Transforming data into knowledge for intelligent decision-making in early drug discovery

Paricharak, S.A.

Propositions

Transforming Data into Knowledge for Intelligent Decision-making in Early Drug Discovery

1. A screening paradigm involving an iterative expansion around hits using biological and chemical similarity metrics in parallel can result in greatly improved efficiency over conventional high-throughput screening campaigns.

   This thesis, *Chapter four*

2. Descriptors derived from atom topology show a strong correlation with fingerprint-based descriptors with respect to diversity assessment of a number of compound sets ranging in size, diversity and origin.

   This thesis, *Chapter three*

3. The Bayes Affinity Fingerprints showed the best performance in selecting compound sets that are diverse in bioactivity space, compared to twelve other commonly used fingerprint-based, pharmacophore-based, shape-based, connectivity-matrix-based, and physicochemical-property-based descriptors.

   This thesis, *Chapter three*

4. The concept of active learning can be used to derive compound sets that, once screened and used as training sets for machine learning, predict bioactivity of unknown compounds better than when randomly selected training sets are used.

   This thesis, *Chapter five*

5. Despite improvements in screening technology, high-throughput screening campaigns are still costly due to the large amount of resources required in relation to the number of active compounds discovered.

   Phatak, S. S., Stephan, C. C., and Cavasotto, C. N. (2009) High-throughput and in silico screenings in drug discovery. *Expert. Opin. Drug Discov. 4*, 947–959.

6. Even though over 10^{63} drug-like molecules possibly exist, it is likely that only a fraction of these molecules is therapeutically relevant. This is evidenced by the success of high-throughput screening campaigns comprising "only" 10^6 molecules.

   Lipinski, C., Hopkins, A. (2004) Navigating Chemical Space for Biology and Medicine. *Nature 432*, 855–861.

7. Efficient exploration of relevant chemical space is important for targets with few known active chemotypes or phenotypic assays. Diversity-based library design is aimed at addressing this need by optimizing biological relevance and compound diversity to provide multiple starting points for further development.

   Roth, H. J. (2005) There is no such thing as "diversity"! *Curr. Opin. Chem. Biol. 9*, 293–295

8. Chemical diversity is an ambiguous concept, as it can be based on a wide range of chemical or biological descriptors.

   Roth, H. J. (2005) There is no such thing as "diversity"! *Curr. Opin. Chem. Biol. 9*, 293–295

9. At the strategic level, the game of chess bears resemblance to a PhD: competence in the opening (action plan), middlegame (execution), and endgame (completion and communication of results) is essential for success.

10. Although the personal contribution of a PhD candidate is paramount, the overall productivity is highly dependent on effective collaboration with

the advisors and fellow researchers. From this point of view, an analogy to an a cappella choir can be drawn: the vocal support of all artists involved makes or breaks the impact of the lead artist.