

Design and synthesis of diverse leadlike libraries

Jan Schultz, Alexandrine Cheronet, Herman Verheij, Valery Maltsev and Norimasa Morita

Introduction

The current trend in early drug discovery shows a shift towards screening of smaller and more focused libraries that provide good quality leads rather than just hits. In library design more and more emphasis is put on introducing druglike motifs in the compounds as well as maintaining a good profile of the physicochemical properties. In the past years many papers have been published addressing the concepts of druglikeness^{1, 2, 3} and leadlikeness^{4, 5, 6} and there is a clear trend in lead discovery to keep key physicochemical properties of leads below the limits of druglike properties, and thereby provide the 'window' for increases in these properties in lead optimization⁷.

Library design

Pyxis Discovery is developing a computational approach to describe structures of compounds as combinations of rings, linkers and substituents⁸. The software that Pyxis Discovery develops allows the systematic identification and classification of rings and linkers in structures. By definition linkers have at both terminal ends rings. The basis for the design of the first libraries have been drugs in clinical development and approved drugs that have been derived from the Comprehensive Medicinal Chemistry (CMC) and MDL Drug Data Report (MDDR) databases. Prior to the identification of the rings and linkers the databases have been cleaned in order to only assess structures of compounds that are orally available. The identified rings and linkers from these compounds have been used to virtually 'construct' new structures. The main principle is illustrated in Figure 1. From this database of virtual structures a diverse subset of compounds has been synthesized meeting the criteria for leadlikeness as defined by MW ≤ 450, clogP ≤ 4.5, cLogS ≥ -7.0, PSA ≤ 150 Å², RTB ≤ 10, HACC ≤ 10 and HDON ≤ 5.

