

Łukasz P. Kozłowski











Email: lukasz.kozlowski.lpk@gmail.com

<http://bioinformatics.netmark.pl>
<http://www.solidscientometrics.com>

Research Interests

- computational biology (bioinformatics, structural biology, proteomics, intrinsic protein disorder, next-generation sequencing, single nucleotide polymorphism)
- big data & machine learning (deep learning, SVM, genetic algorithms, numeric optimization)
- bibliometrics, scientometrics

Projects





-  **Proteome-pI 2.0 - Proteome Isoelectric Point Database** (pI for over 61 million sequences from 20,115 proteomes calculated by 22 methods. Contribution: 100% Status: *operational since 2021*)
-  **Unicorn Papers** – a list with top $\frac{1}{10,000}$ (permyriad, "for every ten thousand") cited papers from PUBMED. Contribution: 100% Status: *operational since 2021*
-  **IPC 2.0** – prediction of isoelectric point and pKa dissociation constants using deep learning and SVR. Contribution: 100% Status: *operational since 2020*
-  **fCite – a fractional citation tool to quantify an individual's scientific research output** (divide the citations by the number of authors; the metrics are calculated based on 650 million citations and references, 34 million publications and 1.23 million ORCID profiles). Contribution: 100%. Status: *operational since 2018*
-  **IPC – Isoelectric Point Calculator** (cited >150 times in Web of Science, >200k visitors since 2016, [the stats](#)). Contribution: 100% Status: *operational since 2016*
-  **Proteome-pI - Proteome Isoelectric Point Database** (pI for over 21 million sequences from 5,029 proteomes calculated by 18 methods, cited >80 times in Web of Science). Contribution: 100% Status: *operational since 2016*
-  **MetaDisorder** – prediction of intrinsic disorder in proteins (in 2008-2010 the best program according [CASP8](#) defeating >20 programs for protein disorder prediction in a blind test, see [here](#), a group 153, the publication cited >200 times Web of Science). Contribution: ~80% Status: *operational since 2010*
- GeneSilico fold recognition server** – the development and the maintenance; in ten years I integrated there >100 bioinformatics tools (1.7TB for the programs, 2TB for biological databases; >40 GB MySQL database; over 3500 registered users). Contribution: ~50% Status: retired (*management from 2008 to 2017*)
-  **CompRNA** – benchmarking of RNA structure prediction methods. Contribution: ~15% Status: *operational since 2010*
-  **GDFuzz3D** – protein contact map to 3D structure retrieval service. Contribution: ~35% Status: *operational since 2015*
-  **mRNA3db** - Human pre-mRNA 3'-end processing machinery database. Contribution: ~80% Status: *operational since 2013*
-  **Shannon Entropy Calculator** (~100 visitors per day – [the stats](#)). Contribution: 100% Status: *operational since 2012*

Technical skills

Problem solving: analysis of big data using advanced statistics & machine learning (e.g. deep learning, support vector machines, genetic algorithms, numeric optimization); **Software development:** *programming languages:* PYTHON, JavaScript, PHP, R, C++, C, shell; *databases:* MySQL, PostgreSQL; **Scientific packages:** in-depth knowledge of bioinformatics programs (e.g. BLAST, HHSearch, ClustalW, UCSD Chimera, Modeller, MEME, CD-HIT, ABYSS, Ray, Velvet and few hundred others); **Languages:** Polish (native), English (CAE, fluent spoken and written), German (basic)

Educational Background





























Ph.D. (2013, *biochemistry, cum laude*), BSc (2007, *informatics*), MSc (2006, *genetics*), BSc (2004, *biology*)



- 2008 – 2013 **Ph.D. (biochemistry, bioinformatics, cum laude)** Institute of Biochemistry and Biophysics of the Polish Academy of Science (Warsaw), **Thesis title:** *Integrated bioinformatics platform for protein analysis. Prediction of protein domain and intrinsic protein disorder* 
- 2005 – 2007 **Bachelor's degree:** Jan Kochanowski University (Kielce), Institute of Physics, Division Computer Science, **Thesis title:** *Determining amino acid composition in proteins using genetic algorithm* 
- 2004 – 2006 **Master's degree:** Jan Kochanowski University (Kielce), Institute of Biology, Department of Biochemistry and Genetics, **Thesis title:** *Phylogenetic analysis of the linker histones in vertebrates* 
- 2001 – 2004 **Bachelor's degree:** Jan Kochanowski University (Kielce), Institute of Biology, Department of Biochemistry and Genetics, **Thesis title:** *The diversity of linker histones in vertebrates* 



