

# Alex Bateman

## List of Publications by Citations

**Source:** <https://exaly.com/author-pdf/5737384/alex-bateman-publications-by-citations.pdf>

**Version:** 2024-04-25

This document has been generated based on the publications and citations recorded by exaly.com. For the latest version of this publication list, visit the link given above.

The third column is the impact factor (IF) of the journal, and the fourth column is the number of citations of the article.

230  
papers

83,300  
citations

86  
h-index

258  
g-index

258  
ext. papers

106,035  
ext. citations

13  
avg, IF

7.86  
L-index

#	Paper	IF	Citations
230	Initial sequencing and analysis of the human genome. <i>Nature</i> , <b>2001</b> , 409, 860-921	50.4	17366
229	Pfam: the protein families database. <i>Nucleic Acids Research</i> , <b>2014</b> , 42, D222-30	20.1	3975
228	The Pfam protein families database: towards a more sustainable future. <i>Nucleic Acids Research</i> , <b>2016</b> , 44, D279-85	20.1	3678
227	miRBase: microRNA sequences, targets and gene nomenclature. <i>Nucleic Acids Research</i> , <b>2006</b> , 34, D140-4	20.1	3494
226	HMMER web server: interactive sequence similarity searching. <i>Nucleic Acids Research</i> , <b>2011</b> , 39, W29-37	20.1	2917
225	The Pfam protein families database. <i>Nucleic Acids Research</i> , <b>2012</b> , 40, D290-301	20.1	2844
224	The Pfam protein families database. <i>Nucleic Acids Research</i> , <b>2004</b> , 32, D138-41	20.1	2720
223	Complete genome sequence of the model actinomycete <i>Streptomyces coelicolor</i> A3(2). <i>Nature</i> , <b>2002</b> , 417, 141-7	50.4	2513
222	The Pfam protein families database. <i>Nucleic Acids Research</i> , <b>2010</b> , 38, D211-22	20.1	2467
221	The Pfam protein families database in 2019. <i>Nucleic Acids Research</i> , <b>2019</b> , 47, D427-D432	20.1	2298
220	The EMBL-EBI search and sequence analysis tools APIs in 2019. <i>Nucleic Acids Research</i> , <b>2019</b> , 47, W636-W641	20.1	2132
219	The Pfam protein families database. <i>Nucleic Acids Research</i> , <b>2002</b> , 30, 276-80	20.1	1839
218	The Pfam protein families database. <i>Nucleic Acids Research</i> , <b>2008</b> , 36, D281-8	20.1	1806
217	Pfam: clans, web tools and services. <i>Nucleic Acids Research</i> , <b>2006</b> , 34, D247-51	20.1	1784
216	InterPro: the integrative protein signature database. <i>Nucleic Acids Research</i> , <b>2009</b> , 37, D211-5	20.1	1379
215	The Pfam protein families database. <i>Nucleic Acids Research</i> , <b>2000</b> , 28, 263-6	20.1	1074
214	UniProt: the universal protein knowledgebase in 2021. <i>Nucleic Acids Research</i> , <b>2021</b> , 49, D480-D489	20.1	1073

213	Rfam: annotating non-coding RNAs in complete genomes. <i>Nucleic Acids Research</i> , <b>2005</b> , 33, D121-4	20.1	990
212	InterPro in 2017-beyond protein family and domain annotations. <i>Nucleic Acids Research</i> , <b>2017</b> , 45, D190-D199		970
211	Rfam: an RNA family database. <i>Nucleic Acids Research</i> , <b>2003</b> , 31, 439-41	20.1	970
210	The InterPro protein families database: the classification resource after 15 years. <i>Nucleic Acids Research</i> , <b>2015</b> , 43, D213-21	20.1	954
209	InterPro in 2011: new developments in the family and domain prediction database. <i>Nucleic Acids Research</i> , <b>2012</b> , 40, D306-12	20.1	844
208	InterPro in 2019: improving coverage, classification and access to protein sequence annotations. <i>Nucleic Acids Research</i> , <b>2019</b> , 47, D351-D360	20.1	835
207	Rfam 12.0: updates to the RNA families database. <i>Nucleic Acids Research</i> , <b>2015</b> , 43, D130-7	20.1	723
206	MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. <i>Nucleic Acids Research</i> , <b>2012</b> , 40, D343-50	20.1	686
205	MEROPS: the peptidase database. <i>Nucleic Acids Research</i> , <b>2010</b> , 38, D227-33	20.1	670
204	Rfam: updates to the RNA families database. <i>Nucleic Acids Research</i> , <b>2009</b> , 37, D136-40	20.1	652
203	The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. <i>Nucleic Acids Research</i> , <b>2018</b> , 46, D624-D632	20.1	643
202	Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. <i>Nucleic Acids Research</i> , <b>2013</b> , 41, e121	20.1	629
201	Pfam: The protein families database in 2021. <i>Nucleic Acids Research</i> , <b>2021</b> , 49, D412-D419	20.1	620
200	MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. <i>Nucleic Acids Research</i> , <b>2014</b> , 42, D503-9	20.1	602
199	HMMER web server: 2018 update. <i>Nucleic Acids Research</i> , <b>2018</b> , 46, W200-W204	20.1	596
198	Rfam 11.0: 10 years of RNA families. <i>Nucleic Acids Research</i> , <b>2013</b> , 41, D226-32	20.1	596
197	HMMER web server: 2015 update. <i>Nucleic Acids Research</i> , <b>2015</b> , 43, W30-8	20.1	590
196	The InterPro Database, 2003 brings increased coverage and new features. <i>Nucleic Acids Research</i> , <b>2003</b> , 31, 315-8	20.1	556

