COMPAS (COMmon PATternS) is a software tool that was especially developed to harness the technology of next generation sequencing (NGS) to bring light into the black box of *in vitro* selection experiments. The functionalities of COMPAS enable (A) quality control of combinatorial starting libraries, (B) improved analysis of enriched libraries as well as (C) optimization of identified leads.

**Workflow**

COMPAS can be used to leverage and optimize existing selection platforms. After the selection procedure the PCR-amplicons of successive selection rounds can be indexed and combined for an economic NGS. COMPAS provides functionalities to assign dataset sequences to its origin-experiment. Now, analysis can be performed on data of defined experiments (e.g. selection rounds).
Check of random distribution of nucleotides

Diverse combinatorial starting libraries with random regions built up by homogeneously distributed nucleotides motifs are essential for the development of ligands with the desired binding characteristics.

COMPAS gives the probability of nucleotides over the random regions (diagram and table view), thus helps to control and optimize the synthesis of optimal libraries.

Check random distribution of short motifs

The distribution of motifs gives a measure for the diversity of the starting pool on the level of motifs. As motifs are the parts that finally mediate binding to the target molecule of interest, equally random distributed libraries are the better source for high quality ligands.

The non optimized library shows an unbalanced distribution of motifs (here 6 Nt).

Whereas the optimal pool shows an Gaussian distribution of motifs.

The better the distribution of nucleotides, the better the distribution of motifs, the better is the ligand space, the better the chance to end up with high quality ligands.
Check length of random region lengths at high resolution

Also the length of the random region can be checked. In terms of random region length the figured library is fine. About 1.8 Mio sequences are of correct length.

COMPAS gives an overview of lengths of the variable regions of interest

Identification of sequences with higher copy number/check for contaminations

A typical starting libraries is expected to consist to 100% out of unique sequences. If you work with over-represented libraries you can check the copy number of full length sequences here.

Also contaminations can be identified: Load name-databases with already known clones (e.g. plastic binders, clones binding to His-tags,...) to identify them by name.
COMPAS Modes for Identification of Ligands

COMPAS has different methods for ligand identification. The conventional modes “Sequence Counting” and “Fuzzy Search” as well as the advanced “Co-occurrence mode”

The most powerful method is the “Co-occurrence mode”. It identifies ligands in the context of sequence families (patterns). Non-identical sequences that contain two frequent motifs build up so-called proto-patterns, which are subsequently grouped into sequence families according to a user-defined degree of similarity of the entire sequence. This mode is even capable to identify a relevant sequence family if each of their members is still unique.
Identification of aptamers in the context of families

Example of three sequence families identified with the ‘co-occurrence mode’. Non identical but related sequences (with a user defined degree of homology) are grouped into the same family. The patterns contain additional information: The consensus sequence including their relative and absolute number as well as the sequences of each family member including their absolute and relative number.

Radioactive filter binding analysis of aptamers predicted by COMPAS reveals:

44 aptamers could be identified after the first SELEX round with the COMPAS approach (●) whereas only 15 aptamers could be identified after nine SELEX cycles with the conventional approach (○).

✓ Affinity and selectivity of software predicted aptamers are in the same range as conventionally identified clones.
✓ All clones identified with the conventional approach could be also identified with the COMPAS approach.
COMPAS compares different experiments on a monoclonal level. Experiments can be the subsequent selection cycles (see A) as well as completely independent selection cycles that have been performed against closely related targets (see B)

A) **In silico** tracking of monoclonal sequences over consecutive SELEX cycles

- Ligand prediction from the first selection round can be confirmed on the *in silico* level.
- Evolution behavior of ligands can be analyzed on monoclonal level.
- The increase or decrease of defined candidates can be ascribed to the stringency that was applied in the respective SELEX round. Learn how to select.

B) **In silico** identification of aptamers addressing defined binding sites

Differential selection experiments against closely related targets - against wild type protein (wt) as well as against the target protein mutated at binding site of interest (mt) - enable the identification of differentially binding monoclones by comparative *in silico* analysis.

**wet lab:** selection experiment

Eight SELEX cycles against a Membran Protein (MP1):

- MP1 [wt] → MP1 [wt] 
- Cycle 1 - 6
- MP1 [mt] → MP1 [mt] 
- Cycle 7 - 8

**wet lab:** filter binding assay

Clone „MP 1-4“

- Discriminatory binding of clone MP 1-4 could already be predicted on the *in silico* level.
- Selectivity for the wild-type protein could be confirmed in binding assay.
Having a closer look on defined families: clustering of all members belonging to a relevant aptamer family

Lead Optimization

Secondary screening of relevant sequence families enables the identification of conserved and variable regions

- COMPAS gives additional information to which extend nucleotides can be replaced against others.
- Already information on the in silico level that would have to be generated in conventional approaches in laborious doped selection experiments.

Once a relevant monoclonal sequence has been identified, the focus can be set on the other family members of this family.
Summary

The integration of NGS and COMPAS analysis into the in vitro selection workflow enables…

- **... quality control of starting libraries.**
  - Diversity
  - Distribution of NTs over sequence positions
  - Distribution of motifs
  - Distribution of random region length

- **... improved analysis of enriched libraries.**
  - Ligand prediction in very early enrichment cycles
  - Identification of rare ligands (not accessible by conventional cloning and sequencing)
  - Prediction of ligands addressing defined binding sites
  - Clustering of ligands in families

- **... lead optimization and advanced patenting strategies.**
  - Comprehensive information of conserved and variable sequence regions at high resolution

about **aptaIT GmbH**

With its complementary team of scientists in the fields of physics, molecular biology, informatics and medicine, AptaIT bundles knowledge for the development of high performance software solutions to leverage biomedical research and drug discovery.

AptaIT developed the software COMPAS, which efficiently harnesses the technology of deep sequencing for the discovery of therapeutic lead structures. In doing so, AptaIT’s software “COMPAS” optimizes and shortens the laborious and cost intensive drug development process.

**AptaIT offers:**

- expertise and support to implement COMPAS in drug discovery procedures
- service analysis of NGS datasets derived from *in vitro* selection experiments
- licenses of the software tool COMPAS (coming soon)