

more inefficient, while our $3D$ -score is designed to also find these hits. (see some examples for both cases in Fig. 2).

For this type of hits the sophisticated approach is clearly superior to the fast algorithm. Comparing the numbers in Table 5 with those in Table 3 and Table 4 shows that most of the hits for which the two $3D$ -approaches differ can be found in the class discussed here. The main reason for this is that the fast approach is not able to find hits for two molecules that have different overall geometry, in particular small molecules that are substructures of larger ones are not found (see Fig. 3).

Algorithm		ATC code		
$3D$ -soph.	$3D$ -fast	N05A,N06,R06,D04	others	Σ
> 0.75	> 0.75	130	22	152
> 0.75	< 0.75	27	21	48
< 0.75	> 0.75	1	3	4

Table 5: Number of hits: fast $3D$ vs. soph. $3D$

Ad c. In this set we find those hits, that are quite similar in both chemical structure and size. They are reported as relevant by both approaches. For this type of hits both strategies are most similar. Here again, as in case a., the fast and the sophisticated $3D$ -algorithm perform comparably.

From our point of view we have therefore seen several strong arguments in favour of the $3D$ -superposition algorithms. It can be clearly seen that the $3D$ -approach is able to detect similar activity and similar adverse reaction, even with this seemingly simple, purely geometry-based scoring function.

For large data sets a fast $3D$ -superposition algorithm combined with Tanimoto coefficients helps to increase the set of relevant hits.

If one aims to really find all, at least geometrically relevant hits – this may be important for smaller and more specific sets of molecules – it is worth while to follow the sophisticated $3D$ -approach (with a somewhat smaller threshold for relevance). We were able to find really relevant hits that can not be found by simple $2D$ -methods or by the fast $3D$ -algorithm.

Discussion

In agreement with our results it is shown in [27] and [28] that $3D$ similarity searches retrieve compounds with more diverse topology while $2D$ similarity works best when the query molecule contains relatively rare and distinct topological features that are responsible for the biological activity. $2D$ similarity works poorly when common functional groups as in peptides are considered. A similar fragment- or topomer-based steric shape screening was shown to be more selective than $2D$ similarity [13], especially advantageous "lead-hopping" was observed. A reasonable speed for the in silico screening of large compound libraries can be achieved by full-atom superposition procedures as presented in

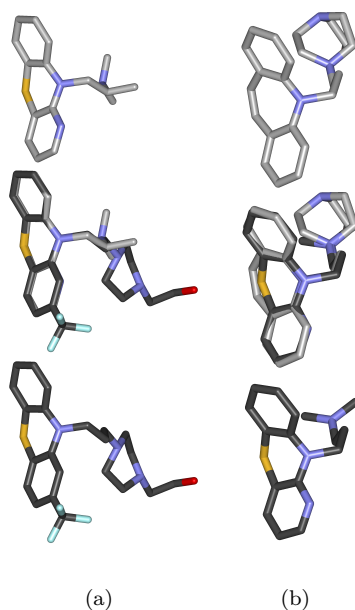


Figure 2: Comparisons with low Tanimoto coefficients and 3D scores above threshold.

(a) Superposition of fluphenazine (bottom, ATC code: N05AB02) with isothipendyl (top, ATC codes: D04AA22, R06AD09) with a 3D score of 0.81 and a Tanimoto coefficient of 0.69. The resemblance to isothipendyl, an antihistaminic agent, is neglected by the 2D similarity measure because of missing chemical groups (trifluoromethyl, piperazin) and quite different sizes of the molecules.

(b) Superposition of prothipendyl (bottom, ATC code: N05AX07) and opipramol (top, ATC code: N06AA05) with a 3D score of 0.76 and a Tanimoto coefficient of 0.72. The similarity to opipramol, an antidepressant, is missed by 2D comparison because the middle ring is seven membered in opipramol (dibenzazepine derivative) and six membered in prothipendyl (azaphenothiazine derivative).

