

Alex Bateman

List of Publications by Year in descending order

Source: <https://exaly.com/author-pdf/5737384/publications.pdf>

Version: 2024-02-01

229
papers

124,268
citations

3515

90
h-index

1250

226
g-index

258
all docs

258
docs citations

258
times ranked

114912
citing authors

#	ARTICLE	IF	CITATIONS
1	Initial sequencing and analysis of the human genome. <i>Nature</i> , 2001, 409, 860-921.	13.7	21,074
2	The Pfam protein families database. <i>Nucleic Acids Research</i> , 2007, 36, D281-D288.	6.5	6,372
3	Pfam: the protein families database. <i>Nucleic Acids Research</i> , 2014, 42, D222-D230.	6.5	5,425
4	The Pfam protein families database: towards a more sustainable future. <i>Nucleic Acids Research</i> , 2016, 44, D279-D285.	6.5	5,391
5	UniProt: the universal protein knowledgebase in 2021. <i>Nucleic Acids Research</i> , 2021, 49, D480-D489.	6.5	4,709
6	HMMER web server: interactive sequence similarity searching. <i>Nucleic Acids Research</i> , 2011, 39, W29-W37.	6.5	4,492
7	miRBase: microRNA sequences, targets and gene nomenclature. <i>Nucleic Acids Research</i> , 2006, 34, D140-D144.	6.5	4,192
8	The Pfam protein families database in 2019. <i>Nucleic Acids Research</i> , 2019, 47, D427-D432.	6.5	3,937
9	The EMBL-EBI search and sequence analysis tools APIs in 2019. <i>Nucleic Acids Research</i> , 2019, 47, W636-W641.	6.5	3,820
10	The Pfam protein families database. <i>Nucleic Acids Research</i> , 2012, 40, D290-D301.	6.5	3,306
11	The Pfam protein families database. <i>Nucleic Acids Research</i> , 2004, 32, 138D-141.	6.5	3,084
12	Pfam: The protein families database in 2021. <i>Nucleic Acids Research</i> , 2021, 49, D412-D419.	6.5	3,068
13	Complete genome sequence of the model actinomycete <i>Streptomyces coelicolor</i> A3(2). <i>Nature</i> , 2002, 417, 141-147.	13.7	2,940
14	The Pfam protein families database. <i>Nucleic Acids Research</i> , 2010, 38, D211-D222.	6.5	2,693
15	The Gene Ontology resource: enriching a GOld mine. <i>Nucleic Acids Research</i> , 2021, 49, D325-D334.	6.5	2,416
16	The Pfam Protein Families Database. <i>Nucleic Acids Research</i> , 2002, 30, 276-280.	6.5	2,067
17	Pfam: clans, web tools and services. <i>Nucleic Acids Research</i> , 2006, 34, D247-D251.	6.5	2,030
18	Highly accurate protein structure prediction for the human proteome. <i>Nature</i> , 2021, 596, 590-596.	13.7	1,773

#	ARTICLE	IF	CITATIONS
19	InterPro: the integrative protein signature database. <i>Nucleic Acids Research</i> , 2009, 37, D211-D215.	6.5	1,712
20	HMMER web server: 2018 update. <i>Nucleic Acids Research</i> , 2018, 46, W200-W204.	6.5	1,432
21	The InterPro protein families and domains database: 20 years on. <i>Nucleic Acids Research</i> , 2021, 49, D344-D354.	6.5	1,385
22	InterPro in 2017â€”beyond protein family and domain annotations. <i>Nucleic Acids Research</i> , 2017, 45, D190-D199.	6.5	1,358
23	Rfam: an RNA family database. <i>Nucleic Acids Research</i> , 2003, 31, 439-441.	6.5	1,296
24	InterPro in 2019: improving coverage, classification and access to protein sequence annotations. <i>Nucleic Acids Research</i> , 2019, 47, D351-D360.	6.5	1,291
25	Rfam: annotating non-coding RNAs in complete genomes. <i>Nucleic Acids Research</i> , 2004, 33, D121-D124.	6.5	1,255
26	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> , 2018, 46, D624-D632.	6.5	1,234
27	Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. <i>Nucleic Acids Research</i> , 2013, 41, e121-e121.	6.5	1,214
28	The InterPro protein families database: the classification resource after 15 years. <i>Nucleic Acids Research</i> , 2015, 43, D213-D221.	6.5	1,205
29	The Pfam Protein Families Database. <i>Nucleic Acids Research</i> , 2000, 28, 263-266.	6.5	1,173
30	MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. <i>Nucleic Acids Research</i> , 2012, 40, D343-D350.	6.5	1,047
31	Rfam 12.0: updates to the RNA families database. <i>Nucleic Acids Research</i> , 2015, 43, D130-D137.	6.5	1,000
32	InterPro in 2011: new developments in the family and domain prediction database. <i>Nucleic Acids Research</i> , 2012, 40, D306-D312.	6.5	921
33	A new genomic blueprint of the human gut microbiota. <i>Nature</i> , 2019, 568, 499-504.	13.7	901
34	HMMER web server: 2015 update. <i>Nucleic Acids Research</i> , 2015, 43, W30-W38.	6.5	849
35	Rfam: updates to the RNA families database. <i>Nucleic Acids Research</i> , 2009, 37, D136-D140.	6.5	820
36	Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. <i>Nucleic Acids Research</i> , 2018, 46, D335-D342.	6.5	819