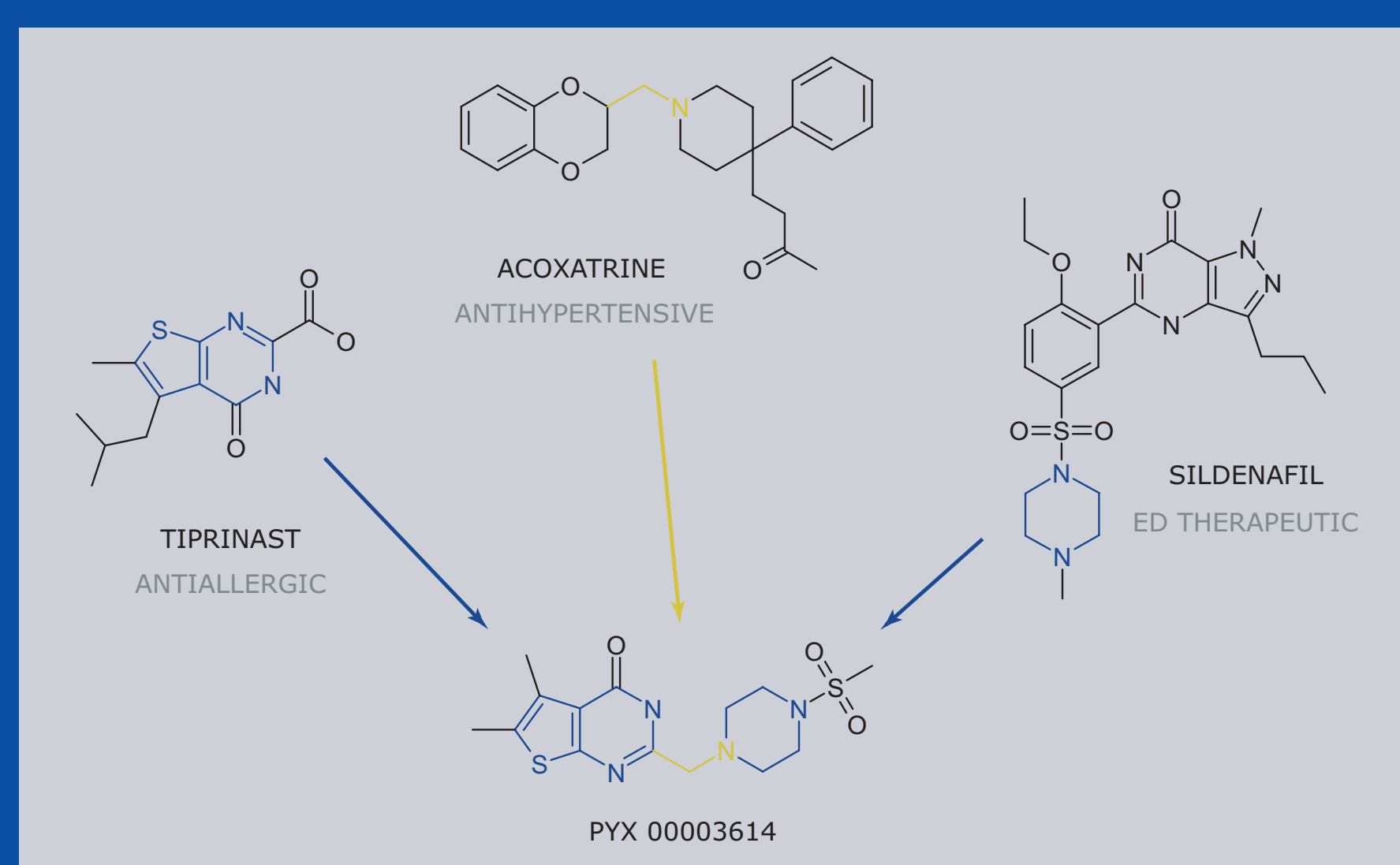


Figure 1 The main principle for constructing new structures on a basis of identified rings and linkers

Profile of the libraries

In order to assess the diversity of the libraries the Scaffold-based Classification Approach (SCA)⁹ has been used, which describes molecules as scaffolds and acyclic fragments. The reason for using this approach is that it is intuitively better understandable for the medicinal chemist than other methods that assess chemical diversity. The topological scaffold of the structure is described by two descriptors named 'complexity' and 'cyclicality'. A SCA plot can be created by plotting the cyclicality against the complexity of the structures. In Figure 2 the SCA plot of the CMC database, including the position of some example structures is presented.

