

PhotoProteinTech database on the properties of luciferases and Ca²⁺-regulated photoproteins, technologies and areas of their application created by automatic analysis of scientific publications and patents

Ivanisenko V.A.^{1, 2, 3, 4*}, Burakova L.P.⁴, Ereemeeva E.V.⁴, Natashin P.V.⁴,
Larionova M.D.⁴, Malikova N.P.¹, Ivanisenko T.V.^{1, 2, 3, 4}, Demenkov P.S.^{1, 2, 3},
Venzel A.S.^{1, 2, 3, 4}, Ivanisenko N.V.^{1, 4, 5}, Vysotski E.S.⁴

¹ Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

² Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

³ Novosibirsk State University, Novosibirsk, Russia

⁴ Institute of Biophysics, SB RAS, Federal Research Center “Krasnoyarsk Science Center of the Siberian Branch of the Russian Academy of Sciences”, Krasnoyarsk, Russia

⁵ AIRI, Moscow, Russia

* salix@bionet.nsc.ru

A relational database PhotoProteinTech has been developed, containing data on the properties of coelenterazine-dependent luciferases and Ca²⁺-regulated photoproteins (sequences, physicochemical properties, melting temperatures, emission spectra, and bioluminescent activity), as well as information on technologies and areas of their application.

The database is implemented using MySQL 5.7.26 DBMS and consists of interconnected tables: (i) “main_data” contains data on native luciferases and photoproteins; (ii) “mutants_data” – data on mutant forms of proteins from the “main_data” table; (iii) “homologs_data” – data on homologs of proteins from the “main_data” table; (iv) “func_site_data” – data on functional sites in the protein structure; (v) “ligand_data” – data on molecular ligands; (v) “taxonomy” – taxonomy; (vi) “dictionaries” – dictionaries of names and synonyms of proteins and ligands; (vii) “ReferenceCross” – references to literature; (viii) “References” – description of literature; (ix) “Patents” – description of patents.

To populate the database, an analysis of scientific publications and patents was carried out using natural language text mining methods implemented in our previously developed cognitive software and information system ANDSystem [1]. The current version of the database contains information about the light-sensitive Ca²⁺-regulated photoprotein berovin from *Beroe abyssicola* and its 78 mutants, Ca²⁺-regulated photoproteins aequorin and its 96 mutants and obelin and its 87 mutants from *Aequorea victoria* and *Obelia longissima*, respectively, as well as the coelenterazine-dependent Renilla luciferase and its 130 mutants.

Analysis of international patents made it possible to extract information about a wide range of technologies in which photoproteins and luciferases were used, including “BRET-based antibody sensors”, “luminescent antibody sensing (LUMABS) technology”, “bioluminescence resonance energy transfer assays (BRET)”, “NanoBRET technology”, “High-throughput screening”, “in vivo imaging” technologies, etc., as well as their biomedical applications, including areas such as drug resistance, cardiovascular diseases, tumors, and others.

In conclusion, PhotoProteinTech provides data on the physicochemical and structural properties of luciferases and photoproteins, as well as related information on technologies and areas of their practical application, which determines the novelty of the created database. The presented data can be used to design new forms of luciferases and photoproteins.

Funding: The work was supported by the grant of the Russian Science Foundation No. 22-14-00125 “Rational design of coelenterazine-dependent bioluminescent proteins with novel properties using mutagenesis and molecular modeling”.

References

1. Ivanisenko V.A., Demenkov P.S., Ivanisenko T.V., Mishchenko E.L., Saik O.V. A new version of the ANDSystem tool for automatic extraction of knowledge from scientific publications with expanded functionality for reconstruction of associative gene networks by considering tissue-specific gene expression. *BMC Bioinformatics*. 2019;20(Suppl. 1):H.34