195	Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. <i>Nucleic Acids Research</i> , <b>2018</b> , 46, D335-D342	20.1	533
194	Pfam: multiple sequence alignments and HMM-profiles of protein domains. <i>Nucleic Acids Research</i> , <b>1998</b> , 26, 320-2	20.1	527
193	The Gene Ontology resource: enriching a GOld mine. <i>Nucleic Acids Research</i> , <b>2021</b> , 49, D325-D334	20.1	494
192	Pfam 3.1: 1313 multiple alignments and profile HMMs match the majority of proteins. <i>Nucleic Acids Research</i> , <b>1999</b> , 27, 260-2	20.1	488
191	A new genomic blueprint of the human gut microbiota. <i>Nature</i> , <b>2019</b> , 568, 499-504	50.4	484
190	InterPro, progress and status in 2005. <i>Nucleic Acids Research</i> , <b>2005</b> , 33, D201-5	20.1	426
189	The structure of a domain common to archaeobacteria and the homocystinuria disease protein. <i>Trends in Biochemical Sciences</i> , <b>1997</b> , 22, 12-3	10.3	399
188	Highly accurate protein structure prediction for the human proteome. <i>Nature</i> , <b>2021</b> , 596, 590-596	50.4	399
187	New developments in the InterPro database. <i>Nucleic Acids Research</i> , <b>2007</b> , 35, D224-8	20.1	397
186	The InterPro protein families and domains database: 20 years on. <i>Nucleic Acids Research</i> , <b>2021</b> , 49, D344-D354	20.1	358
185	The structure of a LysM domain from E. coli membrane-bound lytic murein transglycosylase D (MltD). <i>Journal of Molecular Biology</i> , <b>2000</b> , 299, 1113-9	6.5	343
184	Ensembl Genomes 2018: an integrated omics infrastructure for non-vertebrate species. <i>Nucleic Acids Research</i> , <b>2018</b> , 46, D802-D808	20.1	330
183	Domains in gene silencing and cell differentiation proteins: the novel PAZ domain and redefinition of the Piwi domain. <i>Trends in Biochemical Sciences</i> , <b>2000</b> , 25, 481-2	10.3	326
182	Rfam: Wikipedia, clans and the "decimal" release. <i>Nucleic Acids Research</i> , <b>2011</b> , 39, D141-5	20.1	321
181	Tissue-specific splicing of disordered segments that embed binding motifs rewires protein interaction networks. <i>Molecular Cell</i> , <b>2012</b> , 46, 871-83	17.6	272
180	iPfam: visualization of protein-protein interactions in PDB at domain and amino acid resolutions. <i>Bioinformatics</i> , <b>2005</b> , 21, 410-2	7.2	263
179	The tify family previously known as ZIM. <i>Trends in Plant Science</i> , <b>2007</b> , 12, 239-44	13.1	221
178	QuickTree: building huge Neighbour-Joining trees of protein sequences. <i>Bioinformatics</i> , <b>2002</b> , 18, 1546-7.2		214