Employment


- 2018 – Assistant Professor (University of Warsaw, Institute of Informatics), **Poland**
- 2018 – Chairman, CEO and Owner ([Solid Scientometrics Sp. z o.o.](#)), **Poland**
- 2015 – 2017 Postdoctoral Researcher, Max Planck Institute for Biophysical Chemistry, **Germany**
- 2013 – 2015 Postdoctoral Researcher, The International Institute of Molecular and Cell Biology, **Poland**
- 2010 – 2011 Self-employed trader (E-commerce), **Poland**
- 2008 – 2013 Ph.D. Student, The International Institute of Molecular and Cell Biology, **Poland**
- 2007 – 2008 Research assistant, Bioinformatician, The International Institute of Molecular and Cell Biology, **Poland**
- 2006 – 2007 Research assistant, Genetics, Jan Kochanowski University in Kielce, **Poland**


Publications & Preprints






-    Morzywolek, ..., **Kozłowski LP**, ..., et al. *Novel lytic enzyme of prophage origin from Clostridium botulinum E3 strain Alaska E43 with bactericidal activity against clostridial cells.* **Int. J. Mol. Sci** **2021**; doi: 10.3390/ijms22179536
-  Aevansson, ..., **Kozłowski LP**, ..., et al. *Going to extremes - a metagenomic journey into the dark matter of life.* **FEMS Microbiol. Lett** **2021**; doi: 10.1093/femsle/fnab067 (review)
-    **Kozłowski LP***  *IPC 2.0 – prediction of isoelectric point and pKa dissociation constants.* **Nucleic Acids Res.** **2021**, 49 (W1): W285–W292; doi: 10.1093/nar/gkab295
-   Erijman A, **Kozłowski L**, et al. *A High-Throughput Screen for Transcription Activation Domains Reveals Their Sequence Features and Permits Prediction by Deep Learning.* **Molecular Cell** **2020**; 78, 1–13
-   **Kozłowski LP***  *fCite: a fractional citation tool to quantify an individual's scientific research output.* **bioRxiv** 771485; 2019; doi: 10.1101/771485 (**preprint**)
-    Plotka M, Sancho-Vaello E, Dorawa S, Kaczorowska AK, **Kozłowski LP**, Kaczorowski T, Zeth K. *Structure and function of the Ts2631 endolysin of Thermus scotoductus phage vB_Tsc2631 with unique N-terminal extension used for peptidoglycan binding.* **Sci Rep.** **2019**, 9(1):1261
-    **Kozłowski LP***  *Proteome-pI - Proteome Isoelectric Point Database.* **Nucleic Acids Res.** **2017**, 45 (D1): D1112–D1116
-    **Kozłowski LP***  *IPC - Isoelectric point calculator.* **Biology Direct** **2016**, 11:55 **Highly accessed**  Highly Cited Paper
-  Pawłowski M, **Kozłowski L**, Kloczkowski A. *MQAPsingle: A quasi single-model approach for estimation of the quality of individual protein structure models.* **Proteins** **2016**, 84(8):1021–8
-    Plotka M, Kaczorowska AK, Morzywolek A, Makowska J, **Kozłowski LP**, et al. *Biochemical Characterization and Validation of a Catalytic Site of a Highly Thermostable Ts2631 Endolysin from the Thermus scotoductus Phage vB_Tsc2631.* **PLoS One** **2015**, 10(9):e0137374
-    Pietal MJ, Bujnicki JM, **Kozłowski LP** *GDFuzz3D: a method for protein 3D structure reconstruction from contact maps, based on a non-Euclidean distance function.* **Bioinformatics** **2015**, 31(21), 3499–3505
-    Głow D, Pianka D, Sulej AA, **Kozłowski LP**, et al. *Sequence-specific cleavage of dsRNA by Mini-III RNase.* **Nucleic Acids Res.** **2015**, 43(5):2864–73
-   Sierocka I, **Kozłowski LP**, Bujnicki JM, Jarmolowski A, Szweykowska-Kulinska Z *Female-specific gene expression in dioecious liverwort Pellia endiviifolia is developmentally regulated and connected to archegonia production.* **BMC Plant Biol.** **2014**, 14:168
-    Plotka M, Kaczorowska AK, Stefanska A, Morzywolek A, Fridjonsson O, Dunin-Horkawicz S, **Kozłowski L**, et al. *Discovery and characterization of a novel highly thermostable endolysin from Thermus scotoductus MAT2119 bacteriophage Ph2119 that shows amino-acid sequence similarity to eukaryotic peptidoglycan recognition proteins (PGRPs).* **Appl Env Microbiol** **2013** 80(3):886–95
-  Tuszynska I, Matelska D, Magnus M, Chojnowski G, Kasprzak JM, **Kozłowski LP**, et al. *Computational modeling of protein-RNA complex structures.* **Methods** **2013**, 65(3):310–9 (review)
-    Puton T, **Kozłowski L**, Rother KM, Bujnicki JM *CompaRNA: a server for continuous benchmarking of automated methods for RNA secondary structure prediction.* **Nucleic Acids Res.** **2013**, 41(7):4307–23
-    **Kozłowski LP**, Bujnicki JM *MetaDisorder: a meta-server for the prediction of intrinsic disorder in proteins.* **BMC Bioinformatics** **2012**, 13:111 **Highly accessed**  Highly Cited Paper
-   Al-Haggar M, Madej-Pilarczyk A, **Kozłowski L**, et al. *A novel homozygous p.Arg527Leu LMNA mutation in two unrelated Egyptian families causes overlapping mandibuloacral dysplasia and progeria syndrome.* **Eur J Human Genet** **2012**, 20 (11),