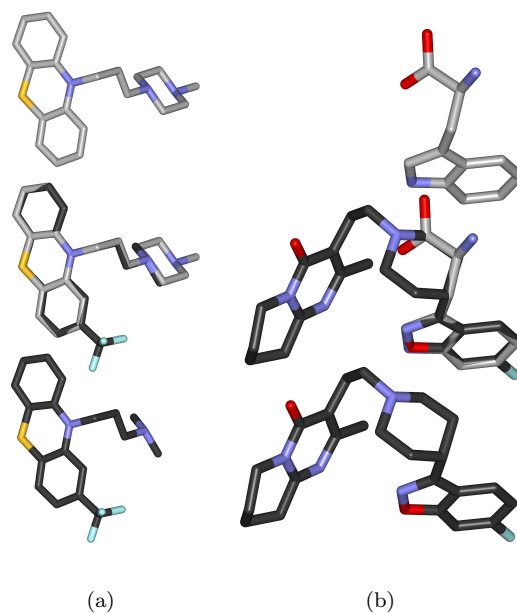


Figure 3: Differences between fast and sophisticated superpositions.

(a) Superposition of perazine (top, ATC code: N05AD10) with triflupromazine (bottom, ATC code: N05AA05) with a $3D$ score of 0.77 (sophisticated), 0.50 (fast) and a Tanimoto coefficient of 0.84. The resemblance between the two neuroleptics is neglected by the fast superposition algorithm because the centers of gravity do not fit.

(b) Superposition of tryptophan (top, ATC code: N06AX02) and risperidone (bottom, ATC code: N05AX08) with a $3D$ score of 0.77 (sophisticated), 0.50 (fast) and a Tanimoto coefficient of 0.44. The similarity to tryptophan, an antidepressant, is missed by the fast superposition algorithm because of the very different overall geometry of the molecules.

this analysis.

With receptor structures available ligand-docking programs have been shown to enrich hit lists of in silico screening approaches [45] but in the case of psycholeptics a number of structurally unknown receptors are engaged. Most of the processes involved in ADME are driven by rather unspecific interactions between drugs and macromolecules but drug transporters and cytochromes gained increased interest in early ADME profiling via similarity based structure activity relation (SIBAR) [46]. The increased predictive power of the 3D- vs. 2D-similarity for side effects demonstrated in this analysis gives rise to the hope that improvements in ADME and toxicity profiling will be possible.

Limitations of the fast 3D superposition approach are spherical compounds for which it might fail to find proper assignments. The known size bias and size limitation of 2D similarity measures [44] also may cause problems for the fast algorithm.

The conformer generation is a general problem because the 3D similarity between two structural ensembles depends critically on the original structures, the conformer generation [22] and clustering [47] algorithm, the parameters like energy threshold, and the number of conformers per compound. In particular the number of rotatable bonds will restrict the 3D similarity approach or will require new algorithms [48].

References

- [1] Voigt, J. H.; Bienfait, B.; Wang, S.; Nicklaus, M. C. Comparison of the NCI open database with seven large chemical structural databases. *J Chem Inf Comput Sci* 2001, 41, 702-712.
- [2] Bradley, M. P. An overview of the diversity represented in commercially-available databases. *Mol Divers* 2002, 5, 175-183.
- [3] Matter, H. Selecting optimally diverse compounds from structure databases: a validation study of two-dimensional and three-dimensional molecular descriptors. *J Med Chem* 1997, 40, 1219-1229.
- [4] Brown, R. D.; Martin, Y. C. An evaluation of structural descriptors and clustering methods for use in diversity selection. *SAR QSAR Environ Res* 1998, 8, 23-39.
- [5] Delaney, J. S. Assessing the ability of chemical similarity measures to discriminate between active and inactive compounds. *Mol Divers* 1996, 1, 217-222.
- [6] Martin, Y. C.; Kofron, J. L.; Traphagen, L. M. Do structurally similar molecules have similar biological activity? *J Med Chem* 2002, 45, 4350-4358.