177	Genome of <i>Acanthamoeba castellanii</i> highlights extensive lateral gene transfer and early evolution of tyrosine kinase signaling. <i>Genome Biology</i> , <b>2013</b> , 14, R11	18.3	205
176	Skyline: a tool for creating informative, interactive logos representing sequence alignments and profile hidden Markov models. <i>BMC Bioinformatics</i> , <b>2014</b> , 15, 7	3.6	196
175	The systematic functional analysis of <i>Plasmodium</i> protein kinases identifies essential regulators of mosquito transmission. <i>Cell Host and Microbe</i> , <b>2010</b> , 8, 377-87	23.4	195
174	The PASTA domain: a beta-lactam-binding domain. <i>Trends in Biochemical Sciences</i> , <b>2002</b> , 27, 438	10.3	193
173	A unified catalog of 204,938 reference genomes from the human gut microbiome. <i>Nature Biotechnology</i> , <b>2021</b> , 39, 105-114	44.5	185
172	The CHAP domain: a large family of amidases including GSP amidase and peptidoglycan hydrolases. <i>Trends in Biochemical Sciences</i> , <b>2003</b> , 28, 234-7	10.3	182
171	The structure of a PKD domain from polycystin-1: implications for polycystic kidney disease. <i>EMBO Journal</i> , <b>1999</b> , 18, 297-305	13	161
170	Large-scale screening for novel low-affinity extracellular protein interactions. <i>Genome Research</i> , <b>2008</b> , 18, 622-30	9.7	157
169	Non-Coding RNA Analysis Using the Rfam Database. <i>Current Protocols in Bioinformatics</i> , <b>2018</b> , 62, e51	24.2	152
168	Dfam: a database of repetitive DNA based on profile hidden Markov models. <i>Nucleic Acids Research</i> , <b>2013</b> , 41, D70-82	20.1	147
167	DUFs: families in search of function. <i>Acta Crystallographica Section F: Structural Biology Communications</i> , <b>2010</b> , 66, 1148-52		145
166	The Hotdog fold: wrapping up a superfamily of thioesterases and dehydratases. <i>BMC Bioinformatics</i> , <b>2004</b> , 5, 109	3.6	137
165	InterPro: an integrated documentation resource for protein families, domains and functional sites. <i>Briefings in Bioinformatics</i> , <b>2002</b> , 3, 225-35	13.4	137
164	Alternative splicing of intrinsically disordered regions and rewiring of protein interactions. <i>Current Opinion in Structural Biology</i> , <b>2013</b> , 23, 443-50	8.1	134
163	EBI Metagenomics in 2017: enriching the analysis of microbial communities, from sequence reads to assemblies. <i>Nucleic Acids Research</i> , <b>2018</b> , 46, D726-D735	20.1	130
162	The evolution of the DLK1-DIO3 imprinted domain in mammals. <i>PLoS Biology</i> , <b>2008</b> , 6, e135	9.7	130
161	MGNify: the microbiome analysis resource in 2020. <i>Nucleic Acids Research</i> , <b>2020</b> , 48, D570-D578	20.1	127
160	Predicting active site residue annotations in the Pfam database. <i>BMC Bioinformatics</i> , <b>2007</b> , 8, 298	3.6	124

159	iPfam: a database of protein family and domain interactions found in the Protein Data Bank. <i>Nucleic Acids Research</i> , <b>2014</b> , 42, D364-73	20.1	120
158	RNAcentral: a comprehensive database of non-coding RNA sequences. <i>Nucleic Acids Research</i> , <b>2017</b> , 45, D128-D134	20.1	119
157	Comparative analysis of the polycystic kidney disease 1 (PKD1) gene reveals an integral membrane glycoprotein with multiple evolutionary conserved domains. <i>Human Molecular Genetics</i> , <b>1997</b> , 6, 1483-9	5.6	116
156	The ENTH domain. <i>FEBS Letters</i> , <b>2002</b> , 513, 11-8	3.8	115
155	Rfam 14: expanded coverage of metagenomic, viral and microRNA families. <i>Nucleic Acids Research</i> , <b>2021</b> , 49, D192-D200	20.1	108
154	The PLAT domain: a new piece in the PKD1 puzzle. <i>Current Biology</i> , <b>1999</b> , 9, R588-90	6.3	104
153	Pfam 10 years on: 10,000 families and still growing. <i>Briefings in Bioinformatics</i> , <b>2008</b> , 9, 210-9	13.4	101
152	Membrane-bound progesterone receptors contain a cytochrome b5-like ligand-binding domain. <i>Genome Biology</i> , <b>2002</b> , 3, RESEARCH0068	18.3	101
151	TreeFam v9: a new website, more species and orthology-on-the-fly. <i>Nucleic Acids Research</i> , <b>2014</b> , 42, D922-5	20.1	96
150	Cache Domains That are Homologous to, but Different from PAS Domains Comprise the Largest Superfamily of Extracellular Sensors in Prokaryotes. <i>PLoS Computational Biology</i> , <b>2016</b> , 12, e1004862	5	95
149	Protein interactions in human genetic diseases. <i>Genome Biology</i> , <b>2008</b> , 9, R9	18.3	94
148	The European Bioinformatics Institute in 2016: Data growth and integration. <i>Nucleic Acids Research</i> , <b>2016</b> , 44, D20-6	20.1	91
147	RNAcentral: a hub of information for non-coding RNA sequences. <i>Nucleic Acids Research</i> , <b>2019</b> , 47, D221-D229	20.1	90
146	RNAcentral: an international database of ncRNA sequences. <i>Nucleic Acids Research</i> , <b>2015</b> , 43, D123-9	20.1	89
145	New knowledge from old: in silico discovery of novel protein domains in <i>Streptomyces coelicolor</i> . <i>BMC Microbiology</i> , <b>2003</b> , 3, 3	4.5	89
144	Tandem repeats lead to sequence assembly errors and impose multi-level challenges for genome and protein databases. <i>Nucleic Acids Research</i> , <b>2019</b> , 47, 10994-11006	20.1	86
143	A comparison of dense transposon insertion libraries in the <i>Salmonella</i> serovars Typhi and Typhimurium. <i>Nucleic Acids Research</i> , <b>2013</b> , 41, 4549-64	20.1	86
142	Metazoan Scc4 homologs link sister chromatid cohesion to cell and axon migration guidance. <i>PLoS Biology</i> , <b>2006</b> , 4, e242	9.7	83