#	ARTICLE	IF	CITATIONS
37	MEROPS: the peptidase database. <i>Nucleic Acids Research</i> , 2010, 38, D227-D233.	6.5	786
38	<i>MEROPS</i>: the database of proteolytic enzymes, their substrates and inhibitors. <i>Nucleic Acids Research</i> , 2014, 42, D503-D509.	6.5	782
39	Rfam 11.0: 10 years of RNA families. <i>Nucleic Acids Research</i> , 2013, 41, D226-D232.	6.5	745
40	Pfam: multiple sequence alignments and HMM-profiles of protein domains. <i>Nucleic Acids Research</i> , 1998, 26, 320-322.	6.5	644
41	The InterPro Database, 2003 brings increased coverage and new features. <i>Nucleic Acids Research</i> , 2003, 31, 315-318.	6.5	640
42	A unified catalog of 204,938 reference genomes from the human gut microbiome. <i>Nature Biotechnology</i> , 2021, 39, 105-114.	9.4	628
43	Pfam 3.1: 1313 multiple alignments and profile HMMs match the majority of proteins. <i>Nucleic Acids Research</i> , 1999, 27, 260-262.	6.5	533
44	Ensembl Genomes 2018: an integrated omics infrastructure for non-vertebrate species. <i>Nucleic Acids Research</i> , 2018, 46, D802-D808.	6.5	489
45	InterPro, progress and status in 2005. <i>Nucleic Acids Research</i> , 2004, 33, D201-D205.	6.5	478
46	Rfam 14: expanded coverage of metagenomic, viral and microRNA families. <i>Nucleic Acids Research</i> , 2021, 49, D192-D200.	6.5	475
47	The structure of a domain common to archaeobacteria and the homocystinuria disease protein. <i>Trends in Biochemical Sciences</i> , 1997, 22, 12-13.	3.7	459
48	New developments in the InterPro database. <i>Nucleic Acids Research</i> , 2007, 35, D224-D228.	6.5	444
49	The structure of a LysM domain from E. coli membrane-bound lytic murein transglycosylase D (MltD) 1 Edited by P. E. Wight. <i>Journal of Molecular Biology</i> , 2000, 299, 1113-1119.	2.0	402
50	Domains in gene silencing and cell differentiation proteins: the novel PAZ domain and redefinition of the Piwi domain. <i>Trends in Biochemical Sciences</i> , 2000, 25, 481-482.	3.7	371
51	Rfam: Wikipedia, clans and the "decimal" release. <i>Nucleic Acids Research</i> , 2011, 39, D141-D145.	6.5	355
52	Tissue-Specific Splicing of Disordered Segments that Embed Binding Motifs Rewires Protein Interaction Networks. <i>Molecular Cell</i> , 2012, 46, 871-883.	4.5	344
53	Non-coding RNA Analysis Using the Rfam Database. <i>Current Protocols in Bioinformatics</i> , 2018, 62, e51.	25.8	309
54	Skylogn: a tool for creating informative, interactive logos representing sequence alignments and profile hidden Markov models. <i>BMC Bioinformatics</i> , 2014, 15, 7.	1.2	298

