

Mutation Mining: Automated Extraction of Mutation Data from Scientific Publications.

Krzysztof Rataj¹, Jagna Witek¹, Stefan Mordalski¹, Kurt Kristiansen²,
Sabina Smusz¹, Andrzej J. Bojarski¹

¹ Department of Medicinal Chemistry, Institute of Pharmacology,
Polish Academy of Science, Smętna 12, Kraków, Poland

² Medical Pharmacology and Toxicology, Department of Medical Biology,
Faculty of Health Science, University of Tromsø, N-9037 Tromsø, Norway
e-mail: krzysztof_rataj@o2.pl

Manual extracting of mutation data from scientific publications is a slow and laborious process. Automated extraction is complicated, since mutations are described in different ways and need to be checked manually. Often, the publications contain data on mutagenesis of other proteins than the target one, rendering standard automated methods helpless. Most available software require manually created mutation databases¹ and do not in fact analyse given publication. Creation of error-free, internet-accessible mutation extracting software would become a major asset in gathering scientific data and database construction. Therefore, we came up with an idea for this task – the Mutation Miner. Mutation Miner software, collaborating with Uniprot² and HyperCLDB³ databases, is the solution for this task. Upon uploading your publication in PDF format, the program uses resourceful algorithms to find both mutations and the target protein, enabling cross-checking of acquired results. Even though still in its alpha-testing phase and few more options are yet to be implemented, Mutation Miner is a useful, time-saving tool.

[1] Horn F., Lau A.L., Cohen F.E.: *Bioinformatics* **20** (4) (2004), 557-68.

[2] Magrane M. and the UniProt consortium: **UniProt Knowledgebase: a hub of integrated protein data**. *Database* 2011, bar009.

[3] Romano P., Manniello A., Aresu O., Armento M., Cesaro M., Parodi B.: *Nucleic Acids Res.* **37** (Database issue) (2009), D925-D932. DOI: doi:10.1093/nar/gkn730; PMID: 18927105.