Robert D. Finn

List of Publications by Citations

203 papers	72,840 citations	77 h-index	226 g-index
226 ext. papers	94,098 ext. citations	13.8 avg, IF	7.77 L-index

#	Paper	IF	Citations
203	Initial sequencing and analysis of the human genome. *Nature*, **2001**, 409, 860-921	50.4	17366
202	Pfam: the protein families database. *Nucleic Acids Research*, **2014**, 42, D222-30	20.1	3975
201	The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Research*, **2016**, 44, D279-85	20.1	3678
200	HMMER web server: interactive sequence similarity searching. *Nucleic Acids Research*, **2011**, 39, W29-37	20.1	2917
199	The Pfam protein families database. *Nucleic Acids Research*, **2012**, 40, D290-301	20.1	2844
198	The Pfam protein families database. *Nucleic Acids Research*, **2004**, 32, D138-41	20.1	2720
197	The Pfam protein families database. *Nucleic Acids Research*, **2010**, 38, D211-22	20.1	2467
196	The Pfam protein families database in 2019. *Nucleic Acids Research*, **2019**, 47, D427-D432	20.1	2298
195	The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research*, **2019**, 47, W636-W641	20.1	2132
194	The Pfam protein families database. *Nucleic Acids Research*, **2002**, 30, 276-80	20.1	1839
193	The Pfam protein families database. *Nucleic Acids Research*, **2008**, 36, D281-8	20.1	1806
192	Pfam: clans, web tools and services. *Nucleic Acids Research*, **2006**, 34, D247-51	20.1	1784
191	InterPro: the integrative protein signature database. *Nucleic Acids Research*, **2009**, 37, D211-5	20.1	1379
190	The Pfam protein families database. *Nucleic Acids Research*, **2000**, 28, 263-6	20.1	1074
189	UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Research*, **2021**, 49, D480-D489	20.1	1073
188	Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Research*, **2005**, 33, D121-4	20.1	990
187	InterPro in 2017-beyond protein family and domain annotations. *Nucleic Acids Research*, **2017**, 45, D190-D199	20.1	970

186	Rfam: an RNA family database. *Nucleic Acids Research*, **2003**, 31, 439-41	20.1	970
185	The InterPro protein families database: the classification resource after 15 years. *Nucleic Acids Research*, **2015**, 43, D213-21	20.1	954
184	InterPro in 2011: new developments in the family and domain prediction database. *Nucleic Acids Research*, **2012**, 40, D306-12	20.1	844
183	InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Research*, **2019**, 47, D351-D360	20.1	835
182	Rfam 12.0: updates to the RNA families database. *Nucleic Acids Research*, **2015**, 43, D130-7	20.1	723
181	MEROPS: the peptidase database. *Nucleic Acids Research*, **2010**, 38, D227-33	20.1	670
180	Rfam: updates to the RNA families database. *Nucleic Acids Research*, **2009**, 37, D136-40	20.1	652
179	The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Research*, **2018**, 46, D624-D632	20.1	643
178	Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Research*, **2013**, 41, e121	20.1	629
177	Pfam: The protein families database in 2021. *Nucleic Acids Research*, **2021**, 49, D412-D419	20.1	620
176	MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Research*, **2014**, 42, D503-9	20.1	602
175	HMMER web server: 2018 update. *Nucleic Acids Research*, **2018**, 46, W200-W204	20.1	596
174	Rfam 11.0: 10 years of RNA families. *Nucleic Acids Research*, **2013**, 41, D226-32	20.1	596
173	HMMER web server: 2015 update. *Nucleic Acids Research*, **2015**, 43, W30-8	20.1	590
172	Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. *Nucleic Acids Research*, **2018**, 46, D335-D342	20.1	533
171	Twenty years of the MEROPS database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Research*, **2016**, 44, D343-50	20.1	489
170	A new genomic blueprint of the human gut microbiota. *Nature*, **2019**, 568, 499-504	50.4	484
169	Single-particle electron cryo-microscopy: towards atomic resolution. *Quarterly Reviews of Biophysics*, **2000**, 33, 307-69	7	481