#	ARTICLE	IF	CITATIONS
55	Genome of <i>Acanthamoeba castellanii</i> highlights extensive lateral gene transfer and early evolution of tyrosine kinase signaling. <i>Genome Biology</i> , 2013, 14, R11.	13.9	296
56	MGnify: the microbiome analysis resource in 2020. <i>Nucleic Acids Research</i> , 2020, 48, D570-D578.	6.5	296
57	iPfam: visualization of protein-protein interactions in PDB at domain and amino acid resolutions. <i>Bioinformatics</i> , 2005, 21, 410-412.	1.8	295
58	The tify family previously known as ZIM. <i>Trends in Plant Science</i> , 2007, 12, 239-244.	4.3	283
59	The Systematic Functional Analysis of Plasmodium Protein Kinases Identifies Essential Regulators of Mosquito Transmission. <i>Cell Host and Microbe</i> , 2010, 8, 377-387.	5.1	267
60	QuickTree: building huge Neighbour-Joining trees of protein sequences. <i>Bioinformatics</i> , 2002, 18, 1546-1547.	1.8	249
61	Dfam: a database of repetitive DNA based on profile hidden Markov models. <i>Nucleic Acids Research</i> , 2012, 41, D70-D82.	6.5	243
62	Predicting active site residue annotations in the Pfam database. <i>BMC Bioinformatics</i> , 2007, 8, 298.	1.2	239
63	Tandem repeats lead to sequence assembly errors and impose multi-level challenges for genome and protein databases. <i>Nucleic Acids Research</i> , 2019, 47, 10994-11006.	6.5	236
64	The CHAP domain: a large family of amidases including GSP amidase and peptidoglycan hydrolases. <i>Trends in Biochemical Sciences</i> , 2003, 28, 234-237.	3.7	209
65	The PASTA domain: a β^2 -lactam-binding domain. <i>Trends in Biochemical Sciences</i> , 2002, 27, 438-440.	3.7	204
66	DUFs: families in search of function. <i>Acta Crystallographica Section F: Structural Biology Communications</i> , 2010, 66, 1148-1152.	0.7	203
67	The structure of a PKD domain from polycystin-1: implications for polycystic kidney disease. <i>EMBO Journal</i> , 1999, 18, 297-305.	3.5	186
68	Large-scale screening for novel low-affinity extracellular protein interactions. <i>Genome Research</i> , 2008, 18, 622-630.	2.4	184
69	EBI Metagenomics in 2017: enriching the analysis of microbial communities, from sequence reads to assemblies. <i>Nucleic Acids Research</i> , 2018, 46, D726-D735.	6.5	175
70	RNAcentral: a comprehensive database of non-coding RNA sequences. <i>Nucleic Acids Research</i> , 2017, 45, D128-D134.	6.5	174
71	Alternative splicing of intrinsically disordered regions and rewiring of protein interactions. <i>Current Opinion in Structural Biology</i> , 2013, 23, 443-450.	2.6	166
72	The Evolution of the DLK1-DIO3 Imprinted Domain in Mammals. <i>PLoS Biology</i> , 2008, 6, e135.	2.6	162

#	ARTICLE	IF	CITATIONS
73	The Hotdog fold: wrapping up a superfamily of thioesterases and dehydratases. <i>BMC Bioinformatics</i> , 2004, 5, 109.	1.2	158
74	iPfam: a database of protein family and domain interactions found in the Protein Data Bank. <i>Nucleic Acids Research</i> , 2014, 42, D364-D373.	6.5	156
75	InterPro: An integrated documentation resource for protein families, domains and functional sites. <i>Briefings in Bioinformatics</i> , 2002, 3, 225-235.	3.2	155
76	RNAcentral: a hub of information for non-coding RNA sequences. <i>Nucleic Acids Research</i> , 2019, 47, D221-D229.	6.5	153
77	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> , 2016, 12, e1004862.	1.5	147
78	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> , 1997, 6, 1483-1489.	1.4	141
79	Ensembl Genomes 2022: an expanding genome resource for non-vertebrates. <i>Nucleic Acids Research</i> , 2022, 50, D996-D1003.	6.5	141
80	Using deep learning to annotate the protein universe. <i>Nature Biotechnology</i> , 2022, 40, 932-937.	9.4	133
81	The ENTH domain. <i>FEBS Letters</i> , 2002, 513, 11-18.	1.3	131
82	TreeFam v9: a new website, more species and orthology-on-the-fly. <i>Nucleic Acids Research</i> , 2014, 42, D922-D925.	6.5	125
83	The PLAT domain: a new piece in the PKD1 puzzle. <i>Current Biology</i> , 1999, 9, R588-S2.	1.8	122
84	The HMMER Web Server for Protein Sequence Similarity Search. <i>Current Protocols in Bioinformatics</i> , 2017, 60, 3.15.1-3.15.23.	25.8	119
85	Membrane-bound progesterone receptors contain a cytochrome b5-like ligand-binding domain. <i>Genome Biology</i> , 2002, 3, research0068.1.	13.9	117
86	Pfam 10 years on: 10 000 families and still growing. <i>Briefings in Bioinformatics</i> , 2008, 9, 210-219.	3.2	114
87	Protein interactions in human genetic diseases. <i>Genome Biology</i> , 2008, 9, R9.	13.9	110
88	Computational strategies to combat COVID-19: useful tools to accelerate SARS-CoV-2 and coronavirus research. <i>Briefings in Bioinformatics</i> , 2021, 22, 642-663.	3.2	110
89	A comparison of dense transposon insertion libraries in the <i>Salmonella</i> serovars Typhi and Typhimurium. <i>Nucleic Acids Research</i> , 2013, 41, 4549-4564.	6.5	108
90	The European Bioinformatics Institute in 2016: Data growth and integration. <i>Nucleic Acids Research</i> , 2016, 44, D20-D26.	6.5	108

