CHAPTER 4

Milestones of the drug discovery
4. **Milestones of the drug discovery:**

Highlights the importance of the below critical milestones of the drug discovery and correlated to the current research.

1. Target identification
2. Lead discovery
3. Lead optimization
4. Ligand-based design
5. Receptor-based design (Docking)

Historically, drugs were discovered through identifying the active ingredient from traditional remedies or by serendipitous discovery. Later chemical libraries of synthetic small molecules, natural products or extracts were screened in intact cells or whole organisms to identify substances that have a desirable therapeutic effect in a process known as classical pharmacology. Since sequencing of the human genome which allowed rapid cloning and synthesis of large quantities of purified proteins, it has become common practice to use high throughput screening of large compounds libraries against isolated biological targets which are hypothesized to be disease modifying in a process known as reverse pharmacology. Hits from these screens are then tested in cells and then in animals for efficacy. Even more recently, scientists have been able to understand the shape of biological molecules at the atomic level, and to use that knowledge to design drug candidates.
Modern drug discovery involves the identification of screening hits, medicinal chemistry and optimization of those hits to increase the affinity, selectivity (to reduce the potential of side effects), efficacy/potency, metabolic stability (to increase the half-life), and oral bioavailability. Once a compound that fulfills all of these requirements has been identified, it will begin the process of drug development prior to clinical trials.

Drug development:

In contemporary drug discovery, the following are the major milestones in drug development:

1) Target identification
2) Target validation
3) Lead identification
4) Lead optimization
5) Preclinical pharmacology & toxicology and
6) Clinical trials.
Failure of a candidate drug molecule at the development stage (stages 3, 4) can occur as a result of a combination of reasons, such as poor pharmacokinetics, lack of efficacy, and/or toxicity. Improving the pharmacological profile of a candidate molecule requires the optimization of numerous, often competing objectives (ie, biological or chemical properties), to discover the few improved molecules that represent the best compromise of the multiple criteria important for a successful drug. Typically, when a series of compounds with adequate potency are identified and the remaining objectives have to be taken into account, the pharmaceutical industry strives to optimize one objective at a time – starting with the binding affinity of a molecule – as part of a process involving the sequential optimization of each biological property. The process involves a series of biological tests (screening rounds) performed to determine the biological attributes of molecules e.g. binding affinity with the pharmaceutical target of interest, toxicity tests, etc. After each round of screening it is the goal of teams of highly trained medicinal and computational chemists to decide which compounds to synthesize and/or select for further investigation so as to gradually detect a set of ligands binding to the target with sufficient affinity. The decision is taken based on the results of the previous screening rounds, the expert knowledge of the team of chemists and knowledge about the pharmaceutical target.

4.1 Target Identification:

Target-based drug discovery begins with the identification of the function of a potential therapeutic drug target and understanding its role in the disease process. Most of the drugs fail in clinical trails due to couple of major reasons, the
first is that they do not work and the second is that they are not safe. As such, one of the most important steps in developing a new drug is target identification and validation. A target is a broad term which can be applied to a range of biological entities which may include for example proteins, genes and RNA. A good target needs to be efficacious, safe, meet clinical and commercial needs and, above all, be ‘druggable’. A ‘druggable’ target is accessible to the putative drug molecule, be that a small molecule or larger biologicals and upon binding, elicit a biological response which may be measured both in vitro and in vivo. It is now known that certain target classes are more amenable to small molecule drug discovery, for example, G-protein-coupled receptors (GPCRs), whereas antibodies are good at blocking protein/protein interactions. Good target identification and validation enables increased confidence in the relationship between target and disease and allows us to explore whether target modulation will lead to mechanism-based side effects. At this point, little may be known about proteins identified as potential drugable targets. If the target is an enzyme, suspected catalytic activity needs to be determined as well as downstream substrates. Specific assay methods will not exist, although generic catalytic activity can be ascertained by proteome-scale screens (protease, phosphatase, kinase, etc.). Identifying small molecules that bind to target proteins may help elucidate information on their role in the biological process.

Data mining of available biomedical data has led to a significant increase in target identification. In this context, data mining refers to the use of a bioinformatics approach to not only help in identifying but also selecting and prioritizing potential disease targets. The data which are available come from a variety of sources but include publications and patent information, gene expression
data, proteomics data, transgenic phenotyping and compound profiling data. Identification approaches also include examining mRNA/protein levels to determine whether they are expressed in disease and if they are correlated with disease exacerbation or progression. Another powerful approach is to look for genetic associations, for example, is there a link between a genetic polymorphism and the risk of disease or disease progression or is the polymorphism functional. For example, familial Alzheimer's Disease (AD) patients commonly have mutations in the amyloid precursor protein or presenilin genes which lead to the production and deposition in the brain of increased amounts of the Abeta peptide, characteristic of AD.