168   InterPro, progress and status in 2005. *Nucleic Acids Research*, **2005**, 33, D201-5   20.1   426

167   Highly accurate protein structure prediction for the human proteome. *Nature*, **2021**, 596, 590-596   50.4   399

166   New developments in the InterPro database. *Nucleic Acids Research*, **2007**, 35, D224-8   20.1   397

165   The InterPro protein families and domains database: 20 years on. *Nucleic Acids Research*, **2021**, 49, D344-D354   20.1   358

164   Ensembl Genomes 2018: an integrated omics infrastructure for non-vertebrate species. *Nucleic Acids Research*, **2018**, 46, D802-D808   20.1   330

163   Rfam: Wikipedia, clans and the "decimal" release. *Nucleic Acids Research*, **2011**, 39, D141-5   20.1   321

162   The Dfam database of repetitive DNA families. *Nucleic Acids Research*, **2016**, 44, D81-9   20.1   295

161   iPfam: visualization of protein-protein interactions in PDB at domain and amino acid resolutions. *Bioinformatics*, **2005**, 21, 410-2   7.2   263

160   A human gut bacterial genome and culture collection for improved metagenomic analyses. *Nature Biotechnology*, **2019**, 37, 186-192   44.5   224

159   QuickTree: building huge Neighbour-Joining trees of protein sequences. *Bioinformatics*, **2002**, 18, 1546-7   7.2   214

158   Skylign: a tool for creating informative, interactive logos representing sequence alignments and profile hidden Markov models. *BMC Bioinformatics*, **2014**, 15, 7   3.6   196

157   The systematic functional analysis of Plasmodium protein kinases identifies essential regulators of mosquito transmission. *Cell Host and Microbe*, **2010**, 8, 377-87   23.4   195

156   The PASTA domain: a beta-lactam-binding domain. *Trends in Biochemical Sciences*, **2002**, 27, 438   10.3   193

155   A unified catalog of 204,938 reference genomes from the human gut microbiome. *Nature Biotechnology*, **2021**, 39, 105-114   44.5   185

154   The Human RNA-Binding Proteome and Its Dynamics during Translational Arrest. *Cell*, **2019**, 176, 391-403.e19   36.2   157

153   Non-Coding RNA Analysis Using the Rfam Database. *Current Protocols in Bioinformatics*, **2018**, 62, e51   24.2   152

152   Dfam: a database of repetitive DNA based on profile hidden Markov models. *Nucleic Acids Research*, **2013**, 41, D70-82   20.1   147

151   DUFs: families in search of function. *Acta Crystallographica Section F: Structural Biology Communications*, **2010**, 66, 1148-52   145

(2003-2004)

150	The Hotdog fold: wrapping up a superfamily of thioesterases and dehydratases. *BMC Bioinformatics*, **2004**, 5, 109	3.6	137
149	Microbial abundance, activity and population genomic profiling with mOTUs2. *Nature Communications*, **2019**, 10, 1014	17.4	134
148	Alternative splicing of intrinsically disordered regions and rewiring of protein interactions. *Current Opinion in Structural Biology*, **2013**, 23, 443-50	8.1	134
147	EBI Metagenomics in 2017: enriching the analysis of microbial communities, from sequence reads to assemblies. *Nucleic Acids Research*, **2018**, 46, D726-D735	20.1	130
146	MGnify: the microbiome analysis resource in 2020. *Nucleic Acids Research*, **2020**, 48, D570-D578	20.1	127
145	Predicting active site residue annotations in the Pfam database. *BMC Bioinformatics*, **2007**, 8, 298	3.6	124
144	iPfam: a database of protein family and domain interactions found in the Protein Data Bank. *Nucleic Acids Research*, **2014**, 42, D364-73	20.1	120
143	RNAcentral: a comprehensive database of non-coding RNA sequences. *Nucleic Acids Research*, **2017**, 45, D128-D134	20.1	119
142	Rfam 14: expanded coverage of metagenomic, viral and microRNA families. *Nucleic Acids Research*, **2021**, 49, D192-D200	20.1	108
141	Pfam 10 years on: 10,000 families and still growing. *Briefings in Bioinformatics*, **2008**, 9, 210-9	13.4	101
140	TreeFam v9: a new website, more species and orthology-on-the-fly. *Nucleic Acids Research*, **2014**, 42, D922-5	20.1	96
139	Cache Domains That are Homologous to, but Different from PAS Domains Comprise the Largest Superfamily of Extracellular Sensors in Prokaryotes. *PLoS Computational Biology*, **2016**, 12, e1004862	5	95
138	Protein interactions in human genetic diseases. *Genome Biology*, **2008**, 9, R9	18.3	94
137	Massive expansion of human gut bacteriophage diversity. *Cell*, **2021**, 184, 1098-1109.e9	56.2	93
136	The European Bioinformatics Institute in 2016: Data growth and integration. *Nucleic Acids Research*, **2016**, 44, D20-6	20.1	91
135	RNAcentral: a hub of information for non-coding RNA sequences. *Nucleic Acids Research*, **2019**, 47, D221-D229	20.1	90
134	RNAcentral: an international database of ncRNA sequences. *Nucleic Acids Research*, **2015**, 43, D123-9	20.1	89
133	New knowledge from old: in silico discovery of novel protein domains in Streptomyces coelicolor. *BMC Microbiology*, **2003**, 3, 3	4.5	89