#	ARTICLE	IF	CITATIONS
91	The evolution of protein domain families. <i>Biochemical Society Transactions</i> , 2009, 37, 751-755.	1.6	106
92	RNAcentral: an international database of ncRNA sequences. <i>Nucleic Acids Research</i> , 2015, 43, D123-D129.	6.5	103
93	New knowledge from old: in silico discovery of novel protein domains in <i>Streptomyces coelicolor</i> . <i>BMC Microbiology</i> , 2003, 3, 3.	1.3	98
94	EBI metagenomics in 2016 - an expanding and evolving resource for the analysis and archiving of metagenomic data. <i>Nucleic Acids Research</i> , 2016, 44, D595-D603.	6.5	97
95	Metazoan Scc4 Homologs Link Sister Chromatid Cohesion to Cell and Axon Migration Guidance. <i>PLoS Biology</i> , 2006, 4, e242.	2.6	95
96	Representative Proteomes: A Stable, Scalable and Unbiased Proteome Set for Sequence Analysis and Functional Annotation. <i>PLoS ONE</i> , 2011, 6, e18910.	1.1	94
97	Quantifying the mechanisms of domain gain in animal proteins. <i>Genome Biology</i> , 2010, 11, R74.	13.9	93
98	The BON domain: a putative membrane-binding domain. <i>Trends in Biochemical Sciences</i> , 2003, 28, 352-355.	3.7	89
99	Asparagine Peptide Lyases. <i>Journal of Biological Chemistry</i> , 2011, 286, 38321-38328.	1.6	89
100	The PA domain: A protease-associated domain. <i>Protein Science</i> , 2000, 9, 1930-1934.	3.1	88
101	Dosage Sensitivity Shapes the Evolution of Copy-Number Varied Regions. <i>PLoS ONE</i> , 2010, 5, e9474.	1.1	88
102	Structure and distribution of pentapeptide repeats in bacteria. <i>Protein Science</i> , 1998, 7, 1477-1480.	3.1	87
103	Integrating biological data into the Distributed Annotation System. <i>BMC Bioinformatics</i> , 2008, 9, S3.	1.2	87
104	The SIS domain: a phosphosugar-binding domain. <i>Trends in Biochemical Sciences</i> , 1999, 24, 94-95.	3.7	78
105	The G5 domain: a potential N-acetylglucosamine recognition domain involved in biofilm formation. <i>Bioinformatics</i> , 2005, 21, 1301-1303.	1.8	78
106	Filling out the structural map of the NTF2-like superfamily. <i>BMC Bioinformatics</i> , 2013, 14, 327.	1.2	74
107	Microbial composition of Kombucha determined using amplicon sequencing and shotgun metagenomics. <i>Journal of Food Science</i> , 2020, 85, 455-464.	1.5	72
108	RNIE: genome-wide prediction of bacterial intrinsic terminators. <i>Nucleic Acids Research</i> , 2011, 39, 5845-5852.	6.5	71