![Fig 4.2 Target Identification](image-url)
4.2 Target Validation:

After a drug target has been identified, a rigorous evaluation needs to occur to demonstrate that modulation of the target will have the desired therapeutic effect. This involves intensive *in vitro*, as well as *in vivo* studies that provide information on the effects of the pharmacological intervention. The result of these efforts is to establish sufficient knowledge so that physiologically relevant model systems can be developed into assays for downstream screening. This can have a significant impact on reducing the attrition rate of drug candidates due to clinical failures. Validation techniques range from *in vitro* tools through the use of whole animal models, to modulation of a desired target in disease patients. While each approach is valid in its own, confidence in the observed outcome is significantly increased by a multi-validation approach as depicted below.
4.3 Lead identification:

The aim of this stage of the work is to refine each hit series to try to produce more potent and selective compounds which possess PK properties adequate to examine their efficacy in any in vivo models that are available. Typically, the work now consists of intensive SAR investigations around each core compound structure, with measurements being made to establish the magnitude of activity and selectivity of each compound. This needs to be carried out systematically and, where structural information about the target is known, structure-based drug design techniques using molecular modelling and methodologies such as X-ray crystallography and NMR can be applied to develop the SAR faster and in a more focused way. This type of activity will also often give rise to the discovery of new binding sites on the target proteins.
A screening cascade at this time would generally consist of a relatively high throughput assay establishing the activity of each molecule on the molecular target, together with assays in the same format for sites where selectivity might be known, or expected to be, an issue. A compound meeting basic criteria at this stage would be escalated into a further bank of assays. These should include higher order functional investigations against the molecular target and also whether the compounds were active in primary assays in different species. The HTS assay is generally carried out on protein encoded by human DNA sequences but as animal models are used to validate the activity of compounds in *in vivo* disease models, in pharmacodynamic (PD)/PK modelling and in preclinical toxicity studies, it is important to have data on activity *in vitro* on orthologues. This is also particularly important as it will assist in minimizing dosing levels in toxicology studies which are chosen on the basis multiples of the pharmacologically effective doses.

Attention in this phase has to also turn to more detailed profiling of physicochemical and *in vitro* ADME properties and this series of studies is carried out in parallel, with key compounds being selected for assessment.

Solubility and permeability assessments are crucial in ruling in or out the potential of a compound to be a drug, that is, drug substance often needs access to a patient's circulation and therefore may be injected or more generally has to be adsorbed in the digestive system. Deficiency in one or other parameter in a molecule can sometimes be put right. For example formulation strategies can be used to design a tablet such that it dissolves in a particular region of the gut at a pH in which the compound is more soluble. A compound that lacks both these properties is very unlikely to become a drug no matter how potent it is in the primary screening assay. Microsomal stability is a useful measure of the ability of *in vivo*
metabolizing enzymes to modify and then remove a compound. Hepatocytes are sometimes used in this sort of study instead and these will give more extensive results but are not used routinely as they need to be prepared freshly on a regular basis. CYP450 inhibition is examined as, among other things, it is an important predictor of whether a new compound might have an influence on the metabolism of an existing drug with which it may be co-administered.

If one or more of these properties is less than ideal, then it might be necessary to screen many more compounds specifically for those properties. Each program will end up subtly different in this regard. For example in one recent project to identify novel GPCR antagonists, a number of sub-micromolar hit compounds were identified. The main issues associated with these molecules was that they showed some speciation with poorer receptor affinities in rodent receptors, a general lack of selectivity with >50% inhibition at 10 μM at 30 out of 63 GPCRs and transporters tested in a cross-screening panel as well as broad CYP450 inhibitory activity. It was felt that a number of these deficiencies were associated with the nature of the base common to all the initial structures. Modification of the basic residue resulted in a number of compounds which were as potent as the initial hits at the principal receptor but which were more selective in their actions. In common with many programs, as potency at the principal target improved selectivity issues in this series were left behind.

Key compounds which are beginning to meet the target potency and selectivity, as well as most of the physicochemical and ADME targets, should be assessed for PK in rats. Here one would normally be aiming for a half-life of >60 min when the compound is administered intravenously and a fraction in excess of
20% absorbed following oral dosing although sometimes, different targets require very different PK profiles. In large pharma with inhouse drug metabolism pharmacokinetics (DMPK) departments numerous compounds might be profiled while in academic environments there may be funds for only a predefined number of these expensive investigations. As the receptor antagonist program, described above, advanced through the hit-to-lead phase, a number of compounds were prepared which had potency in the nanomolar range and a benign selectivity profile except for some potency at the hERG channel, a potassium voltage-gated ion channel important for cardiac function and inhibition at which can cause cardiac liability. Ideally for hERG we were aiming for an activity over 30 uM or at least a 1000-fold selectivity for the target. A number of these compounds were examined in PK studies and were found to have a reasonable half-life following intravenous dosing but poor plasma levels were noted when the compound was given orally to rats. It was felt that some of these compounds, representing the end of the hit-to-lead phase of the project were, although not likely themselves to be progressed, capable of answering questions in disease models.