132. Tandem repeats lead to sequence assembly errors and impose multi-level challenges for genome and protein databases. *Nucleic Acids Research*, **2019**, 47, 10994-11006 — 20.1 — 86

131. A comparison of dense transposon insertion libraries in the Salmonella serovars Typhi and Typhimurium. *Nucleic Acids Research*, **2013**, 41, 4549-64 — 20.1 — 86

130. The evolution of protein domain families. *Biochemical Society Transactions*, **2009**, 37, 751-5 — 5.1 — 81

129. EBI metagenomics in 2016--an expanding and evolving resource for the analysis and archiving of metagenomic data. *Nucleic Acids Research*, **2016**, 44, D595-603 — 20.1 — 81

128. Quantifying the mechanisms of domain gain in animal proteins. *Genome Biology*, **2010**, 11, R74 — 18.3 — 78

127. Integrating biological data--the Distributed Annotation System. *BMC Bioinformatics*, **2008**, 9 Suppl 8, S3 — 3.6 — 77

126. Structure and distribution of pentapeptide repeats in bacteria. *Protein Science*, **1998**, 7, 1477-80 — 6.3 — 76

125. The G5 domain: a potential N-acetylglucosamine recognition domain involved in biofilm formation. *Bioinformatics*, **2005**, 21, 1301-3 — 7.2 — 68

124. Representative proteomes: a stable, scalable and unbiased proteome set for sequence analysis and functional annotation. *PLoS ONE*, **2011**, 6, e18910 — 3.7 — 67

123. Asparagine peptide lyases: a seventh catalytic type of proteolytic enzymes. *Journal of Biological Chemistry*, **2011**, 286, 38321-38328 — 5.4 — 65

122. Benchmarking taxonomic assignments based on 16S rRNA gene profiling of the microbiota from commonly sampled environments. *GigaScience*, **2018**, 7, — 7.6 — 61

121. On expert curation and scalability: UniProtKB/Swiss-Prot as a case study. *Bioinformatics*, **2017**, 33, 3454-3460 — 7.4 — 60

120. Integrating sequence and structural biology with DAS. *BMC Bioinformatics*, **2007**, 8, 333 — 3.6 — 57

119. Members of the immunoglobulin superfamily in bacteria. *Protein Science*, **1996**, 5, 1939-41 — 6.3 — 57

118. Computational strategies to combat COVID-19: useful tools to accelerate SARS-CoV-2 and coronavirus research. *Briefings in Bioinformatics*, **2021**, 22, 642-663 — 13.4 — 56

117. The PepSY domain: a regulator of peptidase activity in the microbial environment?. *Trends in Biochemical Sciences*, **2004**, 29, 169-72 — 10.3 — 55

116. RNAcentral: A vision for an international database of RNA sequences. *Rna*, **2011**, 17, 1941-6 — 5.8 — 54

115. HPMCD: the database of human microbial communities from metagenomic datasets and microbial reference genomes. *Nucleic Acids Research*, **2016**, 44, D604-9 — 20.1 — 50

