

Introduction

Fragment screening is a widely used approach for identification of new chemical series in drug discovery projects. It provides an effective means of chemical space coverage and allows the development of compounds that maintain good ligand efficiency.

The success of any screening campaign depends on the quality of the library which is being screened. Fragment library collections are often generated based on physicochemical criteria, such as the Rule of 3.¹ Some consideration of chemotype diversity may also be included. A fragment library, particularly one for use with diverse protein targets, should also cover a broad range of shape and electrostatic space.

Methods are outlined to profile libraries on shape and flexibility using complementary visualisation and cube fingerprint approaches. They are i) fast and ii) easily interpretable.

Approaches

We describe a number of ways to assess and compare library shape profiles and show how they perform against a commercially available fragment library and the recently released, shape-biased BioFocus fragment collection (FRG04).

To use these approaches, we describe molecular shape using the square roots of each of the principal moments of inertia (PMIs). These descriptors will be referred to as length, width and thickness. A file of pre-generated ligand conformations is used as input to each of the methods.

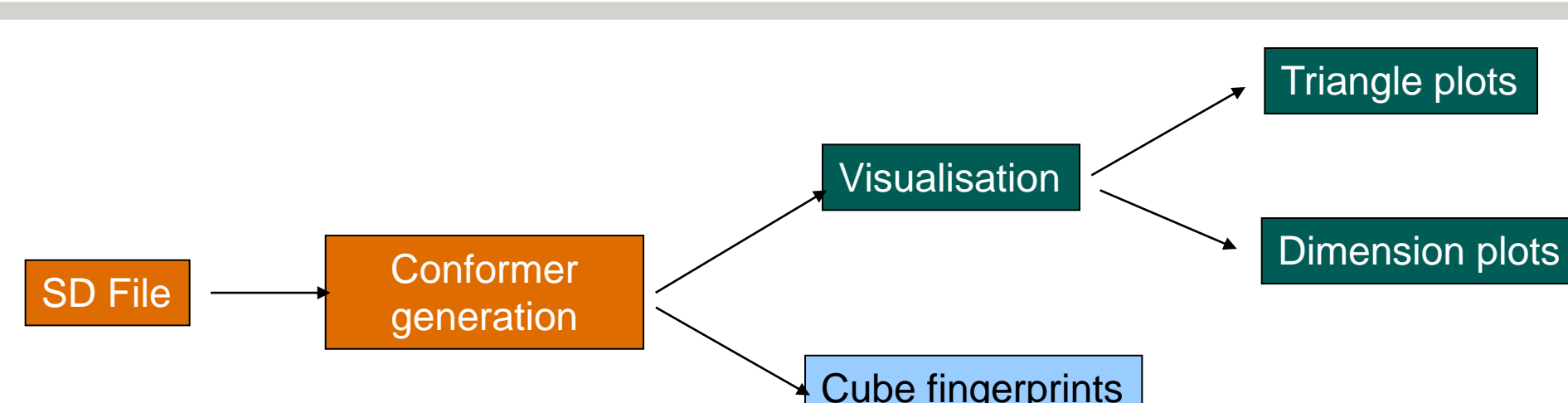


Figure 1. Workflow schematic

Triangle plots

These graphs, originally developed by Sauer and Schwarz² and enhanced by BioFocus, provide a 2D depiction of the shape of a library in a size independent manner.

Dimension plots

3D plots using the square root of each PMI as length, width and thickness provide a readily interpretable means of visualising the shape space of a library. Includes size information.

Cube fingerprints

By dividing the dimension plot into cells, and then unpacking these cells, binary fingerprints can be obtained which code the presence or absence of molecule conformations that have particular shapes.

Whilst the triangle and dimension plots focus on visualisation, the fingerprint method is more suited to library assessment through profiling and comparison with other collections.

Cube fingerprints

Cube fingerprints are derived from dimension plots where maximum values of 8Å, 4Å and 3Å have been set for each of the 3 axes respectively based on a large scale analysis of fragment shapes.

