[P6] A relational database structure for fast and efficient mining of molecular interactions in the PDB

Loïc Dréano[1,2], Ashenafi Legehar[2], Evgeni Grazhdankin[2], Alexandre Borrel[2,3], Henri Xhaard[2]

[1]Molecules Thérapeutiques in silico (MTi), INSERM UMRS-973, University Paris Diderot, Sorbonne Paris Cite, 75205 Paris Cedex 13, France
[2]Faculty of Pharmacy, Division of Pharmaceutical Chemistry and Technology, University of Helsinki, Viikinkaari 5E, P.O. Box 56, FI-00014 Helsinki, Finland
[3] Division of Intramural Research/Biostatistics and Computational Biology Branch, NIH/NIEHS, RTP, North Carolina 27709, United States

There are around 130,000 protein structures in the protein data bank (PDB)[1] that provide a wealth of information about molecular interactions. Filtering the complexes of interest is usually conducted at different levels in order to build datasets where for example the resolution, protein chain and ligand redundancies[2], interactions with metals, or simply associated features have been controlled. This is usually done using collections of scripts organized into workflows[3-4]. Examples of organizing the data into databases are rare[5], yet they represent a powerful and time-efficient ways to conduct data mining studies. Here, we present a relational database, developed in PostgreSQL to mine molecular interactions in the PDB, as well as the key features.

Bibliography:

[1] H.M. Berman; J. Westbrook, Z. Feng; G. Gilliland; T.N. Bhat; H. Weissig; I.N. Shindyalov; P.E. Bourne. (2000) The Protein Data Bank Nucleic Acids Research, 28: 235-242.
[2] U. Hobohm; C. Sander. Protein Science (1994) 3: 522-524.
[3] A. Borrel; AC. Camproux; H. Xhaard. ACS omega (2017) 7359-7374.
[4] Y. Zhang; A. Borrel; L. Ghemtio; L. Regad; G. Boije Af Gennäs; AC. Camproux;J. Yli-Kauhaluoma; H. Xhaard . J Chem Inf Model (2017), 3:499-516.
[5] T. Inhester; S. Bietz; M Hilbig; R. Schmidt; M. Rarey. J. Chem. Inf. Model (2017) 2: 148–158