#	ARTICLE	IF	CITATIONS
109	TheDUTT1Gene, 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> , 1998, 11, 29-35.	1.0	68
110	The PepSY domain: a regulator of peptidase activity in the microbial environment?. <i>Trends in Biochemical Sciences</i> , 2004, 29, 169-172.	3.7	68
111	Integrating sequence and structural biology with DAS. <i>BMC Bioinformatics</i> , 2007, 8, 333.	1.2	68
112	RNAcentral: A vision for an international database of RNA sequences. <i>Rna</i> , 2011, 17, 1941-1946.	1.6	67
113	Members of the immunoglobulin superfamily in bacteria. <i>Protein Science</i> , 1996, 5, 1939-1941.	3.1	66
114	The RNA WikiProject: Community annotation of RNA families. <i>Rna</i> , 2008, 14, 2462-2464.	1.6	66
115	Identifying Protein Domains with the Pfam Database. <i>Current Protocols in Bioinformatics</i> , 2008, 23, Unit 2.5.	25.8	61
116	An Introduction to Hidden Markov Models. , 2007, Appendix 3, Appendix 3A.		60
117	Phospholipid scramblases and Tubby-like proteins belong to a new superfamily of membrane tethered transcription factors. <i>Bioinformatics</i> , 2009, 25, 159-162.	1.8	59
118	Biocurators and Biocuration: surveying the 21st century challenges. <i>Database: the Journal of Biological Databases and Curation</i> , 2012, 2012, bar059-bar059.	1.4	59
119	Cryo-EM structures of human RNA polymerase III in its unbound and transcribing states. <i>Nature Structural and Molecular Biology</i> , 2021, 28, 210-219.	3.6	59
120	Mining the draft human genome. <i>Nature</i> , 2001, 409, 827-828.	13.7	58
121	Estimating the quality of eukaryotic genomes recovered from metagenomic analysis with EukCC. <i>Genome Biology</i> , 2020, 21, 244.	3.8	58
122	R2DT is a framework for predicting and visualising RNA secondary structure using templates. <i>Nature Communications</i> , 2021, 12, 3494.	5.8	58
123	Integrating cultivation and metagenomics for a multi-kingdom view of skin microbiome diversity and functions. <i>Nature Microbiology</i> , 2022, 7, 169-179.	5.9	58
124	Cloud computing. <i>Bioinformatics</i> , 2009, 25, 1475-1475.	1.8	56
125	The COMBREX Project: Design, Methodology, and Initial Results. <i>PLoS Biology</i> , 2013, 11, e1001638.	2.6	54
126	Bacterial retrons encode phage-defending tripartite toxin-antitoxin systems. <i>Nature</i> , 2022, 609, 144-150.	13.7	52

#	ARTICLE	IF	CITATIONS
127	Structure of the <i>Escherichia coli</i> ProQ RNA-binding protein. <i>Rna</i> , 2017, 23, 696-711.	1.6	50
128	SCOOP: a simple method for identification of novel protein superfamily relationships. <i>Bioinformatics</i> , 2007, 23, 809-814.	1.8	49
129	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> , 2003, 100, 4516-4520.	3.3	47
130	The ELIXIR Core Data Resources: fundamental infrastructure for the life sciences. <i>Bioinformatics</i> , 2020, 36, 2636-2642.	1.8	47
131	Reuse of structural domain-domain interactions in protein networks. <i>BMC Bioinformatics</i> , 2007, 8, 259.	1.2	44
132	How to use the MEROPS database and website to help understand peptidase specificity. <i>Protein Science</i> , 2021, 30, 83-92.	3.1	44
133	Databases, data tombs and dust in the wind. <i>Bioinformatics</i> , 2008, 24, 2127-2128.	1.8	42
134	The metagenomic data life-cycle: standards and best practices. <i>GigaScience</i> , 2017, 6, 1-11.	3.3	42
135	Unifying the known and unknown microbial coding sequence space. <i>ELife</i> , 2022, 11, .	2.8	41
136	Distant homology recognition using structural classification of proteins. , 1997, 29, 105-112.		40
137	Using the MEROPS Database for Proteolytic Enzymes and Their Inhibitors and Substrates. <i>Current Protocols in Bioinformatics</i> , 2014, 48, 1.25.1-33.	25.8	39
138	Curators of the world unite: the International Society of Biocuration. <i>Bioinformatics</i> , 2010, 26, 991-991.	1.8	38
139	AntiFam: a tool to help identify spurious ORFs in protein annotation. <i>Database: the Journal of Biological Databases and Curation</i> , 2012, 2012, bas003-bas003.	1.4	38
140	INCREASING THE IMPACT OF BIOINFORMATICS. <i>Bioinformatics</i> , 2005, 21, 1-1.	1.8	37
141	SnoPatrol: how many snoRNA genes are there?. <i>Journal of Biology</i> , 2010, 9, 4.	2.7	37
142	UniProt-DAAC: domain architecture alignment and classification, a new method for automatic functional annotation in UniProtKB. <i>Bioinformatics</i> , 2016, 32, 2264-2271.	1.8	37
143	Ten Simple Rules for Editing Wikipedia. <i>PLoS Computational Biology</i> , 2010, 6, e1000941.	1.5	36
144	Visualizing profile-profile alignment: pairwise HMM logos. <i>Bioinformatics</i> , 2005, 21, 2912-2913.	1.8	35

