K04

In Silico Screening using Active Learning with Descriptor Sampling

([1]NEC Corp., [2]TANABE SEIYAKU Co., Ltd.) ○Yukiko Fujiwara (1), Yoshiko Yamashita (1), Tsutomu Osoda (1), Minoru Asogawa (1), Masaaki Asao (2), Hideshi Shimadzu (2), Kazuya Nakao (2), Chiaki Fukushima (2), Ryo Shimizu (2)

In hit-finding processes, more hits and much structure-activity relationship information, is required to be extracted through biochemical experiments. For this purpose, we developed the active learning method using descriptor sampling, which captures many aspects of chemical information. To evaluate our methods, we used the structure-activity data of the compounds which bind to GPCRs (G-protein coupled receptors). In this study, we have used 1,551 compounds which bind to biogenic amine receptors selected from Pharmaprojects as the positives, and 256,991 compounds selected from Pharmaprojects and Available Chemicals Directory (ACD) as the negatives. We have shown that descriptor sampling is efficient for hit-finding and prediction. Based on our method, we have shown that 11% of biochemical experiments of the whole chemical library are enough for finding 90% of positives, and 32% are enough for finding 99% of positives. On the contrary, for finding 90% (or 99%) of positives, 90% (or 99 %) of biochemical experiments were necessary when random screening methods were conducted, and 22% (or 61%) of biochemical experiments were necessary when similarity-based screening methods were conducted. Using active learning method, we developed rapid and cost-effective systems for hit-finding.