Uncovering Activity Cliff-Forming Compounds using SALI Values

José L. Medina-Franco1,*, Karina Martínez-Mayorga2, Kai Fang Cheng3, Mingzhu He3, Yousef Al-Abed2, Nathalie Meurice3

1Department of Research, Mayo Clinic, Scottsdale, AZ 85295, USA
2Instituto de Química, Universidad Nacional Autónoma de México, Mexico City 04510, Mexico
3Feinstein Institute for Medical Research, Center for Molecular Innovation, Manhasset, NY 11030 USA

Abstract

Activity cliffs have a significant impact on a number of tasks relevant in medicinal chemistry and chemoinformatics such as lead optimization, development of methods to predict biological activities and selection of queries for similarity searching. Therefore, the identification of compounds highly associated with activity cliffs, in biological data sets i.e., ‘activity cliff generators’, is of prime significance. The identification of activity cliff generators based on Structure-Activity Similarity (SAS) maps and frequency counts has been reported previously. Herein, we propose a complementary approach to identify activity cliff generators based on the distribution of the Structure-Activity Landscape Index (SALI) values. We discuss the SAR and the activity cliff generators identified in screening data sets from our effort to identify inhibitors of the migration inhibitory factor (MIF).

Methods

A data set of 79 synthetic compounds with reported K values against MIF was analyzed. This set of compounds was previously used to develop QSAR models. All compounds were evaluated using comparable bioassay methodologies.

Seven 2D fingerprint representations were computed with Canvas (Schrödinger); MACCS keys, radial, atom pairs, topological, Molprint 2D, atom triplets, and dendritic. Similar structures were computed with the Tanimoto coefficient. The presence of activity cliffs was evaluated quantitatively computing the SALI values proposed by Guha where A and A are the activities of the jth and jth molecules and sim(i,j) is the similarity coefficient between the two molecules. Mean SALI values were computed using the mean structure similarity of four selected 2D and 3D molecular representations.

In order to identify activity cliff generators all pairs of molecules with a high mean SALI value were identified. To define 'high' SALI value we computed the mean and standard deviation of the distribution of the mean SALI values and then selected the pairs of compounds with SALI values greater than two standard deviations. Then, for each compound, we calculated the frequency of occurrence among the pairs highly-ranked. Activity cliff generators were selected as the molecules with the highest frequencies. This approach is reminiscent of the method proposed by Méndez-Lucio et al. to identify individual molecules associated with high selectivity in a chemogenomics data set. Of note, the 'high' mean SALI value depends on the distribution of the activity and similarity data of the particular data set.

As reference we also generated the consensus SAS maps using a standard protocol for each of the 3,081 pairs of compounds we plot the absolute value of the potency difference vs. the mean structure similarity. The later measure was the same used to compute the mean SALI values. Each data point was further distinguished by the most active compound in the pair.

Results and discussion

Fig. 1 shows the different design of the 4 selected fingerprints. Fig. 2 and Table 1 indicate that molecule 50 is the top-ranked activity cliff generator and that compound 34 (Fig. 3) is the most active activity cliff generator. Four out of the 10 top-ranked activity cliff generators are active (K < 1 μM).

Fig. 3 illustrates the chemical diversity of activity cliff-forming compounds in this data set as determined by the 'high' values of mean SALI. It is worth noting the very large potency difference of each pair of compounds.

Fig. 4A shows the position of the 5 representative activity cliffs in a consensus SAS map. All data points are located in the activity cliff quadrant of the plot. In contrast, Figs. 4B, C that show the distribution of the 78 pairs of molecules with two selected cliff-forming compounds, illustrate that several, but not all data points have high SALI values.

Conclusions

- An approach is proposed to identify activity cliff-forming compounds based on the distribution of SALI values. The method is focused on pairs of compounds with the highest SALI values.
- A mean SALI value is calculated to reduce the dependence of chemical space with structural representation.
- The first activity landscape modeling of MIF inhibitors is reported.
- Several activity cliff-forming MIF inhibitors were identified, including the most active compound in the data set.
- Activity cliff-forming compounds depend on the data set. A visual inspection is recommended to confirm interpretability.

References


Acknowledgments

We thank Jacob Waddell for writing scripts used in this work.

© 2013 Mayo Foundation for Medical Education and Research