What rings do medicinal chemists use, and why?

M. D. Mackey

Cresset, New Cambridge House, Bassingbourn Road, Litlington, Cambridgeshire, SG8 0SS, UK

mark@cresset-group.com www.cresset-group.com

Abstract

The vast majority of small molecule drugs contain at least one ring. Ring systems are more than just scaffolds waiting to be elaborated: the electrostatic and pharmacophoric properties of ring systems are usually crucial to the biological activity of the molecules that contain them.

In this study we examine what rings have been used by medicinal chemists and cluster them based on their substitution patterns and electrostatic properties.

Methods and results

We used ChEMBL[1] as the data source, and extracted all ring systems and their attachment point geometries. The 49 most common ring systems in ChEMBL (regardless of substitution pattern) are shown in Figure 1.

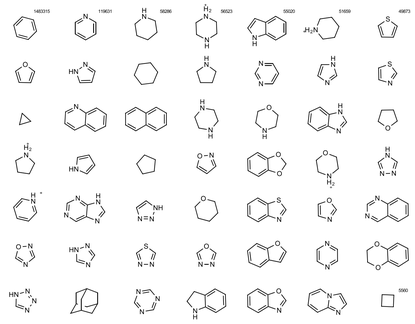

Figure 1: The top 49 most common ring systems in ChEMBL. Numbers indicate frequencies.

The rings were sorted by number of attachment points. For each of these subsets, Spark[2,3] was used to compute a similarity matrix (see Figure 2). UPGMA hierarchical clustering was then used on each subset.

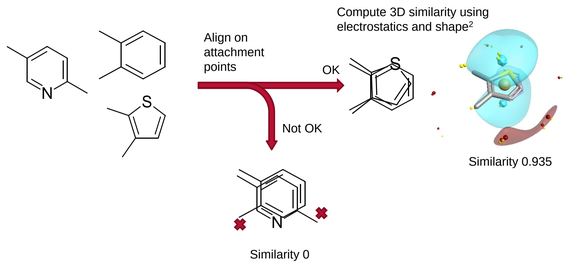

Figure 2: Computing values in the similarity matrix.

The data set consisting of rings with 2 attachment points was investigated further. The 49 most common rings in this set are shown in Figure 3, and the hierarchical clustering of the 500 most common rings in this data set is shown in Figure 4.

