Using Spark to Generate Library Design Ideas

M. Slater[†], T. Cheeseright[†]

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

Abstract

The design of a focused library depends very much on the starting point. In this example, Spark is used to generate ideas for possible starting points for a library based on S-adenosyl methionine (SAM), a co-factor used as a biological methylation synthon.

Method

S-adenosyl methionine (SAM) is a co-factor used as a biological methylation synthon. It is employed in a host of enzymatic methyl transferase processes which are important in a number of disease areas. In the area of epigenetics the lysine methyl transferases KMTs are responsible for methylating lysine groups on histones – a process which mediates gene expression by changing the stability of the nucleosome.

A quick analysis of the binding conformation of SAM across the PDB (figure 1) reveals that a small number of clusters of SAM bioactive conformations are observed. The conformation of SAM found in KMTs form a tight cluster, which is distinct from the more diverse generic SAM utilizing enzymes. Interestingly, the analysis shows that DOT1L, which is also thought to be a KMT, is an outlier and more closely related to the generic enzyme set than to the other KMTs.

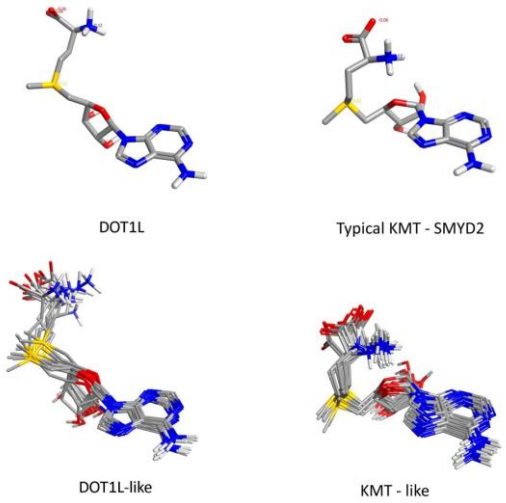

Figure 1: SAM conformations from SAM-utilizing enzymes observed from the PDB.

Assuming that it is optimal to pursue a SAM mimetic design as a paradigm for KMT or DOT1L inhibitor generation, then from a molecular design point of view there are a number of issues which would need to be addressed. One major issue already given is that SAM is ubiquitously used as a cofactor thus a close mimetic may have unwanted side interactions. Clearly a DOT1L SAM mimetic design will have more issues with generic SAM enzyme crossover. A design aimed at other KMTs (e.g. SMYD2) would have selectivity issues just within the specific KMT family.

Designing away from potential crossover activity could be achieved by a full SAM mimetic design since both the adenine and Met chains adopt different vectors and shapes in the

different sub-classes. Alternatively, concentrating on the adenine mimetic alone, the H-bonding patterns and solvent exposure are distinct in the two enzyme sub-classes as shown in figure 2.

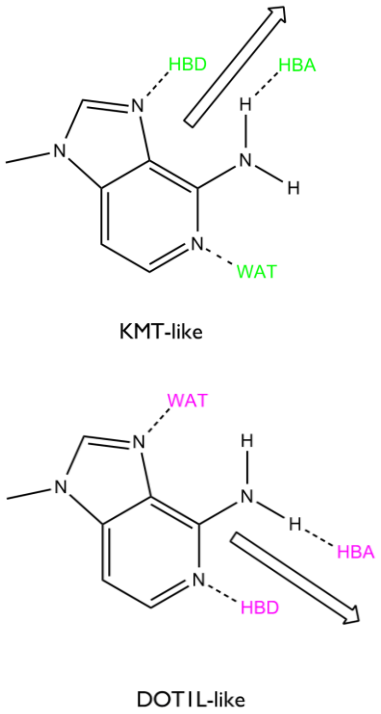

Figure 2: SAM Adenine hydrogen bonding contact vector. Differences in recognition of adenine in the two 'DOT1L-like' v 'KMT-like' systems.

This simple example shows how some background knowledge on the system can impact on the scope and potential success of any given design.

Cresset's fragment replacement tools can be used to search for novel bioisosteric replacements. In this case using the Spark software with adenine as the molecular input means that you can find suitable replacements as seeds for a library. As the template is extracted from a protein context all the ideas would be generated in the same coordinate frame and thus can be visualized and assessed for fit into the protein.