1134-1140
  Puton T, **Kozłowski L***, Tuszyńska I, Rother K, Bujnicki JM *Computational methods for prediction of protein-RNA interactions*. **J Struct Biol** **2012**, 179 (3), 261-268 (review) **Highly accessed**

  Nakagome S, Mano S, **Kozłowski L**, et al. *Crohn's disease risk alleles on the NOD2 locus have been maintained by natural selection on standing variation*. **Mol Biol Evol.** **2012** Jun;29(6):1569-85

 **Kozłowski L**, Orłowski J, Bujnicki JM "Structure prediction of alternatively spliced proteins" in "Alternative pre-mRNA Splicing: Theory and Protocols" Editors: Stamm S, Smith Ch, Luehrmann R, Wiley-Blackwell, **2012**, 582-591. doi: 10.1002/9783527636778 (book chapter)


 Majorek K, **Kozłowski L**, Jakalski M, Bujnicki JM "First steps of protein structure prediction" in "Prediction of Protein Structures, Functions and Interactions" Editors: Bujnicki JM; Wiley&Sons, **2008**, 2, 39–62. doi: 10.1002/9780470741894.ch2 (book chapter)

 the main paper * single authorship publications  corresponding author
 supplementary data & shared first authorship  Highly Cited Paper (Top 1% in WoS)
 Free / Open Access (OA) article

Scientometric data



Other

- Referee in Journals**  Nucleic Acid Res. (21), Bioinformatics (16), International Journal of Molecular Sciences (18), Knowledge-Based Systems (5), Food Technology and Biotechnology (4), Scientometrics (4), PeerJ (2), J Prot Res., Scientific Rep., Viruses, and few others (>80 reviews in total)
- Membership in Societies** International Society for Computational Biology (ISCB, 2009-)
 Society for Bioinformatics in Northern Europe (SocBiN, 2017-)
 Polish Bioinformatics Society (PTBI, 2008-)
 Polish Biochemical Society (PTBioch, 2015-)

Member of an expert panel in the National Center of Research and Development (NCRD, pol. *NCBiR*), Poland (the evaluation of R&D projects with total valuation over \$100M). Occasionally, the assessment for other institutions such as Polish National Agency for Academic Exchange (NAWA), Polish Agency for Enterprise Development (PARP), Register of Expert Peer Reviewers for Italian Scientific Evaluation (REPRISE), Fonds de la Recherche Scientifique (F.R.S.–FNRS).



Conferences

- 2021 **MLCB 2021** (Machine Learning in Computational Biology), **HUPO ReCONNECT 2021** (Human Proteome Organization Congress), **ML in PL 2021** (Machine Learning in Poland), **STI2021** (Conference on Science, Technology and Innovation Indicators), **RECOMB** (poster), **ISMB/ECCB** (poster & talk), **SCFE21, ICCS** (International Conference on Computational Science), **AI in Healthcare**
- 2020 **ECCB, ISMB** (poster & talk), **RECOMB, SCFE20** (Supercomputer Frontiers 2020)
- 2019 **ISMB/ECCB** (poster & talk)
- 2017 **ISMB/ECCB** (poster)
- 2016 **13th Horizons in Molecular Biology Symposium, FRET 2** (International Discussion Meeting on Förster Resonance Energy Transfer), **DAGStat2016** (4th Joint Statistical Meeting)
- 2015 **RECOMB** (19th Annual International Conference on Research in Computational Molecular Biology; poster),
- 2014 **BIO 2014** (1st Congress of the Polish Biochemistry, Cell Biology, Biophysics and Bioinformatics; poster), **ECCB** (poster)
- 2013 **RNA 3' ends: Mechanism and biological function in eukaryotic genomes** (poster), **SocBiN Bioinformatics/BIT13** (poster), **Cancer genetics for medical community** (poster)
- 2012 **CASP10** (poster), **BIT12** (lecture & poster), **PTBI12** (lecture & poster)
- 2011 **MultiPole Approach to Structural Biology** (poster), **PTBI11** (lecture & poster), **BIT11** (lecture & poster)
- 2010 **CASP9** (poster), **EURASNET** (Frontiers in Structural Biology of RNAs and RNPs; poster), **BIT10** (lecture & poster)