141	The evolution of protein domain families. <i>Biochemical Society Transactions</i> , <b>2009</b> , 37, 751-5	5.1	81
140	EBI metagenomics in 2016--an expanding and evolving resource for the analysis and archiving of metagenomic data. <i>Nucleic Acids Research</i> , <b>2016</b> , 44, D595-603	20.1	81
139	Quantifying the mechanisms of domain gain in animal proteins. <i>Genome Biology</i> , <b>2010</b> , 11, R74	18.3	78
138	Integrating biological data--the Distributed Annotation System. <i>BMC Bioinformatics</i> , <b>2008</b> , 9 Suppl 8, S3	3.6	77
137	Dosage sensitivity shapes the evolution of copy-number varied regions. <i>PLoS ONE</i> , <b>2010</b> , 5, e9474	3.7	76
136	Structure and distribution of pentapeptide repeats in bacteria. <i>Protein Science</i> , <b>1998</b> , 7, 1477-80	6.3	76
135	The PA domain: a protease-associated domain. <i>Protein Science</i> , <b>2000</b> , 9, 1930-4	6.3	76
134	The SIS domain: a phosphosugar-binding domain. <i>Trends in Biochemical Sciences</i> , <b>1999</b> , 24, 94-5	10.3	74
133	The G5 domain: a potential N-acetylglucosamine recognition domain involved in biofilm formation. <i>Bioinformatics</i> , <b>2005</b> , 21, 1301-3	7.2	68
132	Representative proteomes: a stable, scalable and unbiased proteome set for sequence analysis and functional annotation. <i>PLoS ONE</i> , <b>2011</b> , 6, e18910	3.7	67
131	Asparagine peptide lyases: a seventh catalytic type of proteolytic enzymes. <i>Journal of Biological Chemistry</i> , <b>2011</b> , 286, 38321-38328	5.4	65
130	The BON domain: a putative membrane-binding domain. <i>Trends in Biochemical Sciences</i> , <b>2003</b> , 28, 352-5	10.3	64
129	The DUTT1 gene, a novel NCAM family member is expressed in developing murine neural tissues and has an unusually broad pattern of expression. <i>Molecular and Cellular Neurosciences</i> , <b>1998</b> , 11, 29-35	4.8	60
128	Integrating sequence and structural biology with DAS. <i>BMC Bioinformatics</i> , <b>2007</b> , 8, 333	3.6	57
127	Members of the immunoglobulin superfamily in bacteria. <i>Protein Science</i> , <b>1996</b> , 5, 1939-41	6.3	57
126	Computational strategies to combat COVID-19: useful tools to accelerate SARS-CoV-2 and coronavirus research. <i>Briefings in Bioinformatics</i> , <b>2021</b> , 22, 642-663	13.4	56
125	The PepSY domain: a regulator of peptidase activity in the microbial environment?. <i>Trends in Biochemical Sciences</i> , <b>2004</b> , 29, 169-72	10.3	55
124	RNAcentral: A vision for an international database of RNA sequences. <i>Rna</i> , <b>2011</b> , 17, 1941-6	5.8	54