Each of the axes in the dimension plot can be split into 1Å sections, to give a set of 1Å³ cells or cubes.

The plot can then be seen as a collection of 8x4x3 cubes, which can be unpacked to give a 96 bit fingerprint.

Each molecule conformation would produce one point on the dimension plot and therefore occupies one cube, corresponding to one bit in the shape fingerprint.

The fingerprints for each conformation of one molecule can be combined (presence/absence) to give a molecule fingerprint.

Fingerprints can be compared at the molecular level or summed to give a library shape profile.

The number of bits set for a molecule is a flexibility measure. Molecules most representative of a particular shape are identified.

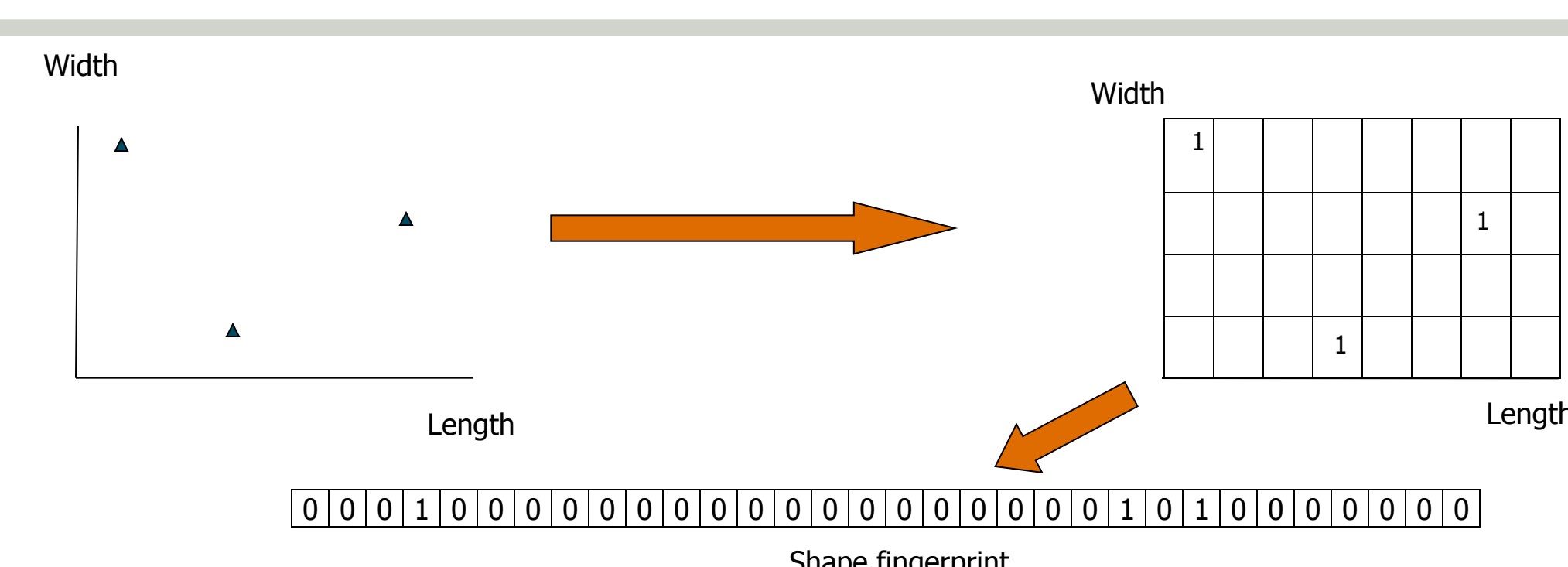


Figure 4. Generation of a cube fingerprint – 2D illustration for molecule with three distinct conformers

Molecule A	00010000000000000100000000010100000100	5
Molecule B	00010101000000000000000000010100000100	3
Molecule C	00000000000100000000000000000000000000	1
Molecule D	00010000000000000000000000000000000110	2
	00030100100100000100000002010000110	

Figure 5. A 32 bit example. Shape profile in orange and flexibility profile in green.