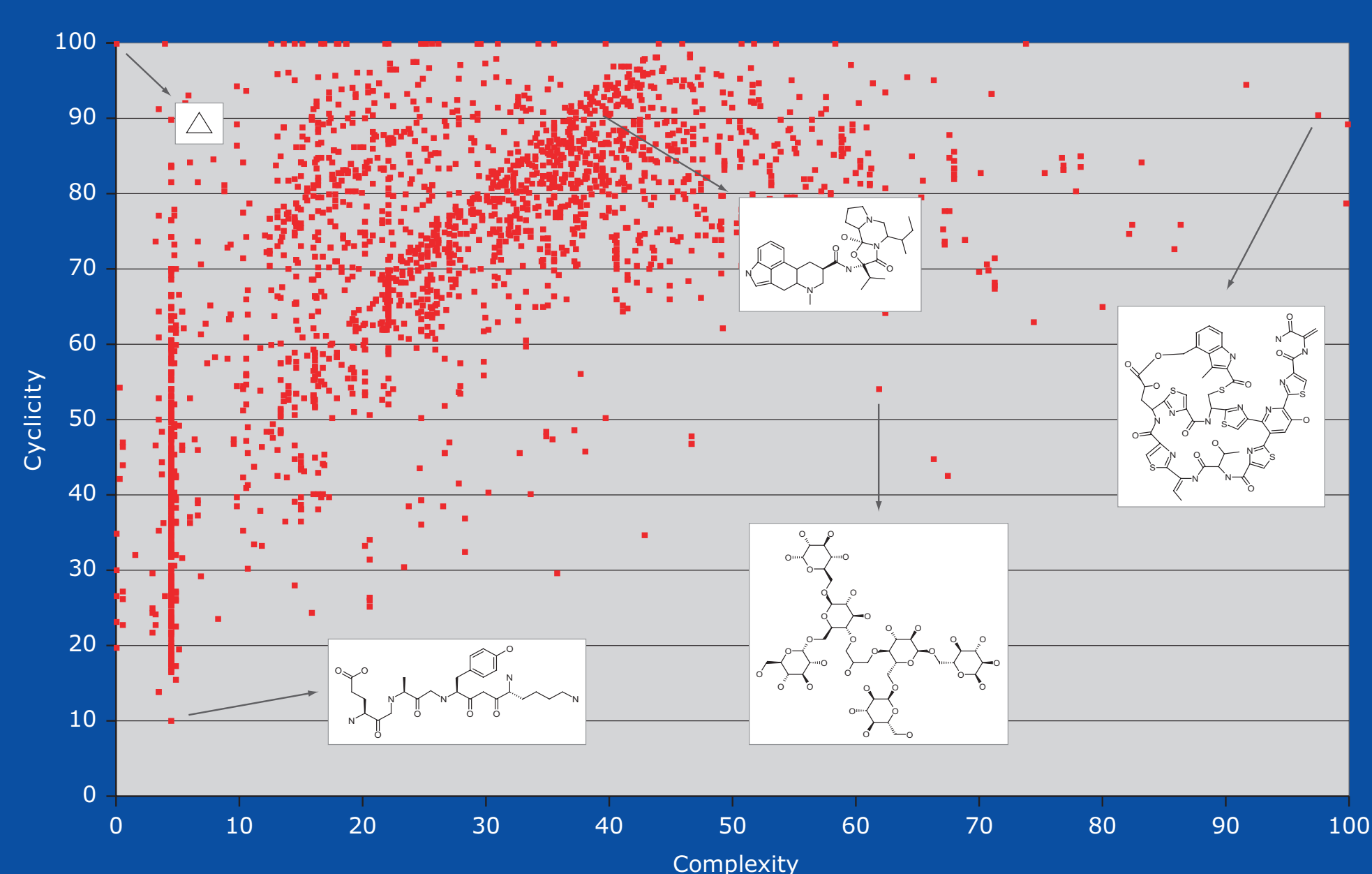


Figure 2 SCA plot of the Comprehensive Medicinal Chemistry (CMC) database, including the position of some example structures.

In Figure 3 the combined SCA plot of 3 databases is presented, the CMC database, the leadlike part of the CMC database and Smart Library Two designed and synthesized by Pyxis Discovery. It becomes clear from Figure 3 that Smart Library Two is sampling the leadlike part of the CMC database rather well.

