# Chemotype bias in virtual screening: the elephant in the room

M Mackey [1]
T Cheeseright [1], J Melville [1], S Rose[1], A Vinter [1]
[1] *Cresset BMD, Welwyn Garden City,UK*

The analysis of virtual screening methods is complicated by the "chemotype problem": the sets of actives used are often present as clusters of highly similar molecules. Failure to correct for this can lead to spuriously high enrichment rates being performed. For example, if a large proportion of the actives come from a congeneric series, then a 2D fingerprint-based search using a molecule from that series will appear to perform exceptionally well, while in reality providing little useful information.

Very few literature assessments of virtual screening studies have attempted such a correction. These all err either in the direction of downweighting clustered actives too much[1], or by inadvertently applying a bias which can make random results appear to be highly significant.[1,2,3] In addition, the commonly used metrics used to assess virtual screens (enrichment factors, ROC, BEDROC etc) all have flaws: either they are insensitive to performance on the early part of the retrieval curve or they are oversensitive to the precise number of actives and inactives used.[2] We present a modification of the BEDROC metric called BAROC which is intuitive, concentrates on early performance, is interpretable, and is insensitive to the size of the dataset, and suggest its use as the standard metric for measuring early enrichment in VS studies.

The application of the chemotype correction to BAROC leads to the CC-BAROC metric (Figure 1).

$$CC - BAROC(\alpha) = \sum_{i=1}^{m} w_i \sum_{j=1}^{k_i} e^{-\alpha \beta_{ij}}$$

Figure 1. $m$ is the number of chemotypes, $k_m$ is number of actives assigned to chemotype $i$, $\beta_{ij}$ is the false positive of the $j^{th}$ active in the $i^{th}$ chemotype, and $w_i$ is a weighting factor depending on the correlation between the VS scoring method and the clustering method

Applying CC-BAROC to literature data sets[4,5] reveals that ignoring chemotype bias often leads to erroneous conclusions about the efficacy of VS methods, and suggests that many published VS studies greatly overstate the effectiveness of VS techniques.

1. Clark, R. D., Webster-Clark, D. J., *J. Comput. Aided Mol. Des*., in press, DOI 10.1007/s10822-008-9181-z
2. Good, A. C. *et al.*, *J. Mol. Graph. Model.*, **2003**, 22, 31-40
3. Warren, G. L. *et al*, *J. Med. Chem.*, **2006**, 49, 5912-5931
4. N. Huang, B. Shoichet, J.J. Irwin; *J. Med. Chem*. **2006**, 49, 6789-6801
5. Jain, A. N., *J. Comput. Aided Mol. Des*., in press, DOI 10.1007/s10822-007-9151-x