123	Mining the draft human genome. <i>Nature</i> , <b>2001</b> , 409, 827-8	50.4	51
122	RNIE: genome-wide prediction of bacterial intrinsic terminators. <i>Nucleic Acids Research</i> , <b>2011</b> , 39, 5845-520.1	50.1	50
121	The RNA WikiProject: community annotation of RNA families. <i>Rna</i> , <b>2008</b> , 14, 2462-4	5.8	50
120	The HMMER Web Server for Protein Sequence Similarity Search. <i>Current Protocols in Bioinformatics</i> , <b>2017</b> , 60, 3.15.1-3.15.23	24.2	49
119	Biocurators and biocuration: surveying the 21st century challenges. <i>Database: the Journal of Biological Databases and Curation</i> , <b>2012</b> , 2012, bar059	5	49
118	The COMBEX project: design, methodology, and initial results. <i>PLoS Biology</i> , <b>2013</b> , 11, e1001638	9.7	47
117	Phospholipid scramblases and Tubby-like proteins belong to a new superfamily of membrane tethered transcription factors. <i>Bioinformatics</i> , <b>2009</b> , 25, 159-62	7.2	45
116	Identifying protein domains with the Pfam database. <i>Current Protocols in Bioinformatics</i> , <b>2008</b> , Chapter 2, Unit 2.5	24.2	43
115	Enhanced protein domain discovery by using language modeling techniques from speech recognition. <i>Proceedings of the National Academy of Sciences of the United States of America</i> , <b>2003</b> , 100, 4516-20	11.5	42
114	Filling out the structural map of the NTF2-like superfamily. <i>BMC Bioinformatics</i> , <b>2013</b> , 14, 327	3.6	40
113	Reuse of structural domain-domain interactions in protein networks. <i>BMC Bioinformatics</i> , <b>2007</b> , 8, 259	3.6	40
112	Structure of the ProQ RNA-binding protein. <i>Rna</i> , <b>2017</b> , 23, 696-711	5.8	38
111	Distant homology recognition using structural classification of proteins <b>1997</b> , 29, 105-112		37
110	SCOOP: a simple method for identification of novel protein superfamily relationships. <i>Bioinformatics</i> , <b>2007</b> , 23, 809-14	7.2	36
109	SnoPatrol: how many snoRNA genes are there?. <i>Journal of Biology</i> , <b>2010</b> , 9, 4		34
108	An introduction to hidden Markov models. <i>Current Protocols in Bioinformatics</i> , <b>2007</b> , Appendix 3, Appendix 3A	24.2	34
107	Visualizing profile-profile alignment: pairwise HMM logos. <i>Bioinformatics</i> , <b>2005</b> , 21, 2912-3	7.2	34
106	Using the MEROPS Database for Proteolytic Enzymes and Their Inhibitors and Substrates. <i>Current Protocols in Bioinformatics</i> , <b>2014</b> , 48, 1.25.1-33	24.2	33



105	ProServer: a simple, extensible Perl DAS server. <i>Bioinformatics</i> , <b>2007</b> , 23, 1568-70	7.2	31
104	Modifier effects between regulatory and protein-coding variation. <i>PLoS Genetics</i> , <b>2008</b> , 4, e1000244	6	30
103	The ELIXIR Core Data Resources: fundamental infrastructure for the life sciences. <i>Bioinformatics</i> , <b>2020</b> , 36, 2636-2642	7.2	29
102	The metagenomic data life-cycle: standards and best practices. <i>GigaScience</i> , <b>2017</b> , 6, 1-11	7.6	29
101	Microbial composition of Kombucha determined using amplicon sequencing and shotgun metagenomics. <i>Journal of Food Science</i> , <b>2020</b> , 85, 455-464	3.4	28
100	The Importance of Biological Databases in Biological Discovery. <i>Current Protocols in Bioinformatics</i> , <b>2015</b> , 50, 1.1.1-1.1.8	24.2	27
99	Pepsin homologues in bacteria. <i>BMC Genomics</i> , <b>2009</b> , 10, 437	4.5	27
98	InterPro in 2011: new developments in the family and domain prediction database. <i>Nucleic Acids Research</i> , <b>2012</b> , 40, 4725-4725	20.1	27
97	Bacterial pleckstrin homology domains: a prokaryotic origin for the PH domain. <i>Journal of Molecular Biology</i> , <b>2010</b> , 396, 31-46	6.5	25
96	The SGS3 protein involved in PTGS finds a family. <i>BMC Bioinformatics</i> , <b>2002</b> , 3, 21	3.6	25
95	INCREASING THE IMPACT OF BIOINFORMATICS. <i>Bioinformatics</i> , <b>2005</b> , 21, 1-1	7.2	25
94	Cryo-EM structures of human RNA polymerase III in its unbound and transcribing states. <i>Nature Structural and Molecular Biology</i> , <b>2021</b> , 28, 210-219	17.6	25
93	A member of the immunoglobulin superfamily in bacteriophage T4. <i>Virus Genes</i> , <b>1997</b> , 14, 163-5	2.3	24
92	Novel protein domains and motifs in the marine planctomycete <i>Rhodopirellula baltica</i> . <i>FEMS Microbiology Letters</i> , <b>2004</b> , 236, 333-340	2.9	24
91	Making your database available through Wikipedia: the pros and cons. <i>Nucleic Acids Research</i> , <b>2012</b> , 40, D9-12	20.1	23
90	HMM-based databases in InterPro. <i>Briefings in Bioinformatics</i> , <b>2002</b> , 3, 236-45	13.4	23
89	Fibronectin type III domains in yeast detected by a hidden Markov model. <i>Current Biology</i> , <b>1996</b> , 6, 1544-8.3	7.3	23
88	UniProt-DAAC: domain architecture alignment and classification, a new method for automatic functional annotation in UniProtKB. <i>Bioinformatics</i> , <b>2016</b> , 32, 2264-71	7.2	22

