

## Knowledge Base of Basic Active Structures from Twenty-eight-day Repeated Dose Toxicity Test Data in Rats

Takashi Okada, Norihito Ohmori, Sachio Mori, Hiroshi Horikawa, Masumi Yamakawa, Takashi Okada, Yuki Sakuratani, Takashi Yamada, Makoto Hayashi

Chemical compounds affecting a bioactivity can usually be classified into several groups, each of which shares a characteristic substructure. We call these substructures “basic active structures” or BASs. Data mining technology has enabled the systematic elaboration of BASs. We applied the method to 141 low-molecular-weight organic compounds selected from twenty-eight-day repeated dose toxicity test data in NEDO SAR project database. The results are now disclosed on the knowledge BASiC.

The process of BAS extraction is as follows. (i) Linear fragments were first extracted from the molecules. (ii) Cascade model, a data mining method, was applied to create characteristic rules for each activity. (iii) Each rule was then examined using a structural refinement system, in which an initial core substructure appeared in the rule incorporates the surrounding atoms and bonds, and increases the discriminating capability between active and inactive compounds. (iv) Experienced users easily recognize BAS candidates in the supporting structures, as the structural diversity and the number of compounds become relatively limited. Finally, a BAS candidate is run through the refinement system to confirm its ability to perform the desired activity. Steps (iii) to (iv) are repeated until the extracted BASs are those found in most of the active compounds.

Application to hemolytic anemia found glycol ether as a BAS in addition to a well known aromatic amine. An independent survey of toxicity mechanisms in the project has also found this substructure as the potential cause of toxicity and approved the usefulness of the method.