Early Stages of Drug Discovery in the Pharmaceutical Industry

Daniel Seeliger / Jan Kriegl, Discovery Research, Boehringer Ingelheim
September 29, 2016
Historical Drug Discovery
From Accidental Discovery to a Drug

Cattle, 1920s, North Dakota

Dead Cattle, 1920s, North Dakota

Coumarin

Fungi

4-Hydroxy-Coumarin

CH₂O

Di-Coumarol

Strong anti-coagulant
Historical Drug Discovery
From Accidental Discovery to a Drug

Accident (Decaying Hay)

Observation of a pharmacological effect

Link of the pharmacological effect to a substance

Chemical variation of the substance

Warfarin

Treatment of diseases
Modern Drug Discovery
From the Disease to the Drug

Disease → Connection between the disease and molecular mechanisms → Definition of “drug targets” → Search for molecules which act on the drug target → Approval for Treatment → Clinical evaluation of the new drug
Drug Discovery and Development

- Discovery Research
- Clinical Development
- Approval

10-12 years
Drug Discovery Project Phases

Discovery Research

Target Discovery → Assay Development → Lead Identification → Lead Optimization → Development
What are Drug Targets?
The Life Cycle of the HI Virus

Blocking a step in the virus life cycle stops viral reproduction
Cleavage of the HIV poly-protein is essential for viral reproduction.

Inhibition of the catalytic mechanism will prevent viral reproduction.

Lower viral reproduction rate will slow down or even halt the progression to AIDS.

Many drug discovery programs fail because the hypothesis about the target/disease link turns out to be wrong!
Inhibition of Drug Targets
HIV-1 Protease Inhibitors

Inhibitors:
- Bind with high affinity to the active site of the enzyme
- Prevent entry and cleavage of the normal substrate of the enzyme
Inhibition of Drug Targets
Molecular Interactions
Principle of Drug Design:
Design molecules which are able to form multiple favorable interactions with the target and which are complementary in shape.
Drug Discovery Project Phases

- Target Discovery
- Assay Development
- Lead Identification
- Lead Optimization

Development
Assays in Drug Discovery

Assay Development

- Develop a robust and reliable assay!
  - Reproducible results
  - Sufficient throughput
  - Physiologically relevant

No drug discovery program without reliable assay
Drug Discovery Project Phases

Discovery Research

Target Discovery
Assay Development
Lead Identification
Lead Optimization

Development
Lead Identification
High-Throughput Screening

Big pharma companies have large compound collections
- Historically grown over decades of research
- Permanently updated with compounds from vendor catalogs
- Permanently updated through inhouse syntheses (combinatorial chemistry)

Boehringer Ingelheim has ~ 2,000,000 compounds registered in the central database

Boehringer Ingelheim’s large screening pool currently contains ~ 850,000 compounds.
Lead Identification
High-Throughput Screening

Throughput ~ 50 000 cpds/day
HTS usually yields several thousand hits

HTS data analysis:
- Clustering of molecules based on chemical similarity
- Identify common patterns among hits (structure-activity relationships, SAR)
Lead Identification
From Data to Information – HTS Analysis

Find Attractive Starting Points for Optimization -> Lead Classes
Drug Discovery Project Phases

Discovery Research

- Target Discovery
- Assay Development
- Lead Identification
- Lead Optimization

Development
Lead Optimization
Multiple Challenges for a Molecule

-MOUTH INTESTINE (pH=7)-
STOMACH (pH=1)
Gut wall

TARGET

Portal vein
Liver
BLOOD
Metabolism

BRAIN
Drug Design is a Multi-Parameter Optimization

- Potency
- Selectivity
- Bioavailability
- Solubility
- Metabolic Stability
- Plasma Protein Binding
- Cytochrome Inhibition (Drug/Drug Interactions)
- Brain Permeation
- Toxicity
- Pharmacokinetic
- .........
Lead Optimization essentially means synthesis of close analogs of an active molecule.

- „Easy“ chemistry -> Variation straightforward
- Cheap!
- Difficult chemistry
- Different synthetic routes
- Expensive!
Synthetic Chemistry can be very expensive (on average 2000€/molecule)

Make Predictions of Molecule Properties!

„Predictions are difficult, especially about the future“ (Niels Bohr)

Structure-based design
- Xray structure(s) required
- Physics-based approaches

Ligand-based design
- Known ligand required
- Physics-based approaches
- Chemoinformatics

Data-driven design
- Lots of data required
- Chemoinformatics
- Machine Learning
Xray Crystallography is a key technology for potency optimization, but....

Free Energy of Binding is an ensemble property -> Cannot be computed from a single structure
Lead Optimization
Computing Binding Free Energies

\[ \Delta G = -RT \ln \left( \frac{1}{8\pi^2} \frac{C_P C_L}{C_{PL}} \frac{\int e^{-(U(r_P)+W(r_P))/RT} dr_P}{\int e^{-(U(r_L)+W(r_L))/RT} dr_L} \right) \]

- Molecular Dynamics Simulations
- Alchemical Free-Energy calculations
Activity of a molecule against a target is determined by their 3-dimensional structure.

Ligand-based design: Search for molecules which are similar to a template molecule.
Search for similar molecules in a virtual chemical space

BI Screening Library

BI virtual library

$\sim 10^{14}$

$10^6$

HTS

Virtual Screening
BI’s virtual library consist of molecules which can be easily synthesized

- Suzuki Coupling (boronic acids)
- Buchwald-Hartwig (anilines, amines)

BI in house
prim Amines (4000)
prim. Anilines (3500)
sec. Amines (6000)
boronic acids (2500)

Millions of possible combinations for each core.
Lead Optimization
Ligand-Based Design
Lead Optimization
Ligand-Based Design
Drug Design is a Multi-Parameter Optimization

- Potency
- Selectivity
- Bioavailability
- Solubility
- Metabolic Stability
- Plasma Protein Binding
- Cytochrome Inhibition (Drug/Drug Interactions)
- Brain Permeation
- Toxicity
- Pharmacokinetic
- .......

Structure/Ligand-based Design (target-specific)

Data-driven Design (often target-independent)
Lead Optimization
Prediction of Molecule Properties with Machine Learning

What is Machine Learning?

Data (Big Data) → Artificial Intelligence → Prediction
Prediction of Molecule Properties
Machine Learning
Prediction of Molecule Properties

Machine Learning

- Solubility?
- Metabolic Stability?
- hERG inhibition?
- CYP inhibition?
- Plasma Protein Binding?

Project (Target)-independent properties

- Solubility: > 50,000 data points
- Metabolic Stability: > 80,000 data points
- hERG inhibition > 8,000 data points
- CYP inhibition > 40,000 data points
- Plasma Protein Binding > 4,000 data points

Large data sets assembled over years and different research projects
Drug Discovery Project Phases

Discovery Research

Target Discovery
Assay Development
Lead Identification
Lead Optimization

Development
• The paradigm of modern drug discovery is to connect diseases and symptoms to molecular mechanisms

• Drug discovery programs are target centric

• Drug design is a multi-parameter optimization

• Modern computational technologies and hardware developments allow reasonable predictions of activity and other molecular properties

• Drug discovery is a very interdisciplinary field of science