87	Experience using web services for biological sequence analysis. <i>Briefings in Bioinformatics</i> , <b>2008</b> , 9, 493-505	5.4	22
86	The TROVE module: a common element in Telomerase, Ro and Vault ribonucleoproteins. <i>BMC Bioinformatics</i> , <b>2003</b> , 4, 49	3.6	21
85	The yeast noncoding RNA interaction network. <i>Rna</i> , <b>2017</b> , 23, 1479-1492	5.8	19
84	The characterisation of three types of genes that overlie copy number variable regions. <i>PLoS ONE</i> , <b>2011</b> , 6, e14814	3.7	19
83	CASP2 knowledge-based approach to distant homology recognition and fold prediction in CASP4. <i>Proteins: Structure, Function and Bioinformatics</i> , <b>2001</b> , Suppl 5, 76-85	4.2	19
82	The use of structure information to increase alignment accuracy does not aid homologue detection with profile HMMs. <i>Bioinformatics</i> , <b>2002</b> , 18, 1243-9	7.2	19
81	Estimating the quality of eukaryotic genomes recovered from metagenomic analysis with EukCC. <i>Genome Biology</i> , <b>2020</b> , 21, 244	18.3	19
80	AntiFam: a tool to help identify spurious ORFs in protein annotation. <i>Database: the Journal of Biological Databases and Curation</i> , <b>2012</b> , 2012, bas003	5	18
79	Characterization of the nodulation plasmid encoded chemoreceptor gene mcpG from <i>Rhizobium leguminosarum</i> . <i>BMC Microbiology</i> , <b>2003</b> , 3, 1	4.5	17
78	Outline structures for the extracellular domains of the fibroblast growth factor receptors. <i>Nature Structural and Molecular Biology</i> , <b>1995</b> , 2, 1068-74	17.6	17
77	Domain atrophy creates rare cases of functional partial protein domains. <i>Genome Biology</i> , <b>2015</b> , 16, 88	18.3	16
76	GO annotation in InterPro: why stability does not indicate accuracy in a sea of changing annotations. <i>Database: the Journal of Biological Databases and Curation</i> , <b>2016</b> , 2016,	5	16
75	Tandem domain swapping: determinants of multidomain protein misfolding. <i>Current Opinion in Structural Biology</i> , <b>2019</b> , 58, 97-104	8.1	16
74	The challenge of increasing Pfam coverage of the human proteome. <i>Database: the Journal of Biological Databases and Curation</i> , <b>2013</b> , 2013, bat023	5	16
73	R2DT is a framework for predicting and visualising RNA secondary structure using templates. <i>Nature Communications</i> , <b>2021</b> , 12, 3494	17.4	16
72	Searching databases to find protein domain organization. <i>Advances in Protein Chemistry</i> , <b>2000</b> , 54, 137-57		15
71	Genome properties in 2019: a new companion database to InterPro for the inference of complete functional attributes. <i>Nucleic Acids Research</i> , <b>2019</b> , 47, D564-D572	20.1	15
70	DASMI: exchanging, annotating and assessing molecular interaction data. <i>Bioinformatics</i> , <b>2009</b> , 25, 1321-8	8.2	14