#	ARTICLE	IF	CITATIONS
145	ProServer: a simple, extensible Perl DAS server. <i>Bioinformatics</i> , 2007, 23, 1568-1570.	1.8	34
146	The European Bioinformatics Institute (EMBL-EBI) in 2021. <i>Nucleic Acids Research</i> , 2022, 50, D11-D19.	6.5	34
147	Modifier Effects between Regulatory and Protein-Coding Variation. <i>PLoS Genetics</i> , 2008, 4, e1000244.	1.5	33
148	Sequence analysis of tyrosine recombinases allows annotation of mobile genetic elements in prokaryotic genomes. <i>Molecular Systems Biology</i> , 2021, 17, e9880.	3.2	33
149	Bacterial Pleckstrin Homology Domains: A Prokaryotic Origin for the PH Domain. <i>Journal of Molecular Biology</i> , 2010, 396, 31-46.	2.0	32
150	Towards BioDBcore: a community-defined information specification for biological databases. <i>Nucleic Acids Research</i> , 2011, 39, D7-D10.	6.5	32
151	The rise and fall of supervised machine learning techniques. <i>Bioinformatics</i> , 2011, 27, 3331-3332.	1.8	32
152	The Importance of Biological Databases in Biological Discovery. <i>Current Protocols in Bioinformatics</i> , 2015, 50, 1.1.1-1.1.8.	25.8	32
153	Editorial. <i>Bioinformatics</i> , 2009, 25, 429-429.	1.8	31
154	Making your database available through Wikipedia: the pros and cons. <i>Nucleic Acids Research</i> , 2012, 40, D9-D12.	6.5	31
155	InterPro in 2011: new developments in the family and domain prediction database. <i>Nucleic Acids Research</i> , 2012, 40, 4725-4725.	6.5	31
156	Pepsin homologues in bacteria. <i>BMC Genomics</i> , 2009, 10, 437.	1.2	30
157	Towards BioDBcore: a community-defined information specification for biological databases. Database: the Journal of Biological Databases and Curation, 2011, 2011, baq027-baq027.	1.4	30
158	Origins of peptidases. <i>Biochimie</i> , 2019, 166, 4-18.	1.3	30
159	A member of the immunoglobulin superfamily in bacteriophage T4. <i>Virus Genes</i> , 1997, 14, 163-165.	0.7	28
160	HMM-based databases in InterPro. <i>Briefings in Bioinformatics</i> , 2002, 3, 236-245.	3.2	28
161	The TROVE module: a common element in Telomerase, Ro and Vault ribonucleoproteins. <i>BMC Bioinformatics</i> , 2003, 4, 49.	1.2	28
162	Tandem domain swapping: determinants of multidomain protein misfolding. <i>Current Opinion in Structural Biology</i> , 2019, 58, 97-104.	2.6	28