Thus, compounds were administered intra-peritoneally and results from the experiments gave substantial credence to the developing program. Attention in this phase has to also turn to more detailed profiling of physicochemical and in vitro ADME properties and this series of studies is carried out in parallel, with key compounds being selected for assessment. Solubility and permeability assessments are crucial in ruling in or out the potential of a compound to be a drug, that is, drug substance often needs access to a patient's circulation and therefore may be injected or more generally has to be adsorbed in the digestive system. Deficiency in one or other parameter in a molecule can sometimes be put
right. For example formulation strategies can be used to design a tablet such that it dissolves in a particular region of the gut at a pH in which the compound is more soluble. A compound that lacks both these properties is very unlikely to become a drug no matter how potent it is in the primary screening assay. Microsomal stability is a useful measure of the ability of in vivo metabolizing enzymes to modify and then remove a compound. Hepatocytes are sometimes used in this sort of study instead and these will give more extensive results but are not used routinely as they need to be prepared freshly on a regular basis. CYP450 inhibition is examined as, among other things, it is an important predictor of whether a new compound might have an influence on the metabolism of an existing drug with which it may be co-administered.

If one or more of these properties is less than ideal, then it might be necessary to screen many more compounds specifically for those properties. Each program will end up subtly different in this regard. For example in one recent project to identify novel GPCR antagonists, a number of sub-micromolar hit compounds were identified. The main issues associated with these molecules was that they showed some speciation with poorer receptor affinities in rodent receptors, a general lack of selectivity with >50% inhibition at 10 µM at 30 out of 63 GPCRs and transporters tested in a cross-screening panel as well as broad CYP450 inhibitory activity. It was felt that a number of these deficiencies were associated with the nature of the base common to all the initial structures. Modification of the basic residue resulted in a number of compounds which were as potent as the initial hits at the principal receptor but which were more selective in their actions. In common with many programs, as potency at the principal target improved selectivity issues in this series were left behind.
Key compounds which are beginning to meet the target potency and selectivity, as well as most of the physicochemical and ADME targets, should be assessed for PK in rats. Here one would normally be aiming for a half-life of >60 min when the compound is administered intravenously and a fraction in excess of 20% absorbed following oral dosing although sometimes, different targets require very different PK profiles. In large pharma with inhouse drug metabolism pharmacokinetics (DMPK) departments numerous compounds might be profiled while in academic environments there may be funds for only a predefined number of these expensive investigations. As the receptor antagonist program, described above, advanced through the hit-to-lead phase, a number of compounds were prepared which had potency in the nanomolar range and a benign selectivity profile except for some potency at the hERG channel, a potassium voltage-gated ion channel important for cardiac function and inhibition at which can cause cardiac liability. Ideally for hERG we were aiming for an activity over 30 uM or at least a 1000-fold selectivity for the target. A number of these compounds were examined in PK studies and were found to have a reasonable half-life following intravenous dosing but poor plasma levels were noted when the compound was given orally to rats. It was felt that some of these compounds, representing the end of the hit-to-lead phase of the project were, although not likely themselves to be progressed, capable of answering questions in disease models. Thus, compounds were administered intra-peritoneally and results from the experiments gave substantial credence to the developing program.

It is common for approximately 1% of the compounds from high throughput screening (HTS) to demonstrate some level of antagonistic or agonistic
effects during the assay campaign, hence being categorized as hits. The next step is to validate these by repeating the primary screening assay. The hits that repeat will go into a secondary screening hit selection process to help choose the best candidates to go into lead optimization.

Secondary screening will yield more detailed dosed response information, allowing the comparison of the affinities of the compounds for the target. This is usually represented as IC50 data, a good hit having <10 µM potency. Among the challenges at this point, is to cull the false positives and assign a level of quantitative intelligence to help guide the decision making process. This is accomplished through the use of counter screens and selectivity screens to help reduce the risk of off-target effects. Mechanism of action studies may also be required to ensure that an inhibitor is intervening with the correct pharmacology.

4.4 Lead optimization:

The object of this final drug discovery phase is to maintain favorable properties in lead compounds while improving on deficiencies in the lead structure. Continuing with example above, the aim of the programme was now to modify the structure to minimize hERG liability and to improve the absorption of the compound. Thus, more regular checks of hERG affinity and CACO2 permeation were undertaken and compounds were soon available which maintained their potency and selectivity at the principal target but which had a much reduced hERG affinity and a better apparent permeation than initial lead compounds. When examined for PK properties in rat one of these compounds, with 8 nM affinity at
the receptor of interest, had an oral bioavailability of over 40% in rats and about 80% in dogs.