69	Enhanced protein domain discovery using taxonomy. <i>BMC Bioinformatics</i> , <b>2004</b> , 5, 56	3.6	14
68	A unified sequence catalogue of over 280,000 genomes obtained from the human gut microbiome		14
67	How to use the MEROPS database and website to help understand peptidase specificity. <i>Protein Science</i> , <b>2021</b> , 30, 83-92	6.3	14
66	Origins of peptidases. <i>Biochimie</i> , <b>2019</b> , 166, 4-18	4.6	13
65	Identifying protein domains with the Pfam database. <i>Current Protocols in Bioinformatics</i> , <b>2003</b> , Chapter 2, Unit 2.5	24.2	13
64	Novel protein domains and motifs in the marine planctomycete <i>Rhodopirellula baltica</i> . <i>FEMS Microbiology Letters</i> , <b>2004</b> , 236, 333-40	2.9	13
63	Patterns of database citation in articles and patents indicate long-term scientific and industry value of biological data resources. <i>F1000Research</i> , <b>2016</b> , 5,	3.6	13
62	Using deep learning to annotate the protein universe.. <i>Nature Biotechnology</i> , <b>2022</b> ,	44.5	13
61	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. <i>Protein Science</i> , <b>2010</b> , 19, 2131-40	6.3	12
60	Ensembl Genomes 2022: an expanding genome resource for non-vertebrates. <i>Nucleic Acids Research</i> , <b>2021</b> ,	20.1	12
59	TreeGrafter: phylogenetic tree-based annotation of proteins with Gene Ontology terms and other annotations. <i>Bioinformatics</i> , <b>2019</b> , 35, 518-520	7.2	12
58	The structure of Jann_2411 (DUF1470) from <i>Jannaschia</i> sp. at 1.45 Å resolution reveals a new fold (the ABATE domain) and suggests its possible role as a transcription regulator. <i>Acta Crystallographica Section F: Structural Biology Communications</i> , <b>2010</b> , 66, 1198-204		9
57	Pfam: the protein families database <b>2005</b> ,		9
56	Unifying the known and unknown microbial coding sequence space		9
55	Acetylation of Surface Carbohydrates in Bacterial Pathogens Requires Coordinated Action of a Two-Domain Membrane-Bound Acyltransferase. <i>MBio</i> , <b>2020</b> , 11,	7.8	9
54	Time to underpin Wikipedia wisdom. <i>Nature</i> , <b>2010</b> , 468, 765	50.4	8
53	Predicted Input of Uncultured Fungal Symbionts to a Lichen Symbiosis from Metagenome-Assembled Genomes. <i>Genome Biology and Evolution</i> , <b>2021</b> , 13,	3.9	8
52	Genome3D: integrating a collaborative data pipeline to expand the depth and breadth of consensus protein structure annotation. <i>Nucleic Acids Research</i> , <b>2020</b> , 48, D314-D319	20.1	7

51	Eleven quick tips to build a usable REST API for life sciences. <i>PLoS Computational Biology</i> , <b>2018</b> , 14, e1006542	7	
50	LUD, a new protein domain associated with lactate utilization. <i>BMC Bioinformatics</i> , <b>2013</b> , 14, 341	3.6	6
49	The structure of pyogenecin immunity protein, a novel bacteriocin-like immunity protein from <i>Streptococcus pyogenes</i> . <i>BMC Structural Biology</i> , <b>2009</b> , 9, 75	2.7	6
48	The structure of BVU2987 from <i>Bacteroides vulgatus</i> reveals a superfamily of bacterial periplasmic proteins with possible inhibitory function. <i>Acta Crystallographica Section F: Structural Biology Communications</i> , <b>2010</b> , 66, 1265-73		6
47	A comparison of Pfam and MEROPS: two databases, one comprehensive, and one specialised. <i>BMC Bioinformatics</i> , <b>2003</b> , 4, 17	3.6	6
46	ELIXIR pilot action: Marine metagenomics - towards a domain specific set of sustainable services. <i>F1000Research</i> , <b>2017</b> , 6,	3.6	6
45	Gene Unprediction with Spurio: A tool to identify spurious protein sequences. <i>F1000Research</i> , <b>2018</b> , 7, 261	3.6	6
44	Defining the remarkable structural malleability of a bacterial surface protein Rib domain implicated in infection. <i>Proceedings of the National Academy of Sciences of the United States of America</i> , <b>2019</b> ,	11.5	6
43	Integrating cultivation and metagenomics for a multi-kingdom view of skin microbiome diversity and functions.. <i>Nature Microbiology</i> , <b>2021</b> ,	26.6	6
42	The complexity, challenges and benefits of comparing two transporter classification systems in TCDB and Pfam. <i>Briefings in Bioinformatics</i> , <b>2015</b> , 16, 865-72	13.4	5
41	Creating a specialist protein resource network: a meeting report for the protein bioinformatics and community resources retreat. <i>Database: the Journal of Biological Databases and Curation</i> , <b>2015</b> , 2015, bav063	5	5
40	Computer Tools and Algorithms for Origami Tessellation Design <b>2002</b> , 121-127		4
39	Sequence analysis of tyrosine recombinases allows annotation of mobile genetic elements in prokaryotic genomes. <i>Molecular Systems Biology</i> , <b>2021</b> , 17, e9880	12.2	4
38	TADOSS: computational estimation of tandem domain swap stability. <i>Bioinformatics</i> , <b>2019</b> , 35, 2507-2508	8.2	4
37	Unifying the known and unknown microbial coding sequence space.. <i>ELife</i> , <b>2022</b> , 11,	8.9	4
36	Two Pfam protein families characterized by a crystal structure of protein lpg2210 from <i>Legionella pneumophila</i> . <i>BMC Bioinformatics</i> , <b>2013</b> , 14, 265	3.6	3
35	The YARHG domain: an extracellular domain in search of a function. <i>PLoS ONE</i> , <b>2012</b> , 7, e35575	3.7	3
34	ISCB computational biology Wikipedia competition. <i>PLoS Computational Biology</i> , <b>2013</b> , 9, e1003242	5	3