#	ARTICLE	IF	CITATIONS
163	Folding the unfoldable: using AlphaFold to explore spurious proteins. <i>Bioinformatics Advances</i> , 2022, 2, .	0.9	28
164	Fibronectin type III domains in yeast detected by a hidden Markov model. <i>Current Biology</i> , 1996, 6, 1544-1547.	1.8	27
165	The SGS3 protein involved in PTGS finds a family. <i>BMC Bioinformatics</i> , 2002, 3, 21.	1.2	27
166	Bioimage informatics: a new category in <i>Bioinformatics</i> . <i>Bioinformatics</i> , 2012, 28, 1057-1057.	1.8	27
167	Genome properties in 2019: a new companion database to InterPro for the inference of complete functional attributes. <i>Nucleic Acids Research</i> , 2019, 47, D564-D572.	6.5	27
168	Ten simple rules to make your computing more environmentally sustainable. <i>PLoS Computational Biology</i> , 2021, 17, e1009324.	1.5	27
169	Novel protein domains and motifs in the marine planctomycete <i>Rhodospirillum rubrum</i> . <i>FEMS Microbiology Letters</i> , 2004, 236, 333-340.	0.7	25
170	The yeast noncoding RNA interaction network. <i>Rna</i> , 2017, 23, 1479-1492.	1.6	25
171	Characterization of the nodulation plasmid encoded chemoreceptor gene <i>mcpG</i> from <i>Rhizobium leguminosarum</i> . <i>BMC Microbiology</i> , 2003, 3, 1.	1.3	24
172	Experience using web services for biological sequence analysis. <i>Briefings in Bioinformatics</i> , 2008, 9, 493-505.	3.2	24
173	The Characterisation of Three Types of Genes that Overlie Copy Number Variable Regions. <i>PLoS ONE</i> , 2011, 6, e14814.	1.1	24
174	Outline structures for the extracellular domains of the fibroblast growth factor receptors. <i>Nature Structural and Molecular Biology</i> , 1995, 2, 1068-1074.	3.6	23
175	Domain atrophy creates rare cases of functional partial protein domains. <i>Genome Biology</i> , 2015, 16, 88.	3.8	23
176	The challenge of increasing Pfam coverage of the human proteome. <i>Database: the Journal of Biological Databases and Curation</i> , 2013, 2013, bat023.	1.4	22
177	TreeGrafter: phylogenetic tree-based annotation of proteins with Gene Ontology terms and other annotations. <i>Bioinformatics</i> , 2019, 35, 518-520.	1.8	22
178	Acetylation of Surface Carbohydrates in Bacterial Pathogens Requires Coordinated Action of a Two-Domain Membrane-Bound Acyltransferase. <i>MBio</i> , 2020, 11, .	1.8	22
179	Identifying Protein Domains with the Pfam Database. <i>Current Protocols in Bioinformatics</i> , 2003, 1, 2.5.1-2.5.19.	25.8	20
180	CASP2 knowledge-based approach to distant homology recognition and fold prediction in CASP4. <i>Proteins: Structure, Function and Bioinformatics</i> , 2001, 45, 76-85.	1.5	19

#	ARTICLE	IF	CITATIONS
181	The use of structure information to increase alignment accuracy does not aid homologue detection with profile HMMs. <i>Bioinformatics</i> , 2002, 18, 1243-1249.	1.8	19
182	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> , 2016, 2016, baw027.	1.4	19
183	Predicted Input of Uncultured Fungal Symbionts to a Lichen Symbiosis from Metagenome-Assembled Genomes. <i>Genome Biology and Evolution</i> , 2021, 13, .	1.1	19
184	Searching databases to find protein domain organization. <i>Advances in Protein Chemistry</i> , 2000, 54, 137-157.	4.4	18
185	Enhanced protein domain discovery using taxonomy. <i>BMC Bioinformatics</i> , 2004, 5, 56.	1.2	18
186	Eleven quick tips to build a usable REST API for life sciences. <i>PLoS Computational Biology</i> , 2018, 14, e1006542.	1.5	18
187	Patterns of database citation in articles and patents indicate long-term scientific and industry value of biological data resources. <i>F1000Research</i> , 2016, 5, 160.	0.8	16
188	DASMI: exchanging, annotating and assessing molecular interaction data. <i>Bioinformatics</i> , 2009, 25, 1321-1328.	1.8	15
189	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> , 2019, 116, 26540-26548.	3.3	15
190	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> , 2021, 118, .	3.3	15
191	Novel protein domains and motifs in the marine planctomycete <i>Rhodopirellula baltica</i> . <i>FEMS Microbiology Letters</i> , 2004, 236, 333-340.	0.7	15
192	Key challenges for the creation and maintenance of specialist protein resources. <i>Proteins: Structure, Function and Bioinformatics</i> , 2015, 83, 1005-1013.	1.5	13
193	Genome3D: integrating a collaborative data pipeline to expand the depth and breadth of consensus protein structure annotation. <i>Nucleic Acids Research</i> , 2020, 48, D314-D319.	6.5	13
194	Pfam: the protein families database. , 2005, , .		12
195	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> , 2010, 19, 2131-2140.	3.1	12
196	Time to underpin Wikipedia wisdom. <i>Nature</i> , 2010, 468, 765-765.	13.7	12
197	Computer Tools and Algorithms for Origami Tessellation Design. , 2002, , 121-127.		11
198	Gene Unprediction with Spurio: A tool to identify spurious protein sequences. <i>F1000Research</i> , 2018, 7, 261.	0.8	11