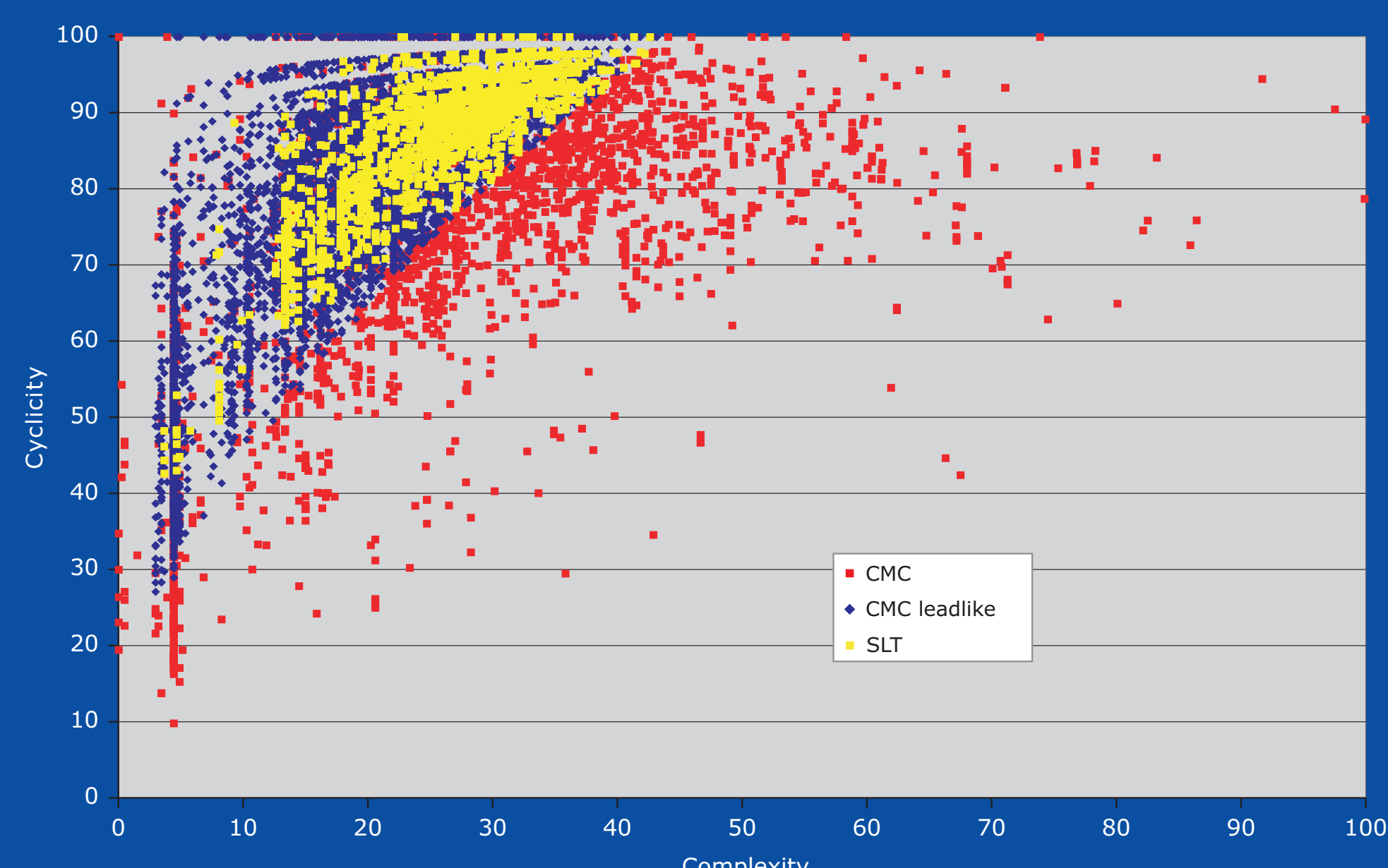


Figure 3 SCA plot of the CMC database, the leadlike part of the CMC database, and Smart Library Two (SLT)

Zooming in on the SCA plot the strength of the SCA method becomes apparent. In Figure 4 it can be seen that different scaffolds are identified as separate vertical 'lines' in the SCA plot. Although there always will be a discussion what the scaffold of a molecule is it is clear that this approach visualizes chemical diversity in a chemically intuitive way. From this analysis it can be concluded that the diversity of the designed and synthesized libraries is good.