114	RNIE: genome-wide prediction of bacterial intrinsic terminators. *Nucleic Acids Research*, **2011**, 39, 5845-52	20.1	50
113	The RNA WikiProject: community annotation of RNA families. *Rna*, **2008**, 14, 2462-4	5.8	50
112	The HMMER Web Server for Protein Sequence Similarity Search. *Current Protocols in Bioinformatics*, **2017**, 60, 3.15.1-3.15.23	24.2	49
111	Biocurators and biocuration: surveying the 21st century challenges. *Database: the Journal of Biological Databases and Curation*, **2012**, 2012, bar059	5	49
110	Clustered coding variants in the glutamate receptor complexes of individuals with schizophrenia and bipolar disorder. *PLoS ONE*, **2011**, 6, e19011	3.7	48
109	Phospholipid scramblases and Tubby-like proteins belong to a new superfamily of membrane tethered transcription factors. *Bioinformatics*, **2009**, 25, 159-62	7.2	45
108	Identifying protein domains with the Pfam database. *Current Protocols in Bioinformatics*, **2008**, Chapter 2, Unit 2.5	24.2	43
107	Enhanced protein domain discovery by using language modeling techniques from speech recognition. *Proceedings of the National Academy of Sciences of the United States of America*, **2003**, 100, 4516-20	11.5	42
106	Filling out the structural map of the NTF2-like superfamily. *BMC Bioinformatics*, **2013**, 14, 327	3.6	40
105	Reuse of structural domain-domain interactions in protein networks. *BMC Bioinformatics*, **2007**, 8, 259	3.6	40
104	Structure of the ProQ RNA-binding protein. *Rna*, **2017**, 23, 696-711	5.8	38
103	Distant homology recognition using structural classification of proteins **1997**, 29, 105-112		37
102	SCOOP: a simple method for identification of novel protein superfamily relationships. *Bioinformatics*, **2007**, 23, 809-14	7.2	36
101	Visualizing profile-profile alignment: pairwise HMM logos. *Bioinformatics*, **2005**, 21, 2912-3	7.2	34
100	Pfam: a domain-centric method for analyzing proteins and proteomes. *Methods in Molecular Biology*, **2007**, 396, 43-58	1.4	34
99	Using the MEROPS Database for Proteolytic Enzymes and Their Inhibitors and Substrates. *Current Protocols in Bioinformatics*, **2014**, 48, 1.25.1-33	24.2	33
98	Eros is a novel transmembrane protein that controls the phagocyte respiratory burst and is essential for innate immunity. *Journal of Experimental Medicine*, **2017**, 214, 1111-1128	16.6	32
97	ProServer: a simple, extensible Perl DAS server. *Bioinformatics*, **2007**, 23, 1568-70	7.2	31

96	COVID-19 pandemic reveals the peril of ignoring metadata standards. *Scientific Data*, **2020**, 7, 188	8.2	30
95	Modifier effects between regulatory and protein-coding variation. *PLoS Genetics*, **2008**, 4, e1000244	6	30
94	Escherichia coli RNA polymerase core and holoenzyme structures. *EMBO Journal*, **2000**, 19, 6833-44	13	30
93	The ELIXIR Core Data Resources: fundamental infrastructure for the life sciences. *Bioinformatics*, **2020**, 36, 2636-2642	7.2	29
92	The metagenomic data life-cycle: standards and best practices. *GigaScience*, **2017**, 6, 1-11	7.6	29
91	Microbial composition of Kombucha determined using amplicon sequencing and shotgun metagenomics. *Journal of Food Science*, **2020**, 85, 455-464	3.4	28
90	Toward unrestricted use of public genomic data. *Science*, **2019**, 363, 350-352	33.3	25
89	Bacterial pleckstrin homology domains: a prokaryotic origin for the PH domain. *Journal of Molecular Biology*, **2010**, 396, 31-46	6.5	25
88	Cryo-EM structures of human RNA polymerase III in its unbound and transcribing states. *Nature Structural and Molecular Biology*, **2021**, 28, 210-219	17.6	25
87	A member of the immunoglobulin superfamily in bacteriophage T4. *Virus Genes*, **1997**, 14, 163-5	2.3	24
86	Novel protein domains and motifs in the marine planctomycete Rhodopirellula baltica. *FEMS Microbiology Letters*, **2004**, 236, 333-340	2.9	24
85	Making your database available through Wikipedia: the pros and cons. *Nucleic Acids Research*, **2012**, 40, D9-12	20.1	23
84	UniProt-DAAC: domain architecture alignment and classification, a new method for automatic functional annotation in UniProtKB. *Bioinformatics*, **2016**, 32, 2264-71	7.2	22
83	Experience using web services for biological sequence analysis. *Briefings in Bioinformatics*, **2008**, 9, 493-505	9.4	22
82	Identifying accurate metagenome and amplicon software via a meta-analysis of sequence to taxonomy benchmarking studies. *PeerJ*, **2019**, 7, e6160	3.1	22
81	Phylogenomics of expanding uncultured environmental Tenericutes provides insights into their pathogenicity and evolutionary relationship with Bacilli. *BMC Genomics*, **2020**, 21, 408	4.5	19
80	The yeast noncoding RNA interaction network. *Rna*, **2017**, 23, 1479-1492	5.8	19
79	Reporting guidelines for human microbiome research: the STORMS checklist. *Nature Medicine*, **2021**, 27, 1885-1892	50.5	19