Shape profiles can be compared for different library subsets eg the molecules occupying one to three shape bins against those occupying >six shape bins.

Visualisation

Triangle plots

Excellent means of 2D visualisation of molecule shapes in a library. Enhanced to improve compound spread along the axes. Variable shading (triangle pixelation) aids in differentiating dense regions of the plot.

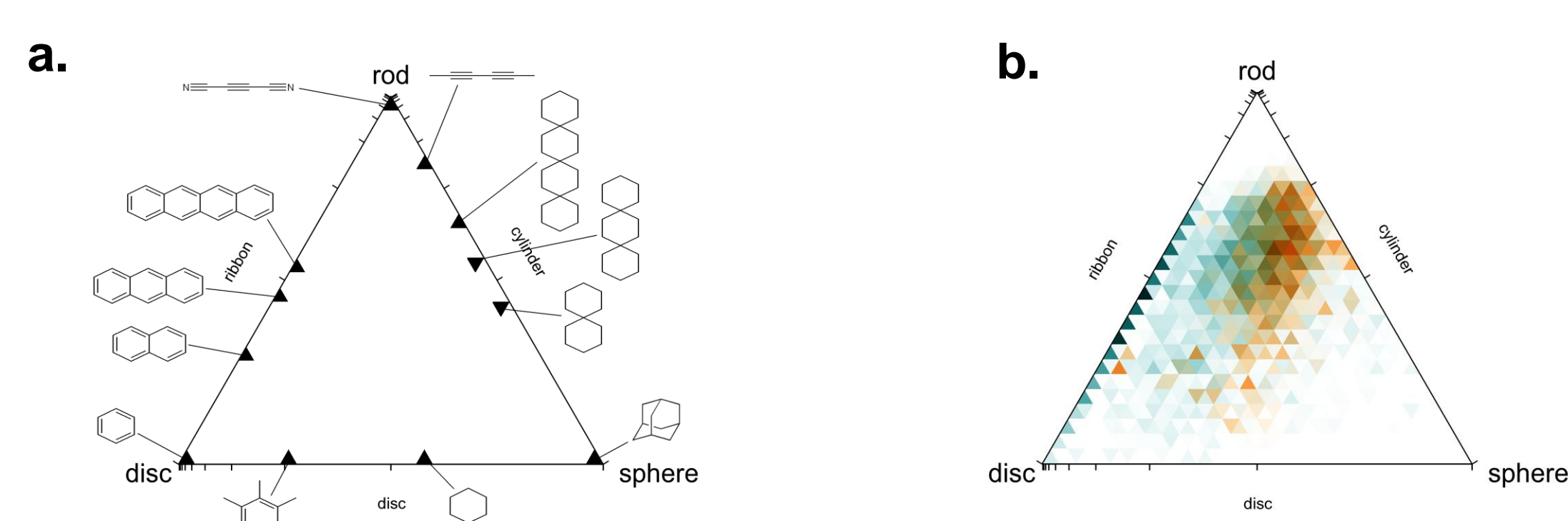


Figure 2. a) Example of original plot for representative compound shapes b) enhanced plot showing comparison of BioFocus shape collection (orange) vs the commercial library (green)

Dimension plots

Dimension plots, where the axes correspond to length, width and thickness, give an intuitive representation of compound shapes in 3D.