33	The SHOCT domain: a widespread domain under-represented in model organisms. <i>PLoS ONE</i> , <b>2013</b> , 8, e57848	3.7	3
32	R2DT: computational framework for template-based RNA secondary structure visualisation across non-coding RNA types		3
31	Exploring Non-Coding RNAs in RNAcentral. <i>Current Protocols in Bioinformatics</i> , <b>2020</b> , 71, e104	24.2	3
30	The Ensembl COVID-19 resource: ongoing integration of public SARS-CoV-2 data. <i>Nucleic Acids Research</i> , <b>2021</b> ,	20.1	3
29	3DPatch: fast 3D structure visualization with residue conservation. <i>Bioinformatics</i> , <b>2019</b> , 35, 332-334	7.2	2
28	New Leadership for Bioinformatics. <i>Bioinformatics</i> , <b>2004</b> , 20, 1821-1821	7.2	2
27	Folding the unfoldable: using AlphaFold to explore spurious proteins. <i>Bioinformatics Advances</i> , <b>2022</b> , 2,		2
26	The European Bioinformatics Institute (EMBL-EBI) in 2021. <i>Nucleic Acids Research</i> , <b>2021</b> ,	20.1	2
25	On expert curation and sustainability: UniProtKB/Swiss-Prot as a case study		2
24	Cryo-EM structures of human RNA polymerase III in its unbound and transcribing states		2
23	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. <i>BMC Bioinformatics</i> , <b>2014</b> , 15, 75	3.6	1
22	Sequence Classification of Protein Families: Pfam and other Resources <b>2013</b> , 25-36		1
21	Automated Sequence-Based Approaches for Identifying Domain Families <b>2013</b> , 1-24		1
20	The Nucleophilic Attack Six-Bladed $\beta$ Propeller (N6P) Superfamily <b>2013</b> , 125-158		1
19	Evolution of Eukaryotic Chromatin Proteins and Transcription Factors <b>2013</b> , 421-502		1
18	Bioinformatics--The new home for protein sequence motifs. <i>Bioinformatics</i> , <b>2006</b> , 22, 2-2	7.2	1
17	Software patents in Bioinformatics. <i>Bioinformatics</i> , <b>2006</b> , 22, 1415-1415	7.2	1
16	Periscope Proteins are variable length regulators of bacterial cell surface interactions		1

15	AF4/FMR2 Family Homolog Regulates Heat-Shock-Induced Gene Expression. <i>Genetics</i> , <b>2020</b> , 215, 1039-1054	10.54	1
14	Periscope Proteins are variable-length regulators of bacterial cell surface interactions. <i>Proceedings of the National Academy of Sciences of the United States of America</i> , <b>2021</b> , 118,	11.5	1
13	The thrombospondin module 1 domain of the matricellular protein CCN3 shows an atypical disulfide pattern and incomplete CWR layers. <i>Acta Crystallographica Section D: Structural Biology</i> , <b>2020</b> , 76, 124-134	5.5	0
12	Metagenomics approach for <i>Polymyxa betae</i> genome assembly enables comparative analysis towards deciphering the intracellular parasitic lifestyle of the plasmodiophorids. <i>Genomics</i> , <b>2021</b> , 114, 9-22	4.3	0
11	Discovery of fibrillar adhesins across bacterial species. <i>BMC Genomics</i> , <b>2021</b> , 22, 550	4.5	0
10	Classifying Proteins into Domain Structure Families <b>2013</b> , 37-68		
9	Structural Annotations of Genomes with Superfamily and G3D <b>2013</b> , 69-97		
8	Phylogenomic Databases and Orthology Prediction <b>2013</b> , 99-124		
7	Functional Diversity of the HUP Domain Superfamily <b>2013</b> , 159-189		
6	The NAD Binding Domain and the Short-Chain Dehydrogenase/Reductase (SDR) Superfamily <b>2013</b> , 191-206		
5	Functional Adaptation and Plasticity in Cytoskeletal Protein Domains: Lessons from the Erythrocyte Model <b>2013</b> , 237-284		
4	Unusual Species Distribution and Horizontal Transfer of Peptidases <b>2013</b> , 285-314		
3	Families of Sequence-Specific DNA-Binding Domains in Transcription Factors across the Tree of Life <b>2013</b> , 383-420		
2	An update from the Bioinformatics Editors. <i>Bioinformatics</i> , <b>2005</b> , 21, 4319-4319	7.2	
1	Modelling structural rearrangements in proteins using Euclidean distance matrices. <i>F1000Research</i> , <b>2019</b> , 9, 728	3.6	