- 2009 **PTBI09** (lecture & poster), **ISMB/ECCB** (Intelligent Systems for Molecular Biology/ European Conference on Computational Biology; poster), **BIT09** (Bioinformatics in Toruń; lecture & poster)
- 2008 **CASP8** (Critical Assessment of Techniques for Protein Structure Prediction; [poster](#) & invited lecture), **PTBI08** (1st Annual Meeting of Polish Bioinformatics Society; lecture & poster), **SocBin Bioinformatics** (lecture & [poster](#))

Workshops & Training

- 2017 Central Anti-Corruption Bureau's (CBA) training (three modules: [Corruption in public administration](#), [Corruption in bussines](#) and [Social consequences of corruption](#))
- 2014 Advanced RNA-Seq and ChIP-Seq Data Analysis Course
- 2013 Automated and Reproducible Analysis of NGS data (ARANGS13) Course
- 2011 EMBO Young Scientists Forum
- 2010 Workshop on Interdisciplinary Grant Application; HealthProt workshop; Grid Computing and Bioinformatics (PL-GRID); Workshop on Scientific Communication; EMBO workshop: RNA and RNA-processing enzymes, School of Molecular Medicine; Cambridge PYTHON's Innovation Academy

Grants

- 2019-2021 National Science Centre (PL, OPUS 2018/29/B/NZ2/01403) – **principal investigator** (~150k Euro) [🔗](#)
- 2016-2020 Virus-X: Viral Metagenomics for Innovation Value (EU, Horizon2020, grant no. 685778) – contractor [🔗](#)
- 2013-2014 Polish Ministry of Science and Higher Education (PL, IUVENTUS PLUS, 0301/IP1/2013/72) – **principal investigator** (~35k Euro)
- 2012-2014 Exgenome Molecular Enzymes (EU, FP7, grant no. 286556) – contractor [🔗](#)
- 2011-2015 European Research Council (EU, ERC StG Grant RNA+P=123, grant no. 26135) – contractor [🔗](#)
- 2009-2012 Health-Prot: Proteins in Health and Disease (EU, FP7, grant no. 229676) – contractor [🔗](#)
- 2008-2013 Polish Ministry of Science and Higher Education (PL, grant no. NN301 190139) – **the main contractor**, so called “promotor’s grant” funding PhD training (~10k Euro)

Fellowships

- 2015-2017 Max Planck Institute Fellowship
- 2008 Masovian Fellowship for Ph.D. Students

Teaching

- 2020/2021 Data analysis and visualization (*Lecture & Labs [x3]*), Architecture of large projects in bioinformatics (*Lecture & Labs [x1]*)
- 2019/2020 Data analysis and visualization (*Lecture & Labs [x2]*)
- 2018/2019 Web applications (*Labs [x1]*), Probability theory and statistics (*Classes [x1] & Labs [x1]*), Statistical data analysis 2 (*Labs [x1]*), Introduction to computer science (*Labs [x1]*)

Classes - whiteboard exercises (old school way); Labs – in front of the computer; [x1], [x2], ... – number of groups;
All teaching activities conducted at University of Warsaw

Varia

- 2016 organization of the tutorial on protein modeling, within Bioinformatics Course for master students in Gottingen University, Gottingen, **Germany**, Nov 1-2 ([Part 1](#), [Part 2](#))
- 2011 co-organization of “MultiPole Approach to Structural Biology”, Warsaw, **Poland**, Nov 16-19 ([badge](#))

In free time active Wikipedian



Online version of CV available at: <http://goo.gl/JgTHzC>

Disclaimer: All information presented in this document are meant to be fully verifiable. Whenever possible there are hyperlinks directing to the source material confirming given information (>100 hyperlinks). All publications are available for download (note that some of them – especially the old one, currently I publish only in OA journals – are not open access, and in the given case, they are for personal use only). Similarly, you can browse the posters and some of the presentations from the past conferences. If there is a piece of information about some of my experience or skill you should be able to easily track it by the analysis of the relevant materials e.g. publications, theses, or/and web pages that I have done in the past.