Compounds at this stage may be deemed to have met the initial goals of the lead optimization phase and are ready for final characterization before being declared as preclinical candidates. Discovery work does not cease at this stage. The team has to continue to explore synthetically in order to produce potential back up molecules, in case the compound undergoing further preclinical or clinical characterization fails and, more strategically, to look for follow-up series.

The stage at which the various elements that constitute further characterization are carried out will vary from company to company and parts of this process may be incorporated into the lead optimization phase. However, in general molecules need to be examined in models of genotoxicity such as the Ames test and in *in vivo* models of general behaviour such as the Irwin's test. High-dose pharmacology, PK/PD studies, dose linearity and repeat dosing PK looking for drug-induced metabolism and metabolic profiling all need to be carried out by the end of this stage. Consideration also needs to be given to chemical stability issues and salt selection for the putative drug substance.

All the information gathered about the molecule at this stage will allow for the preparation of a target candidate profile which with together with toxicological and chemical manufacture and control considerations will form the basis of a regulatory submission to allow human administration to begin.
The process of hit generation to preclinical candidate selection often takes a long time and cannot in any way be considered a routine activity. There are rarely any short cuts and significant, intellectual input is required from scientists from a variety of disciplines and backgrounds. The quality of the hit-to-lead starting point and the expertise of the available team are the key determinants of a successful outcome of this phase of work. Typically, within industry for each project 200 000 to \(>10^6\) compounds might be screened initially and during the following hit-to-lead and lead optimization programmes 100's of compounds are screened to hone down to one or two candidate molecules, usually from different chemical series. In academia screens are more likely to be of a focused nature due to the high cost of an extensive HTS or compounds are derived from a structure-based approach. Only 10% of small molecule projects within industry might make the transition to candidate, failing at multiple stages. These can include the (i) inability to configure a reliable assay; (ii) no developable hits obtained from the HTS; (iii) compounds do not behave as desired in secondary or native tissue assays; (iv) compounds are toxic \textit{in vitro} or \textit{in vivo}; (v) compounds have undesirable side effects which cannot be easily screened out or separated from the mode of action of the target; (vi) inability to obtain a good PK or PD profile in line with the dosing regime required in man, for example, if require a once a day tablet then need the compound to have a half-life \textit{in vivo} suitable to achieve this; and (vii) inability to cross the blood brain barrier for compounds whose target lies within the central nervous system. The attrition rate for protein therapeutics, once the target has been identified, is much lower due to less off target selectivity and prior experience of PK of some proteins, for example, antibodies.
Although relatively less costly than many processes carried out later on in the drug development and clinical phases, preclinical activity is sufficiently high risk and remote from financial return to often make funding it a problem. Ensuring transparency of the cost of each stage/assay within large pharma may help reduce some of their costs and there are some movements towards this as companies instigate a ‘biotech’ mentality and accountability for costs.

Once a candidate is selected, the attrition rate of compounds entering the clinical phase is also high, again only one in 10 candidates reaching the market but at this stage the financial consequences of failure are much higher. There has been considerable debate in industry as to how to improve the success rate, by ‘failing fast and cheap’. Once a candidate reaches the clinical stage, it can become increasingly difficult to kill the project, as at this stage the project has become public knowledge and thus termination can influence confidence in the company and shareholder value. Carrying out more studies prior to clinical development such as improved toxicology screens (using failed drugs to inform these assays), establishing predictive translational models based on a thorough disease understanding and identifying biomarkers may help in this endeavor. It is particularly in these later two areas where academic-industry partnerships could really add value pre-clinically and eventually help bring more effective drugs to patients.

Optimizing leads from hit to lead discovery usually starts off with selecting the leads from secondary screening with proven specificity and the highest binding affinities to the target of interest. Often binding affinities start off in the 1 – 10 µM range, requiring potency improvements of up to five orders on
magnitudes before they would be considered viable drug candidates. Along with engineering in potency and selectivity, other factors such as toxicity and bioavailability need to be considered.

The lead optimization stage will often be accompanied with structural information from ligand-target X-ray crystallization data. Data generated during secondary screening will include $IC_{50}/K_i$ information as well as chemical structures. This information can help in the elucidation of bioavailability and pharmacokinetics and infer structural modification. However, very little information may be available that will give insights into how and why the hits bind to the target. Selecting the best compounds to move into lead optimization will dramatically improve the chances of candidates with desirable potency and selectivity.

Preclinical pharmacology & toxicology and Clinical trials will follow next in drug development.