| 78 | Estimating the quality of eukaryotic genomes recovered from metagenomic analysis with EukCC. *Genome Biology*, **2020**, 21, 244 | 18.3 | 19 |

| 77 | AntiFam: a tool to help identify spurious ORFs in protein annotation. *Database: the Journal of Biological Databases and Curation*, **2012**, 2012, bas003 | 5 | 18 |

| 76 | Outline structures for the extracellular domains of the fibroblast growth factor receptors. *Nature Structural and Molecular Biology*, **1995**, 2, 1068-74 | 17.6 | 17 |

| 75 | Domain atrophy creates rare cases of functional partial protein domains. *Genome Biology*, **2015**, 16, 88 | 18.3 | 16 |

| 74 | GO annotation in InterPro: why stability does not indicate accuracy in a sea of changing annotations. *Database: the Journal of Biological Databases and Curation*, **2016**, 2016, | 5 | 16 |

| 73 | Tandem domain swapping: determinants of multidomain protein misfolding. *Current Opinion in Structural Biology*, **2019**, 58, 97-104 | 8.1 | 16 |

| 72 | The challenge of increasing Pfam coverage of the human proteome. *Database: the Journal of Biological Databases and Curation*, **2013**, 2013, bat023 | 5 | 16 |

| 71 | R2DT is a framework for predicting and visualising RNA secondary structure using templates. *Nature Communications*, **2021**, 12, 3494 | 17.4 | 16 |

| 70 | Microbial community drivers of PK/NRP gene diversity in selected global soils. *Microbiome*, **2019**, 7, 78 | 16.6 | 15 |

| 69 | Genome properties in 2019: a new companion database to InterPro for the inference of complete functional attributes. *Nucleic Acids Research*, **2019**, 47, D564-D572 | 20.1 | 15 |

| 68 | Microbiota Characterization of Agricultural Green Waste-Based Suppressive Composts Using Omics and Classic Approaches. *Agriculture (Switzerland)*, **2020**, 10, 61 | 3 | 14 |

| 67 | DASMI: exchanging, annotating and assessing molecular interaction data. *Bioinformatics*, **2009**, 25, 1321-2 | 7.2 | 14 |

| 66 | Enhanced protein domain discovery using taxonomy. *BMC Bioinformatics*, **2004**, 5, 56 | 3.6 | 14 |

| 65 | A unified sequence catalogue of over 280,000 genomes obtained from the human gut microbiome | | 14 |

| 64 | How to use the MEROPS database and website to help understand peptidase specificity. *Protein Science*, **2021**, 30, 83-92 | 6.3 | 14 |

| 63 | Origins of peptidases. *Biochimie*, **2019**, 166, 4-18 | 4.6 | 13 |

| 62 | Identifying protein domains with the Pfam database. *Current Protocols in Bioinformatics*, **2003**, Chapter 2, Unit 2.5 | 24.2 | 13 |

| 61 | Patterns of database citation in articles and patents indicate long-term scientific and industry value of biological data resources. *F1000Research*, **2016**, 5, | 3.6 | 13 |

