

Ligand.Info, automated system for virtual high-throughput screening

Marcin von Grotthuss¹ and Leszek Rychlewski¹

Keywords: ligand.info, drug design, virtual high-throughput screening, structure-based searching, small-molecule database

1 Introduction.

The Ligand.Info server is based on the assumption that small molecules with similar structure have similar pharmacological properties. The developed system enables search for similar compounds in large databases using structural indices. Index profiles, constructed by averaging indexes of related molecules are used to increase the specificity of the search. The Ligand.Info tool can interactively cluster sets of molecules on the user side and automatically download similar molecules from the server. Ligand.Info databases contain various publicly available sets of small molecules such as Harvard's ChemBank, Hetero Atoms from Protein Data Bank, KEGG Ligand Database, The Open NCI Database, and others. The total size of all databases is more than 1 million entries. All downloaded molecules are automatically displayed using the structure viewer.

2 Ligand.Info Java-based tool.

The system is available at <http://Ligand.Info>. The program has its own desktop, which allows simultaneous analysis of several projects. Figure 1 shows an demo project that can be open from server. Downloaded from the server molecules are presented in a table (left window). The chemical structure of highlighted molecule is displayed by the viewer (right window). The application requires Netscape 6.0 or IE 5.0 and the Java Runtime Environment 1.3, which can be automatically installed during the first use on MS Windows desktops. Manual installation is required by other systems such as Linux, Solaris, Macintosh, and IRIX.

3 Small-molecule Meta-Database.

Ligand.Info Meta-Database is a compilation of databases of biologically annotated small molecules and molecules for which structural information was available. The smallest part of the Meta-Database was taken from Harvard's ChemBank[1]. Despite the fact, that this set contains only 2'344 entries the molecules are well annotated. Other molecules incorporated in the Meta-Database include small-molecules and monomers that are referred to as residues and hetgroups in the Protein Data Bank[2]. The structural information for 4'009 molecules was retrieved using Ligand Chemistry service developed by the Macromolecular Structure Database (MSD) group at the European Bioinformatics Institute[3]. The Meta database was also expanded with compounds (10,006 entries) from the Ligand database of the Kyoto Encyclopedia of Genes and Genomes[4]. An other subset of Ligand.Info Meta-Database came from the public part of the National Cancer Institute (NCI) database. Its public part (250,246 compounds), known as the Open NCI database, has the largest absolute number of unique structures among the eight known large chemical

¹ BioInfoBank Institute, ul. Limanowskiego 24A, 60-744 Poznan, Poland

databases[5]. The next Ligand.Info subset contains 544,391 compounds that were collected from Russia, Ukraine, Lithuania, several other GUS states and Bahrain by AKos Consulting & Solutions GmbH. All AKos molecules have predicted biological activity spectrums including pharmacological effects, mechanisms of action, mutagenicity, carcinogenicity, teratogenicity, embryotoxicity and drug-likeness. The last subset of 348,276 molecules is donated by Asinex Ltd. Asinex's compounds have some solubility and bio-physical prediction. The current size of the Ligand.Info Meta-Database is 1,159,274 entries.

ID	FORMULA	WEIGHT	NSC	LIBAL_SCREEN...
1	C16H14O4S2	334,42 211		Moderate
2	C6H11N4PS	202,22 236		Moderate
3	C6H4N2O4	168,11 1541		Moderate
4	C14H9Cl5O2	386,49 3865		Moderate
5	C14H14N4O2S2	334,42 4492		Moderate
6	C28H18Cl4N4O2S2	648,42 4493		Moderate
7	C2H2N2S3	150,25 4645		Moderate
8	C34H26N6O6S2	678,75 5068		Moderate
9	C22H20O14	508,39 6196		Moderate
10	C12H10O4S2	282,34 7387		Moderate
11	C19H10Br2Cl2O5S	581,06 7816		Moderate
12	C27H28Br2O5S	624,39 7819		Moderate
13	C33H40O19	740,67 9222		Moderate
14	C32H24N6O15S5	892,9 11243		Moderate
15	C34H28N6O14S4	872,89 11247		Moderate
16	C20H26O2	298,43 15432		Moderate
17	C30H18FeN3O6	572,34 16086		Moderate
18	C12H8O8S2	232,33 18859		Moderate
19	C10H10N2O2S	222,27 17800		Moderate
20	C34H32ClFeN4O6	683,95 19664		Moderate
21	C13H19N3O3S12	650,1 20871		Moderate
22	C7H3BrClNO2	248,46 24970		Moderate
23	C14H12ClNO5S	341,77 25834		Moderate
24	C26H22N8O6S	574,58 26252		Moderate

Figure 1: Ligand.Info screenshot. The Ligand.Info table contains downloaded from the server (or read from the user hard disk) molecules (left window). The chemical structure of a selected compound is displayed by the molecular viewer (right window).

References

- [1] Strausberg, R. L.; Schreiber, S. L. *Science* 2003, 300(5617), 294-295.
- [2] Berman H.M.; Westbrook J.; Feng Z.; Gilliland G.; Bhat T.N.; Weissig H.; Shindyalov I.N.; Bourne P.E. *Nucleic Acids Res.* 2002, 28(1), 235-242.
- [3] Boutselakis, H.; Dimitropoulos, D.; Fillon, J.; Golovin, A.; Henrick, K.; Hussain, A.; Ionides, J.; John, M.; Keller, P. A.; Krissinel, E.; McNeil, P.; Naim, A.; Newman, R.; Oldfield, T.; Pineda, J.; Rachedi, A.; Copeland, J.; Sitnov, A.; Sobhany, S.; Suarez-Uruena, A.; Swaminathan, J.; Tagari, M.; Tate, J.; Tromm, S.; Velankar, S.; Vranken, W. *Nucleic Acids Res.* 2003, 31(1), 458-462.
- [4] Goto S.; Okuno Y.; Hattori M.; Nishioka T.; Kanehisa M. *Nucleic Acids Res.* 2002, 30(1), 402-404.
- [5] Voigt, J. H.; Bienfait, B.; Wang, S.; Nicklaus, M. C. *J.Chem.Inf.Comput.Sci.* 2001, 41(3), 702-712.