- [7] Chen, X.; Reynolds, C. H. Performance of similarity measures in 2D fragment-based similarity searching: comparison of structural descriptors and similarity coefficients. *J Chem Inf Comput Sci* 2002, 42, 1407-1414.
- [8] Whittle, M.; Willett, P.; Klaffke, W.; van Noort, P. Evaluation of similarity measures for searching the dictionary of natural products database. *J Chem Inf Comput Sci* 2003, 43, 449-457.
- [9] Dixon, S. L.; Merz, K. M., Jr. One-dimensional molecular representations and similarity calculations: methodology and validation. *J Med Chem* 2001, 44, 3795-3809.
- [10] Wildman, S. A.; Crippen, G. M. Three-dimensional molecular descriptors and a novel QSAR method. *J Mol Graph Model* 2002, 21, 161-170.
- [11] Turner, D. B.; Willett, P. The EVA spectral descriptor. *Eur J Med Chem* 2000, 35, 367-375.
- [12] Hecker, E. A.; Duraiswami, C.; Andrea, T. A.; Diller, D. J. Use of catalyst pharmacophore models for screening of large combinatorial libraries. *J Chem Inf Comput Sci* 2002, 42, 1204-1211.
- [13] Cramer, R. D.; Jilek, R. J.; Andrews, K. M. Dbtop: topomer similarity searching of conventional structure databases. *J Mol Graph Model* 2002, 20, 447-462.
- [14] Patel, Y.; Gillet, V. J.; Bravi, G.; Leach, A. R. A comparison of the pharmacophore identification programs: Catalyst, DISCO and GASP. *J Comput Aided Mol Des* 2002, 16, 653-681.
- [15] Lemmen, C; Lengauer, T. Computational methods for the structural alignment of molecules. *J Comput Aided Mol Des* 2000,14,215-232.
- [16] Lemmen, C.; Lengauer, T.; Klebe, G. FLEXS: a method for fast flexible ligand superposition. *J Med Chem* 1998, 41, 4502-4520.
- [17] Kramer, A.; Horn, H. W.; Rice, J. E. Fast 3D molecular superposition and similarity search in databases of flexible molecules. *J Comput Aided Mol Des* 2003, 17, 13-38.
- [18] Klebe, G.; Mietzner, T.; Weber, F. Methodological developments and strategies for a fast flexible superposition of drug-size molecules. *J Comput Aided Mol Des* 1999, 13, 35-49.
- [19] Furukawa, H.; Hamada, T.; Hayashi, M. K.; Haga, T.; Muto, Y.; Hirota, H.; Yokoyama, S.; Nagasawa, K.; Ishiguro, M. Conformation of ligands bound to the muscarinic acetylcholine receptor. *Mol Pharmacol* 2002, 62, 778-787.

- [20] Mills, J. E.; de Esch, I. J.; Perkins, T. D.; Dean, P. M. SLATE: a method for the superposition of flexible ligands. *J Comput Aided Mol Des* 2001, 15, 81-96.
- [21] Barnum, D.; Greene, J.; Smellie, A.; Sprague, P. Identification of common functional configurations among molecules. *J Chem Inf Comput Sci* 1996, 36, 563-571.
- [22] Smellie, A.; Stanton, R.; Henne, R.; Teig, S. Conformational analysis by intersection: CONAN. *J Comput Chem* 2003, 24, 10-20.
- [23] Weston, J.; Perez-Cruz, F.; Bousquet, O.; Chapelle, O.; Elisseeff, A.; Scholkopf, B. Feature selection and transduction for prediction of molecular bioactivity for drug design. *Bioinformatics* 2003, 19, 764-771.
- [24] Schuffenhauer, A.; Gillet, V. J.; Willett, P. Similarity searching in files of three-dimensional chemical structures: analysis of the BIOSTER database using two-dimensional fingerprints and molecular field descriptors. *J Chem Inf Comput Sci* 2000, 40, 295-307.
- [25] Bostrom, J.; Bohm, M.; Gundertofte, K.; Klebe, G. A 3D QSAR study on a set of dopamine D4 receptor antagonists. *J Chem Inf Comput Sci* 2003, 43, 1020-1027.
- [26] Bostrom, J.; Gundertofte, K.; Liljeforsa, T. A pharmacophore model for dopamine D4 receptor antagonists. *J Comput Aided Mol Des* 2000, 14, 769-786.
- [27] Hahn, M. Three-Dimensional Shape-Based Searching of Conformationally Flexible. *J Chem Inf Comput Sci* 1997, 37, 80-86.
- [28] Guner, O.F.; Hahn, M.; Li, H.; Hassan, M. 2D versus 3D shape similarity: use of molecular shape-based 3D searching techniques for identifying novel compounds. Case study, Accelrys; http://www.accelrys.com/cases/2dvs3di_full.html
- [29] Kubinyi, H. Molecular similarity. 1. Chemical structure and biological action. *Pharm Unserer Zeit* 1998, 27, 92-106.
- [30] Kubinyi, H. Molecular similarity. 2. The structural basis of drug design. *Pharm Unserer Zeit* 1998, 27, 158-172.
- [31] Iwase, K.; Hirono, S. Estimation of active conformations of drugs by a new molecular superposing procedure. *J Comput Aided Mol Des* 1999, 13, 499-512.
- [32] Schneider, G.; Neidhart, W.; Giller, T.; Schmid, G. "Scaffold-Hopping" by Topological Pharmacophore Search: A Contribution to Virtual Screening. *Angew Chem Int Ed Engl* 1999, 38, 2894-2896.