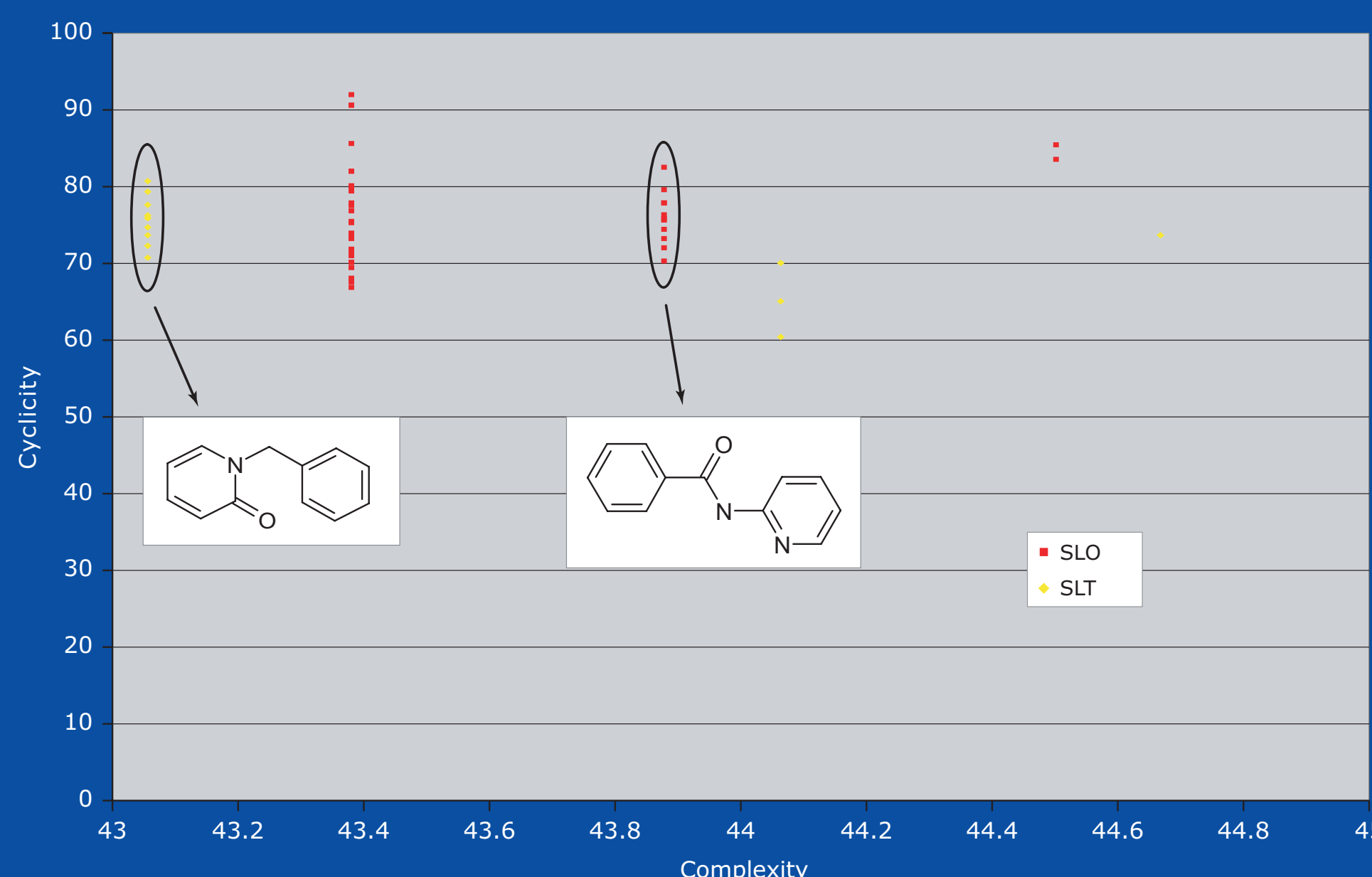


Figure 4 Topological scaffolds identified in the SCA plot in Smart Libraries One (SLO) and Two (SLT)

The calculated average values of the most important physicochemical properties of the libraries are: MW ≈ 350, clogP ≈ 2.6, cLogS ≈ -3.9, PSA ≈ 70 Å², RTB ≈ 4.7, HACC ≈ 2.5 and HDON ≈ 1.1. So the libraries have a leadlike profile that make them very suitable for lead generation projects.

Important in the selection and design of the compounds is the synthetic accessibility of the compounds. On average the number of reaction steps is 4, with a minimum of 2 and a maximum of 6. In the design of the compounds the further development of the compounds is taken into account by ensuring that each compound has on average 2 points of diversity so that the synthesis of analogs for SAR studies is feasible. In Figure 5 and 6 examples are presented how analogs can be synthesized on the basis of compound PYX 00002902 and PYX 00003664.