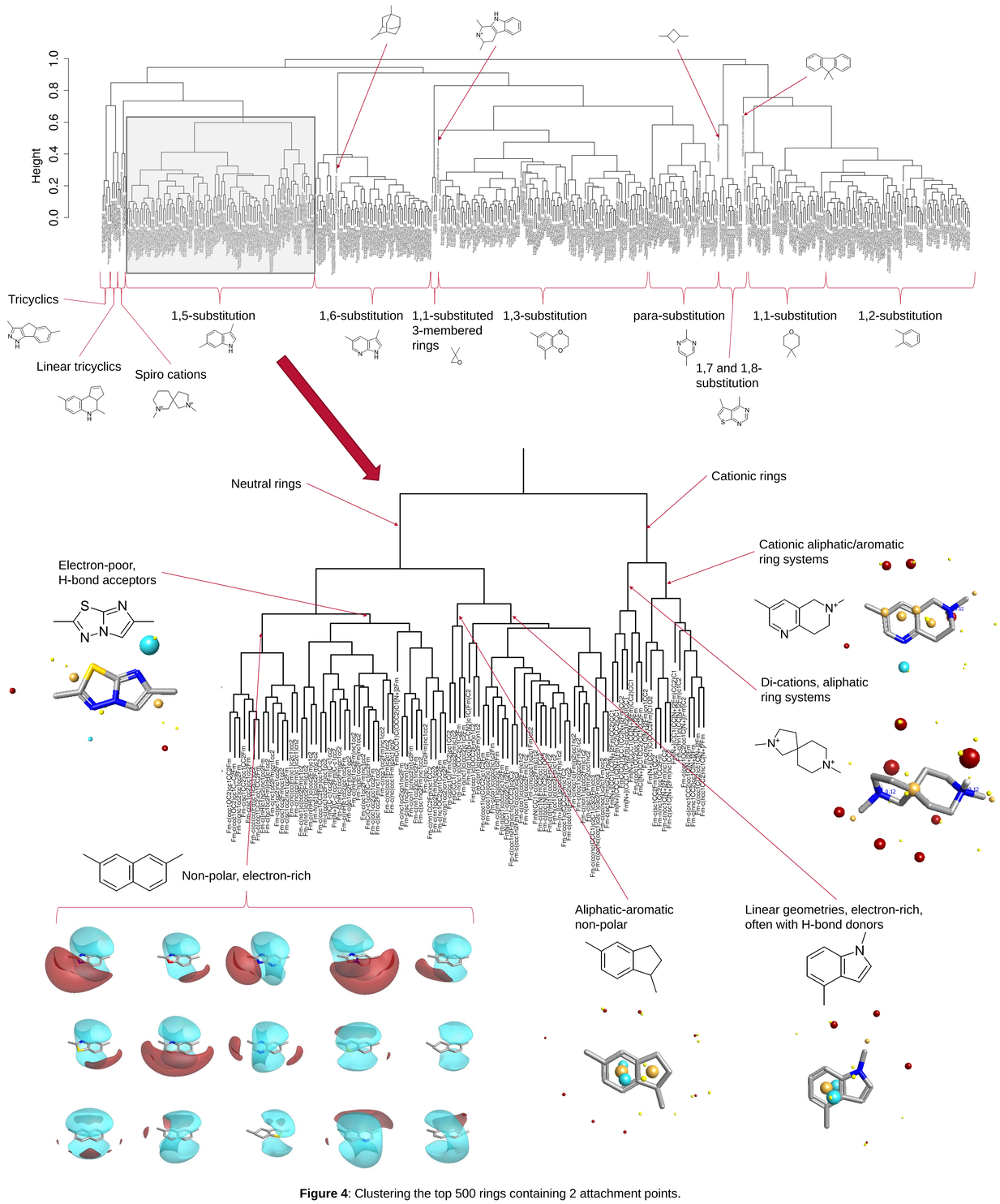

Figure 4: Clustering the top 500 rings containing 2 attachment points.

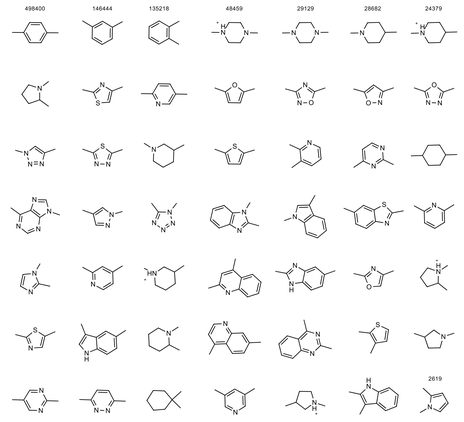

Figure 3: The top 49 most common disubstituted rings in ChEMBL. Numbers indicate frequencies.

Conclusions

The top levels of the clustering are primarily geometric. Rings with the attachment vectors pointing in different directions have a low similarity and appear in different clusters.

The lower-level clusters depend much more on the charge and electrostatic properties of the rings. Charged rings (primarily cations) cluster separately from the neutral rings. Within the neutral rings, the clusters then depend on a combination of H-bond donor and acceptor patterns within the ring, and on whether the ring is electron-rich (with a large negative potential above and below the plane of the ring) or electron-poor.

Extending this clustering to all of the ring systems in ChEMBL (and optionally adding in VEHICLe[4], a database of all possible aromatic rings) allows the chemist to see at a glance what ring systems are equivalent in terms of both attachment point geometry and electrostatic and H-bonding properties.

The current procedure for determining the ring substitution patterns includes simple substituents such as halogens. Modification of the procedure to only use larger substituents may make the results even more appropriate for scaffold-hopping.

References

1. Bento et al., Nucleic Acids Res. **2014**, 42, 1083-1090
2. Cheeseright et al. JCIM **2006**, 46, 665-676
3. Spark 10.4 http://www.cresset-group.com/spark
4. Pitt et al., J. Med. Chem. **2009**, 52, 2952-2963