60	Using deep learning to annotate the protein universe.. *Nature Biotechnology*, **2022**,	44.5	13
59	The crystal structure of a bacterial Sufu-like protein defines a novel group of bacterial proteins that are similar to the N-terminal domain of human Sufu. *Protein Science*, **2010**, 19, 2131-40	6.3	12
58	Ensembl Genomes 2022: an expanding genome resource for non-vertebrates. *Nucleic Acids Research*, **2021**,	20.1	12
57	TreeGrafter: phylogenetic tree-based annotation of proteins with Gene Ontology terms and other annotations. *Bioinformatics*, **2019**, 35, 518-520	7.2	12
56	ELIXIR-EXCELERATE: establishing Europe's data infrastructure for the life science research of the future. *EMBO Journal*, **2021**, 40, e107409	13	11
55	The structure of Jann_2411 (DUF1470) from Jannaschia sp. at 1.45 Å resolution reveals a new fold (the ABATE domain) and suggests its possible role as a transcription regulator. *Acta Crystallographica Section F: Structural Biology Communications*, **2010**, 66, 1198-204		9
54	Pfam: the protein families database **2005**,		9
53	A catalogue of 1,167 genomes from the human gut archaeome.. *Nature Microbiology*, **2021**,	26.6	9
52	Unifying the known and unknown microbial coding sequence space		9
51	Acetylation of Surface Carbohydrates in Bacterial Pathogens Requires Coordinated Action of a Two-Domain Membrane-Bound Acyltransferase. *MBio*, **2020**, 11,	7.8	9
50	Predicted Input of Uncultured Fungal Symbionts to a Lichen Symbiosis from Metagenome-Assembled Genomes. *Genome Biology and Evolution*, **2021**, 13,	3.9	8
49	Genome3D: integrating a collaborative data pipeline to expand the depth and breadth of consensus protein structure annotation. *Nucleic Acids Research*, **2020**, 48, D314-D319	20.1	7
48	Solution structure of human myeloid-derived growth factor suggests a conserved function in the endoplasmic reticulum. *Nature Communications*, **2019**, 10, 5612	17.4	7
47	Eleven quick tips to build a usable REST API for life sciences. *PLoS Computational Biology*, **2018**, 14, e1006542	5	7
46	LUD, a new protein domain associated with lactate utilization. *BMC Bioinformatics*, **2013**, 14, 341	3.6	6
45	The structure of pyogenecin immunity protein, a novel bacteriocin-like immunity protein from Streptococcus pyogenes. *BMC Structural Biology*, **2009**, 9, 75	2.7	6
44	The structure of BVU2987 from Bacteroides vulgatus reveals a superfamily of bacterial periplasmic proteins with possible inhibitory function. *Acta Crystallographica Section F: Structural Biology Communications*, **2010**, 66, 1265-73		6
43	ELIXIR pilot action: Marine metagenomics - towards a domain specific set of sustainable services. *F1000Research*, **2017**, 6,	3.6	6

42	Gene Unprediction with Spurio: A tool to identify spurious protein sequences. *F1000Research*, **2018**, 7, 261	3.6	6
41	Defining the remarkable structural malleability of a bacterial surface protein Rib domain implicated in infection. *Proceedings of the National Academy of Sciences of the United States of America*, **2019**,	11.5	6
40	Integrating cultivation and metagenomics for a multi-kingdom view of skin microbiome diversity and functions.. *Nature Microbiology*, **2021**,	26.6	6
39	The complexity, challenges and benefits of comparing two transporter classification systems in TCDB and Pfam. *Briefings in Bioinformatics*, **2015**, 16, 865-72	13.4	5
38	Creating a specialist protein resource network: a meeting report for the protein bioinformatics and community resources retreat. *Database: the Journal of Biological Databases and Curation*, **2015**, 2015, bav063	5	5
37	Conformational changes of Escherichia coli sigma54-RNA-polymerase upon closed-promoter complex formation. *Journal of Molecular Biology*, **2005**, 354, 201-5	6.5	4
36	Determination of Escherichia coli RNA polymerase structure by single particle cryoelectron microscopy. *Methods in Enzymology*, **2003**, 370, 24-42	1.7	4
35	Sequence analysis of tyrosine recombinases allows annotation of mobile genetic elements in prokaryotic genomes. *Molecular Systems Biology*, **2021**, 17, e9880	12.2	4
34	TADOSS: computational estimation of tandem domain swap stability. *Bioinformatics*, **2019**, 35, 2507-2509	8.2	4
33	Unifying the known and unknown microbial coding sequence space.. *ELife*, **2022**, 11,	8.9	4
32	Two Pfam protein families characterized by a crystal structure of protein lpg2210 from Legionella pneumophila. *BMC Bioinformatics*, **2013**, 14, 265	3.6	3
31	The YARHG domain: an extracellular domain in search of a function. *PLoS ONE*, **2012**, 7, e35575	3.7	3
30	ISCB computational biology Wikipedia competition. *PLoS Computational Biology*, **2013**, 9, e1003242	5	3
29	The C-terminal 12 amino acids of sigma(N) are required for structure and function. *Archives of Biochemistry and Biophysics*, **1999**, 371, 234-40	4.1	3
28	The SHOCT domain: a widespread domain under-represented in model organisms. *PLoS ONE*, **2013**, 8, e57848	3.7	3
27	Estimating the quality of eukaryotic genomes recovered from metagenomic analysis		3
26	R2DT: computational framework for template-based RNA secondary structure visualisation across non-coding RNA types		3
25	Exploring Non-Coding RNAs in RNAcentral. *Current Protocols in Bioinformatics*, **2020**, 71, e104	24.2	3