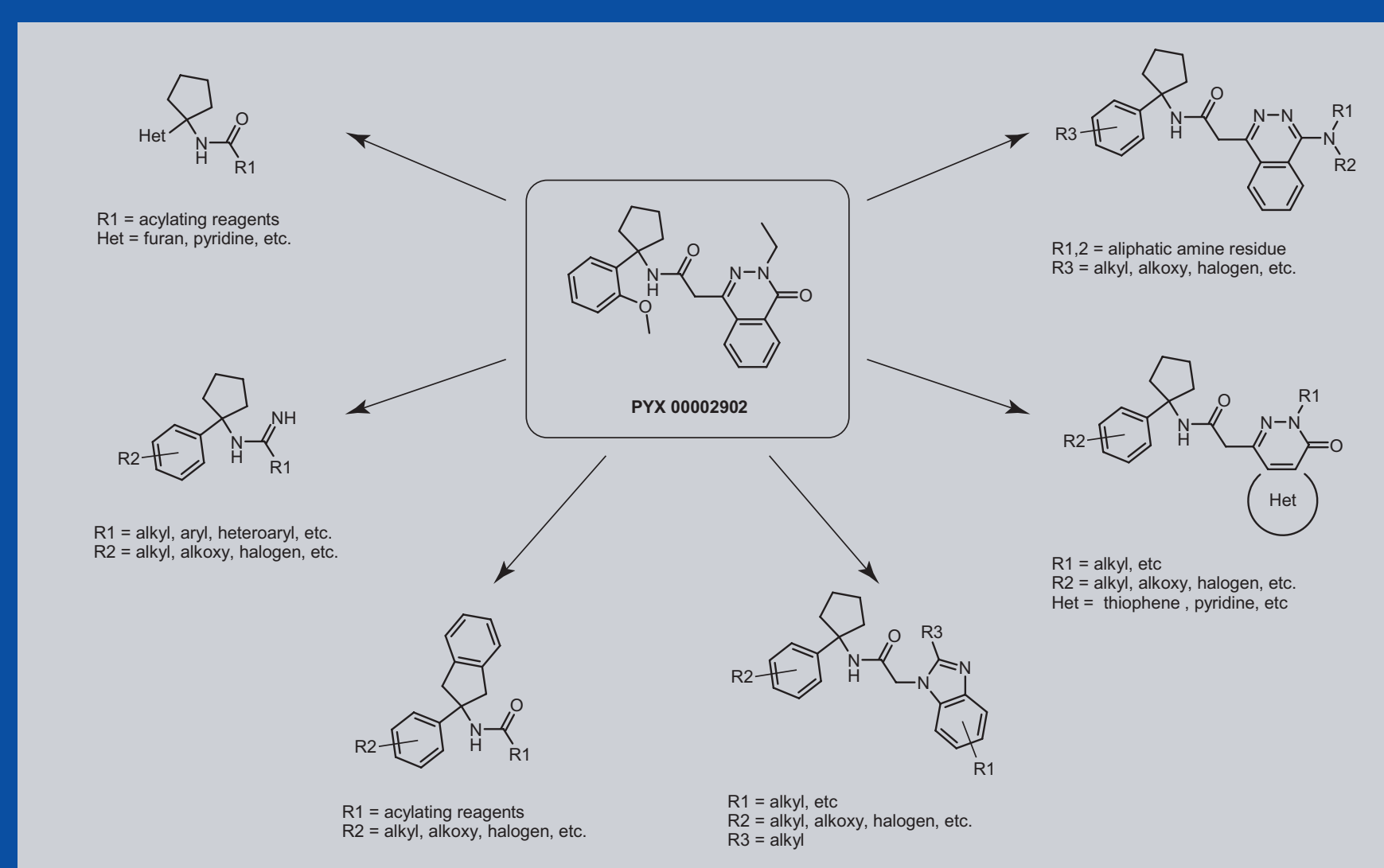


Figure 5 The various possibilities to synthesize analogs for lead optimization on the basis of compound PYX 00002902