#	ARTICLE	IF	CITATIONS
199	The Ensembl COVID-19 resource: ongoing integration of public SARS-CoV-2 data. <i>Nucleic Acids Research</i> , 2022, 50, D765-D770.	6.5	10
200	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> , 2010, 66, 1198-1204.	0.7	9
201	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> , 2010, 66, 1265-1273.	0.7	8
202	LUD, a new protein domain associated with lactate utilization. <i>BMC Bioinformatics</i> , 2013, 14, 341.	1.2	8
203	ISCB Computational Biology Wikipedia Competition. <i>PLoS Computational Biology</i> , 2013, 9, e1003242.	1.5	8
204	Creating a specialist protein resource network: a meeting report for the protein bioinformatics and community resources retreat: Figure 1.. Database: the Journal of Biological Databases and Curation, 2015, 2015, bav063.	1.4	8
205	ELIXIR pilot action: Marine metagenomics – towards a domain specific set of sustainable services. <i>F1000Research</i> , 2017, 6, 70.	0.8	8
206	A comparison of Pfam and MEROPS: two databases, one comprehensive, and one specialised. <i>BMC Bioinformatics</i> , 2003, 4, 17.	1.2	7
207	Discovery of fibrillar adhesins across bacterial species. <i>BMC Genomics</i> , 2021, 22, 550.	1.2	7
208	Structural genomics meets computational biology. <i>Bioinformatics</i> , 2006, 22, 2319-2319.	1.8	6
209	The structure of pyogenecin immunity protein, a novel bacteriocin-like immunity protein from <i>Streptococcus pyogenes</i> . <i>BMC Structural Biology</i> , 2009, 9, 75.	2.3	6
210	The complexity, challenges and benefits of comparing two transporter classification systems in TCDB and Pfam. <i>Briefings in Bioinformatics</i> , 2015, 16, 865-872.	3.2	6
211	TADOSS: computational estimation of tandem domain swap stability. <i>Bioinformatics</i> , 2019, 35, 2507-2508.	1.8	6
212	Exploring Non-Coding RNAs in RNAcentral. <i>Current Protocols in Bioinformatics</i> , 2020, 71, e104.	25.8	6
213	The SHOCT Domain: A Widespread Domain Under-Represented in Model Organisms. <i>PLoS ONE</i> , 2013, 8, e57848.	1.1	6
214	Large-Scale Discovery of Microbial Fibrillar Adhesins and Identification of Novel Members of Adhesive Domain Families. <i>Journal of Bacteriology</i> , 2022, 204, .	1.0	6
215	<i>Caenorhabditis elegans</i> AF4/FMR2 Family Homolog <i>affl-2</i> Regulates Heat-Shock-Induced Gene Expression. <i>Genetics</i> , 2020, 215, 1039-1054.	1.2	5
216	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> , 2020, 76, 124-134.	1.1	5

#	ARTICLE	IF	CITATIONS
217	Bioinformatics--The new home for protein sequence motifs. <i>Bioinformatics</i> , 2006, 22, 2-2.	1.8	4
218	Metagenomics approach for <i>Polymyxa betae</i> genome assembly enables comparative analysis towards deciphering the intracellular parasitic lifestyle of the plasmodiophorids. <i>Genomics</i> , 2022, 114, 9-22.	1.3	4
219	Ten Simple Rules for Chairing a Scientific Session. <i>PLoS Computational Biology</i> , 2009, 5, e1000517.	1.5	3
220	The YARHG Domain: An Extracellular Domain in Search of a Function. <i>PLoS ONE</i> , 2012, 7, e35575.	1.1	3
221	Two Pfam protein families characterized by a crystal structure of protein lpg2210 from <i>Legionella pneumophila</i> . <i>BMC Bioinformatics</i> , 2013, 14, 265.	1.2	3
222	New Leadership for Bioinformatics. <i>Bioinformatics</i> , 2004, 20, 1821-1821.	1.8	2
223	3DPatch: fast 3D structure visualization with residue conservation. <i>Bioinformatics</i> , 2019, 35, 332-334.	1.8	2
224	Software patents in Bioinformatics. <i>Bioinformatics</i> , 2006, 22, 1415-1415.	1.8	1
225	DATABASE, The Journal of Biological Databases and Curation, is now the official journal of the International Society for Biocuration. <i>Database: the Journal of Biological Databases and Curation</i> , 2013, 2013, bat077-bat077.	1.4	1
226	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> , 2014, 15, 75.	1.2	1
227	Rapid identification of novel protein families using similarity searches. <i>F1000Research</i> , 2018, 7, 1975.	0.8	1
228	An update from the Bioinformatics Editors. <i>Bioinformatics</i> , 2005, 21, 4319-4319.	1.8	0
229	Modelling structural rearrangements in proteins using Euclidean distance matrices. <i>F1000Research</i> , 0, 9, 728.	0.8	0