Alternatively the whole SAM 3D conformation from whichever sub-class could be submitted to Blaze to search for commercial vendor molecules that fit specific field patterns from the specific SAM conformation.

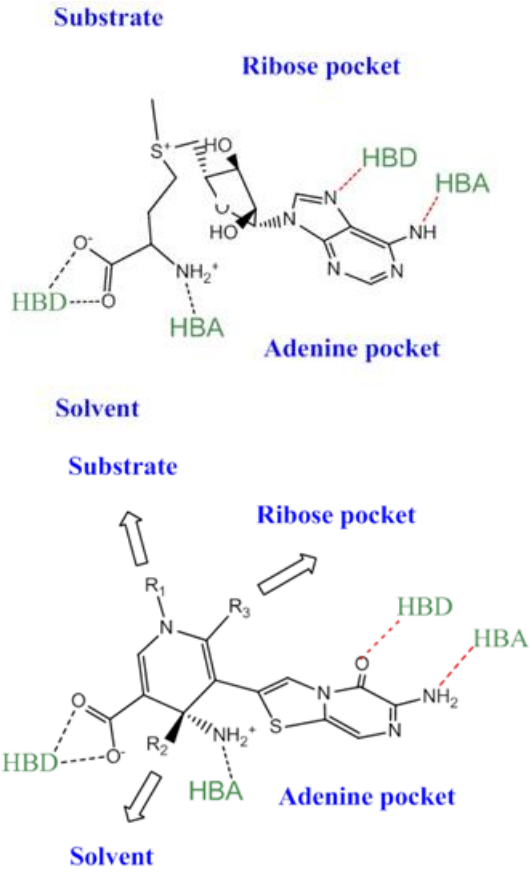

Figure 3: Library design idea for a SMYD-like KMT inhibitor (Top: SAM from SMYD2 and Bottom: virtual molecule).

The output of these virtual exercises, rather than being molecules to test (which is the usual scenario) would be molecular scaffolding ideas that would be potential starting seeds for a design. Ideally we would be looking for a good molecular fit to the interaction patterns (figure 3) and especially to those which also provide appropriate synthetic vectors from which to explore the allowed variation defined from the starting binding pose.

In this case Spark has provided us with a design idea which matches well to the field patterns and interaction patterns required by the KMT SAM conformation in SMYD2 (PDB: 3S7F) and provides three potential vectors for a library: R1 for the substrate pocket, R2 for the open solvated pocket, R3 for the ribose pocket (figures 3 and 4).

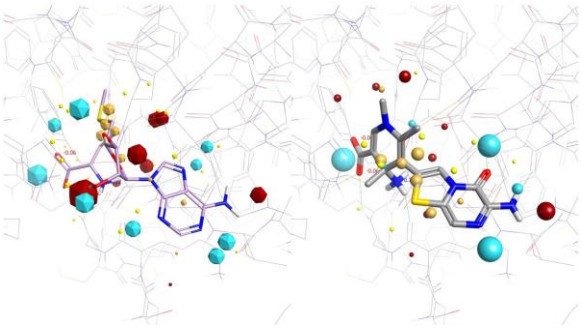

Figure 4: Interaction patterns and putative library design substitution vectors.

A standard protocol for constructing the library might proceed as follows:

Synthetically accessible variants (i.e., commercially available building blocks) of the above library would be gathered and a method outlined, possibly involving intermediate route scouting for incorporating R2 and R3 variants first and then a final array fulfilled by elaborating R1.

A virtual 'all-combinations' library would be constructed and the enumerated library analyzed in terms of predicted 'pharmaceutical-like' properties [MWT, LogP, TPSA, (HBD, HBA, Rot.bnd)-counts etc.]. Combinations which provide poor properties would be discarded.

Chemistry validation of the synthetic route and scope for the decoration transformations would be established followed by stability studies on a sub-set before final synthetic library construction and purification and plating (i.e., 96 well plates for screening).

Conclusion

This example shows how library design ideas can be generated using Spark. The library that is actually generated will depend on a number of choices made by the computational chemist, depending on the desired outcome.