- [33] Holliday, J. D.; Jelfs, S. P.; Willett, P.; Gedeck, P. Calculation of inter-substituent similarity using R-group descriptors. *J Chem Inf Comput Sci* 2003, 43, 406-411.
- [34] Maurer, I.; Volz, H. P. Cell-mediated side effects of psychopharmacological treatment. *Arzneimittelforschung* 2001, 51, 785-792.
- [35] The selection and use of essential medicines. Report of the WHO Expert Committee, 2002 (including the 12th Model list of essential medicines). World Health Organ Tech Rep Ser 2003, 914, i-vi, 1-126.
- [36] Uhl, G. R.; Vandenbergh, D. J.; Miner, L. L. Knockout mice and dirty drugs. *Drug addiction. Curr Biol* 1996, 6, 935-936.
- [37] Specialized Information Services (SIS) Division of the National Library of Medicine (NLM), 8600 Rockville Pike, Bethesda, <http://chem.sis.nlm.nih.gov>
- [38] Accelrys Inc., San Diego, CA. <http://www.accelrys.com>
- [39] Smellie, A.; Stanton, R.; Henne, R.; Teig, S. Conformational analysis by intersection: CONAN. *J Comput Chem.* 2003 Jan 15;24(1):10-20.
- [40] WHO Collaborating Centre for Drug Statistics Methodology, <http://www.whocc.no>
- [41] Willett, P.; Barnard, J. M.; Downs, G. M. Chemical similarity searching. *J. Chem. Inf. Comput. Sci.* 1998, 38, 983-996.
- [42] Daylight Chemical Information Systems, Santa Fe, NM. <http://www.daylight.com>.
- [43] Umeyama, S. Least-Squares Estimation of Transformation Parameters Between Two Point Patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1991, 13, 676-681
- [44] Holliday, J. D.; Salim, N.; Whittle, M.; Willett, P. Analysis and display of the size dependence of chemical similarity coefficients. *J Chem Inf Comput Sci* 2003, 43, 819-828.
- [45] Jenkins, J. L.; Kao, R. Y.; Shapiro, R. Virtual screening to enrich hit lists from high-throughput screening: a case study on small-molecule inhibitors of angiogenin. *Proteins* 2003, 50, 81-93.
- [46] Klein, C.; Kaiser, D.; Kopp, S.; Chiba, P.; Ecker, G. F. Similarity based SAR (SIBAR) as tool for early ADME profiling. *J Comput Aided Mol Des* 2002, 16, 785-793.
- [47] Raymond, J. W.; Blankley, C. J.; Willett, P. Comparison of chemical clustering methods using graph- and fingerprint-based similarity measures. *J Mol Graph Model* 2003, 21, 421-433.

- [48] Raymond, J. W.; Willett, P. Similarity Searching in Databases of Flexible 3D Structures Using Smoothed Bounded Distance Matrices. *J Chem Inf Comput Sci* 2003, 43, 908-916.