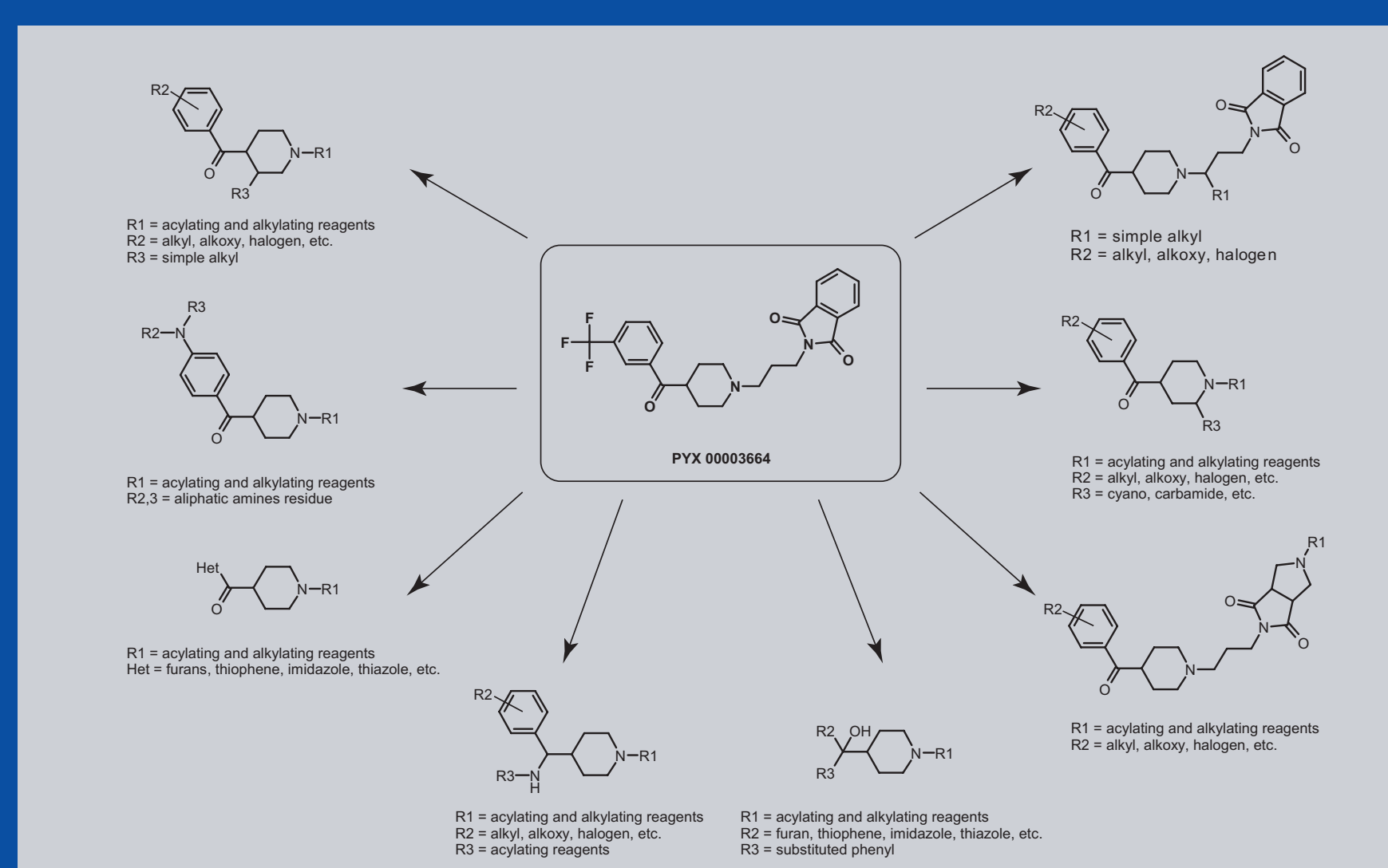


Figure 6 The various possibilities to synthesize analogs for lead optimization on the basis of compound PYX 00003664

The novelty of the compounds is assessed by comparing them with commercially available screening compounds using the Tanimoto similarity as criterion. Compounds that have a Tanimoto similarity that is smaller than 0.85 are considered to be novel. Approximately 2/3 of the compounds in the designed and synthesized libraries meet this criterion.

Therefore it can be concluded that the resulting libraries (each approximately containing 1,900 compounds) are diverse and leadlike thus making them very suitable for lead generation projects.

References

- 1) C.A. Lipinski et al. *Adv. Drug Delivery Rev.* **1997**, 23, 3.
- 2) J. Xu, *J. Chem. Inf. Comput. Sci.* **2000**, 40, 1177.
- 3) I. Muegge et al. *J. Med. Chem.* **2001**, 44, 1841.
- 4) S.J. Teague et al. *Angew. Chem. Int. Ed.* **1999**, 38, 3743.
- 5) M.H. Hann et al. *J. Chem. Inf. Comput. Sci.* **2001**, 41, 856.
- 6) T.I. Oprea et al. *J. Chem. Inf. Comput. Sci.* **2001**, 41, 1308.
- 7) P.D. Leeson et al. *J. Med. Chem.* **2004**, 47, 6338.
- 8) J. Xu *J. Med. Chem.* **2002**, 45, 5311.

European office
Delftechpark 26
2628 XH Delft
The Netherlands
p +31 15 260 09 72

USA office
225 Franklin Street
26th floor
Boston, MA 02110
p +1 617 217 27 80