine</td>
<td>2000</td>
<td>47</td>
</tr>
<tr>
<td>β2-adrenergic</td>
<td>human</td>
<td>2007</td>
<td>49</td>
</tr>
<tr>
<td>β1-adrenergic</td>
<td>turkey</td>
<td>2008</td>
<td>50</td>
</tr>
<tr>
<td>A₂A adenonsine</td>
<td>human</td>
<td>2008</td>
<td>51</td>
</tr>
<tr>
<td>chemokine CXCR4</td>
<td>human</td>
<td>2010</td>
<td>52</td>
</tr>
<tr>
<td>dopamine D₃</td>
<td>human</td>
<td>2010</td>
<td>53</td>
</tr>
<tr>
<td>histamine H₁</td>
<td>human</td>
<td>2011</td>
<td>46</td>
</tr>
<tr>
<td>muscarinic M₂</td>
<td>human</td>
<td>2012</td>
<td>54</td>
</tr>
<tr>
<td>sphingosine 1-phosphate 1</td>
<td>human</td>
<td>2012</td>
<td>55</td>
</tr>
<tr>
<td>muscarinic M3</td>
<td>rat</td>
<td>2012</td>
<td>56</td>
</tr>
<tr>
<td>κ-opioid</td>
<td>human</td>
<td>2012</td>
<td>57</td>
</tr>
<tr>
<td>μ-opioid</td>
<td>mouse</td>
<td>2012</td>
<td>58</td>
</tr>
<tr>
<td>nociceptin/orphanin FQ</td>
<td>human</td>
<td>2012</td>
<td>59</td>
</tr>
<tr>
<td>δ-opioid</td>
<td>mouse</td>
<td>2012</td>
<td>60</td>
</tr>
<tr>
<td>Neurotensin 1</td>
<td>rat</td>
<td>2012</td>
<td>61</td>
</tr>
<tr>
<td>PAR1</td>
<td>human</td>
<td>2010</td>
<td>62</td>
</tr>
<tr>
<td>5HT₁B</td>
<td>human</td>
<td>2013</td>
<td>63</td>
</tr>
</tbody>
</table>
Table 1. continued

<table>
<thead>
<tr>
<th>receptor</th>
<th>Species</th>
<th>year of first structure</th>
<th>REF</th>
</tr>
</thead>
<tbody>
<tr>
<td>5HT_{2B}</td>
<td>Human</td>
<td>2013</td>
<td>64</td>
</tr>
<tr>
<td>smoothened 7TM</td>
<td>Human</td>
<td>2013</td>
<td>65</td>
</tr>
<tr>
<td>CXCR2</td>
<td>N/A</td>
<td>2013</td>
<td>66</td>
</tr>
<tr>
<td>Alpha9 Nicotinic Acetylcholine</td>
<td>Human</td>
<td>2013</td>
<td>67</td>
</tr>
</tbody>
</table>

GPCRs recognize a huge variety of ligands, ranging from small particles, photons and ions, through small ligands, amine and nucleoside, to peptides and lipids and large proteins. Especially aminergic GPCRs are preferred drug targets because of the drug-likeness of their natural ligands.

**Histamine H₄ receptor (H₄R) – a promising new drug target**

The histamine H₄ receptor (H₄R) is the latest identified receptor within the histaminergic receptor family. H₄R is mainly expressed in haematopoietic cells, in particular eosinophils, T-cells, dendritic cells, basophils and mast cells. Several studies show that H₄R antagonists are able to modulate immune responses and inflammatory processes. Inflammation, allergy, colitis, asthma, pain, pruritus, cancer, arthritis and haematopoiesis are proposed therapeutic fields for H₄R targeting ligands.

The number of publications about H₄R steadily increased since its initial identification in 2000 (green bars in Figure 4). Due to the high sequence homology to H₃R the first compounds that were identified to bear considerable affinities for H₄R were imidazole containing H₃R ligands. This fact and the higher propensity of imidazoles to also inhibit CYPs are the reason for other compound classes to be more attractive as potential drugs. In 2003 the indolecarboxamide compound JNJ 7777120 was published as the first selective and potent non-imidazole human H₄R antagonist. This compound is used as a reference ligand for most studies on H₄R. Meanwhile a lot of compounds from different structural classes that show affinity for H₄R are reported (blue bars in Figure 4). Prominent examples are presented in Figure 5. More than 10 years histamine H₄R research has meanwhile led to significant interest of the pharmaceutical industry and resulted in a number of patent applications (red bars in Figure 4). Moreover very recently the compound UR 63325 (undisclosed structure) has entered clinical phase II for respiratory diseases.
Figure 4. Number of H₄R relevant publications (status: September 11th, 2012). Number of publications were extracted from pubmed (http://www.pubmed.com) by entering ‘histamine H₄’ as search term (green bars). Publications were new affinity data for H₄R were reported are additionally shown separately (blue bars). The number of patent applications (red bars) was extracted from esp@cenet (http://www.espacenet.com) by entering the search term ‘histamine H₄’.

Figure 5. Structures and affinities of prominent H₄R ligands sorted by the date of their publications (left to right). 83,86,90-104
The increasing number of GPCR X-ray structures,\textsuperscript{105,106} including the recently solved histamine H\textsubscript{1} receptor (H\textsubscript{1}R) crystal structure,\textsuperscript{46} offers new opportunities for histamine receptor homology modeling and drug discovery.\textsuperscript{40,42} Key residues for ligand binding have been identified by site-directed mutagenesis studies which are useful for binding mode investigations\textsuperscript{40,41,107-109} (see Chapter 6). These insights in protein-ligand interactions are valuable contributions for structure-based discovery and design of improved and new H\textsubscript{1}R ligands.\textsuperscript{37,110-112}

References


Introduction


98. Liu, H.; Altenbach, R. J.; Carr, T. L.; Chandran, P.; Hsieh, G. C.; Lewis, L. G.; Manelli, A. M.; Milicic, I.; Marsh, K. C.; Miller, T. R.; Strakhova, M. I.; Vortherms, T. A.; Wakefield, B. D.; Wetter, J. M.; Witte, D. G.; Honore, P.; Ebenshade, T. A.; Brioni, J. D.; Cowart, M. D. cis-4-(Piperazin-1-yl)-5,6,7a,8,9,10,11,11a-octahydrobenzofuro[2,3-h]quinazolin-2-amine (A-


adjustment and ligand screening with a pseudoreceptor of the human histamine H4 receptor. 