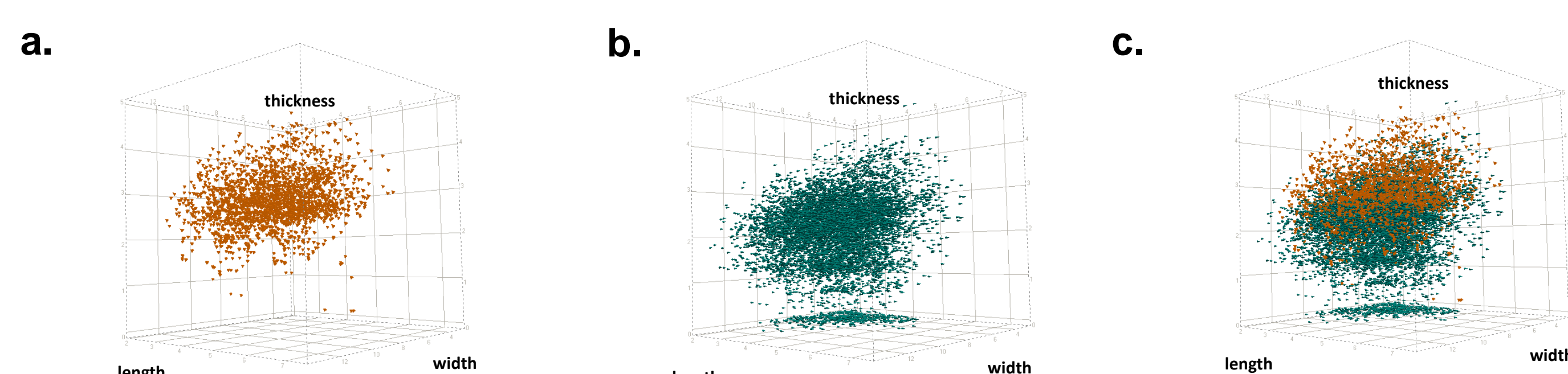


Figure 3. a) Dimension plot for the BioFocus shape collection b) dimension plot for the commercial library; c) comparison of the 2 datasets.

The BioFocus shape collection (FRG04) is one of four proprietary fragment libraries. It contains approximately 100 diverse compounds synthesised to provide bespoke fragments with a higher degree of three-dimensional shape character, consistent with current trends in fragment library composition.³ The comparison with the commercial library shows that the BioFocus shape collection contains few flat compounds, whereas the commercial library has a pronounced cluster of these. This can be seen on both the triangle plot and the dimension plot. The BioFocus shape collection also has thicker compounds, consistent with more 3-dimensionality, as shown in Figure 3.

The visual analysis provides a powerful way to interrogate the shape coverage of a library. However, to compare large numbers of compounds in an automated way, a shape based fingerprint approach, i.e. the cube fingerprints, has been developed.

Test case

The BioFocus shape collection (100 compounds) and the commercially available fragment library (1491 compounds) are compared using cube fingerprints.

Together, they cover 53 shape bins, with three shapes being unique to BioFocus and 12 being unique to the commercial library. This gives an average shape efficiency (coverage/molecule) of 0.56 for the BioFocus set compared to 0.04 for the commercial library, indicating the enrichment in shape of the BioFocus collection.