24	An inter-laboratory study to investigate the impact of the bioinformatics component on microbiome analysis using mock communities. *Scientific Reports*, **2021**, 11, 10590	4.9	3
23	The Ensembl COVID-19 resource: ongoing integration of public SARS-CoV-2 data. *Nucleic Acids Research*, **2021**,	20.1	3
22	3DPatch: fast 3D structure visualization with residue conservation. *Bioinformatics*, **2019**, 35, 332-334	7.2	2
21	The first structure in a family of peptidase inhibitors reveals an unusual Ig-like fold. *F1000Research*, **2013**, 2, 154	3.6	2
20	The European Bioinformatics Institute (EMBL-EBI) in 2021. *Nucleic Acids Research*, **2021**,	20.1	2
19	Cryo-EM structures of human RNA polymerase III in its unbound and transcribing states		2
18	Recovering prokaryotic genomes from host-associated, short-read shotgun metagenomic sequencing data. *Nature Protocols*, **2021**, 16, 2520-2541	18.8	2
17	Structure and computational analysis of a novel protein with metallopeptidase-like and circularly permuted winged-helix-turn-helix domains reveals a possible role in modified polysaccharide biosynthesis. *BMC Bioinformatics*, **2014**, 15, 75	3.6	1
16	Sequence Classification of Protein Families: Pfam and other Resources **2013**, 25-36		1
15	Automated Sequence-Based Approaches for Identifying Domain Families **2013**, 1-24		1
14	The Nucleophilic Attack Six-Bladed β-Propeller (N6P) Superfamily **2013**, 125-158		1
13	Evolution of Eukaryotic Chromatin Proteins and Transcription Factors **2013**, 421-502		1
12	AF4/FMR2 Family Homolog Regulates Heat-Shock-Induced Gene Expression. *Genetics*, **2020**, 215, 1039-1054		1
11	Periscope Proteins are variable-length regulators of bacterial cell surface interactions. *Proceedings of the National Academy of Sciences of the United States of America*, **2021**, 118,	11.5	1
10	The thrombospondin module 1 domain of the matricellular protein CCN3 shows an atypical disulfide pattern and incomplete CWR layers. *Acta Crystallographica Section D: Structural Biology*, **2020**, 76, 124-134	5.5	0
9	Metagenomics approach for Polymyxa betae genome assembly enables comparative analysis towards deciphering the intracellular parasitic lifestyle of the plasmodiophorids. *Genomics*, **2021**, 114, 9-22	4.3	0
8	Discovery of fibrillar adhesins across bacterial species. *BMC Genomics*, **2021**, 22, 550	4.5	0
7	A mouse model of occult intestinal colonization demonstrating antibiotic-induced outgrowth of carbapenem-resistant Enterobacteriaceae.. *Microbiome*, **2022**, 10, 43	16.6	0

List of Publications

6   Classifying Proteins into Domain Structure Families **2013**, 37-68

5   Structural Annotations of Genomes with Superfamily and G3D **2013**, 69-97

13

4   Functional Diversity of the HUP Domain Superfamily **2013**, 159-189

3   The NAD Binding Domain and the Short-Chain Dehydrogenase/Reductase (SDR) Superfamily **2013**, 191-206

2   Families of Sequence-Specific DNA-Binding Domains in Transcription Factors across the Tree of Life **2013**, 383-420

1   Navigating bacterial taxonomy in a world of unchartered microbial organisms. **2021**, 179-197