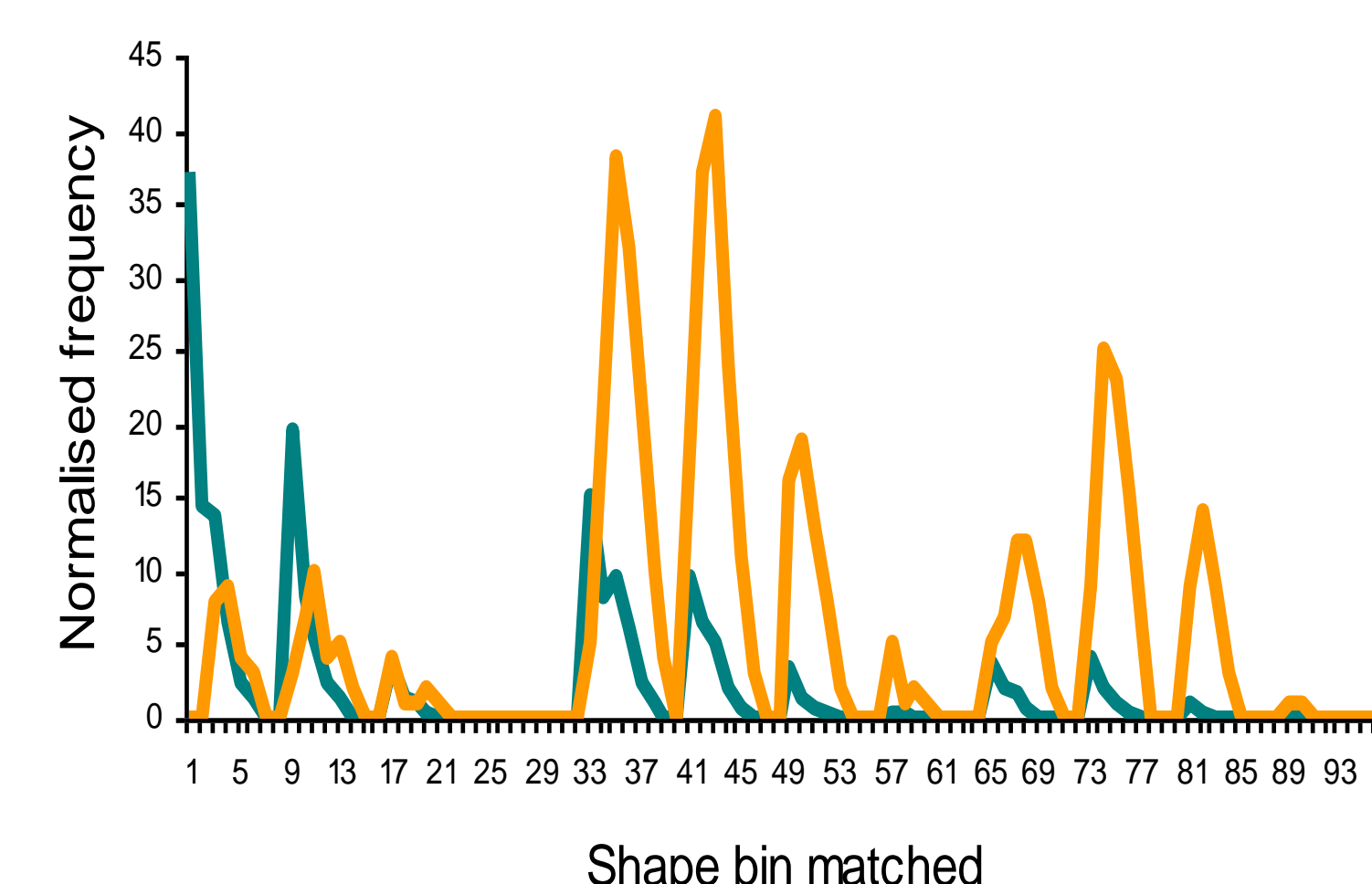


Figure 6. Shape profile analysis. BioFocus shape collection (orange) vs commercial library (green)

The number of shape bins a fragment occupies gives an indication of its flexibility. We consider those fragments that occupy three or fewer shape bins as rigid. Using this metric we observe that the commercial set contains a higher percentage of rigid molecules, which is biased by the proportion of flat aromatic members. Flat compounds (thickness <1Å) comprise 32% and 0% of the commercial and BioFocus collections, respectively. After removal of the flat compounds, approximately half of the members in either the BioFocus or commercial collections occupy one-to-three shape bins indicating that the BioFocus FRG04 collection has a higher shape efficiency but not at the expense of library flexibility.

Conclusions

- We have developed novel ways to analyse the shape profiles of libraries
- Visualisation and fingerprinting approaches give complementary information
- The cube fingerprint approach allows rapid profiling of multiple libraries and can be used to aid hole filling to improve fragment collections
- This analysis demonstrates the non-planar nature of the compounds in the FRG04 collection and their high level of shape efficiency
- The tools are being applied in the development of the next generation of BioFocus collections

References

1. Congreve et al, (2003). *Angew. Chem Int. Ed. Eng.* 42(37), 4479-4482
2. Sauer and Schwarz, (2003), *J. Chem. Inf. Comp. Sci.* 43(3), 987-1003
3. Lovering et al, (2009). *J. Med. Chem.* 52(21), 6752-6756