

Biological Databases

- A biological database is a computerized archive used to **store**, **organize** and **ease retrieval** of sequence data

- A database typically supports the following operations

- ✓ Retrieval
- ✓ Insertion
- ✓ Updating
- ✓ Deletion

Volume 5
Supplement 2
1976

ATLAS of
PROTEIN SEQUENCE
and STRUCTURE

Margaret O. Dayhoff



NATIONAL BIOMOLECULAR RESEARCH FOUNDATION
GEORGETOWN UNIVERSITY MEDICAL CENTER
WASHINGTON, D.C. 20007

Editorial and Review Committee: A. S. Laskowski, T. H. Miller,
A. M. Palmer, and R. J. Roberts
Associate Editor: C. E. Johnson
Production Manager: G. L. Johnson
Design: L. K. Johnson
Cover Design: G. L. Johnson
Typesetting: T. H. Miller, T. L. Johnson
Illustrations: T. H. Miller, T. L. Johnson

Number	Item	Page	Page	Page
1	CYT	1	449	1
2	WAT	2	450	2
3	ALB	3	451	3
4	LEU	4	452	4
5	TRP	5	453	5
6	VAL	6	454	6
7	ILE	7	455	7
8	PRO	8	456	8
9	ASP	9	457	9
10	GLU	10	458	10
11	SER	11	459	11
12	THR	12	460	12
13	TYR	13	461	13
14	PHEN	14	462	14
15	PHENYL	15	463	15
16	PHENYL-ALA	16	464	16
17	PHENYL-GLY	17	465	17
18	PHENYL-VAL	18	466	18
19	PHENYL-LEU	19	467	19
20	PHENYL-TRP	20	468	20
21	PHENYL-ILE	21	469	21
22	PHENYL-PRO	22	470	22
23	PHENYL-SER	23	471	23
24	PHENYL-TYR	24	472	24
25	PHENYL-PHE	25	473	25
26	PHENYL-PHE-ALA	26	474	26
27	PHENYL-PHE-GLY	27	475	27
28	PHENYL-PHE-VAL	28	476	28
29	PHENYL-PHE-LEU	29	477	29
30	PHENYL-PHE-TRP	30	478	30
31	PHENYL-PHE-ILE	31	479	31
32	PHENYL-PHE-PRO	32	480	32
33	PHENYL-PHE-SER	33	481	33
34	PHENYL-PHE-TYR	34	482	34
35	PHENYL-PHE-PHE	35	483	35
36	PHENYL-PHE-PHE-ALA	36	484	36
37	PHENYL-PHE-PHE-GLY	37	485	37
38	PHENYL-PHE-PHE-VAL	38	486	38
39	PHENYL-PHE-PHE-LEU	39	487	39
40	PHENYL-PHE-PHE-TRP	40	488	40
41	PHENYL-PHE-PHE-ILE	41	489	41
42	PHENYL-PHE-PHE-PRO	42	490	42
43	PHENYL-PHE-PHE-SER	43	491	43
44	PHENYL-PHE-PHE-TYR	44	492	44
45	PHENYL-PHE-PHE-PHE	45	493	45
46	PHENYL-PHE-PHE-PHE-ALA	46	494	46
47	PHENYL-PHE-PHE-PHE-GLY	47	495	47
48	PHENYL-PHE-PHE-PHE-VAL	48	496	48
49	PHENYL-PHE-PHE-PHE-LEU	49	497	49
50	PHENYL-PHE-PHE-PHE-TRP	50	498	50
51	PHENYL-PHE-PHE-PHE-ILE	51	499	51
52	PHENYL-PHE-PHE-PHE-PRO	52	500	52
53	PHENYL-PHE-PHE-PHE-SER	53	501	53
54	PHENYL-PHE-PHE-PHE-TYR	54	502	54
55	PHENYL-PHE-PHE-PHE-PHE	55	503	55
56	PHENYL-PHE-PHE-PHE-PHE-ALA	56	504	56
57	PHENYL-PHE-PHE-PHE-PHE-GLY	57	505	57
58	PHENYL-PHE-PHE-PHE-PHE-VAL	58	506	58
59	PHENYL-PHE-PHE-PHE-PHE-LEU	59	507	59
60	PHENYL-PHE-PHE-PHE-PHE-TRP	60	508	60
61	PHENYL-PHE-PHE-PHE-PHE-ILE	61	509	61
62	PHENYL-PHE-PHE-PHE-PHE-PRO	62	510	62
63	PHENYL-PHE-PHE-PHE-PHE-SER	63	511	63
64	PHENYL-PHE-PHE-PHE-PHE-TYR	64	512	64
65	PHENYL-PHE-PHE-PHE-PHE-PHE	65	513	65
66	PHENYL-PHE-PHE-PHE-PHE-PHE-ALA	66	514	66
67	PHENYL-PHE-PHE-PHE-PHE-PHE-GLY	67	515	67
68	PHENYL-PHE-PHE-PHE-PHE-PHE-VAL	68	516	68
69	PHENYL-PHE-PHE-PHE-PHE-PHE-LEU	69	517	69
70	PHENYL-PHE-PHE-PHE-PHE-PHE-TRP	70	518	70
71	PHENYL-PHE-PHE-PHE-PHE-PHE-ILE	71	519	71
72	PHENYL-PHE-PHE-PHE-PHE-PHE-PRO	72	520	72
73	PHENYL-PHE-PHE-PHE-PHE-PHE-SER	73	521	73
74	PHENYL-PHE-PHE-PHE-PHE-PHE-TYR	74	522	74
75	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE	75	523	75
76	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-ALA	76	524	76
77	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-GLY	77	525	77
78	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-VAL	78	526	78
79	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-LEU	79	527	79
80	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-TRP	80	528	80
81	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-ILE	81	529	81
82	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PRO	82	530	82
83	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-SER	83	531	83
84	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-TYR	84	532	84
85	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE	85	533	85
86	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-ALA	86	534	86
87	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-GLY	87	535	87
88	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-VAL	88	536	88
89	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-LEU	89	537	89
90	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-TRP	90	538	90
91	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-ILE	91	539	91
92	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PRO	92	540	92
93	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-SER	93	541	93
94	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-TYR	94	542	94
95	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE	95	543	95
96	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-ALA	96	544	96
97	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-GLY	97	545	97
98	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-VAL	98	546	98
99	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-LEU	99	547	99
100	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-TRP	100	548	100
101	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-ILE	101	549	101
102	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PRO	102	550	102
103	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-SER	103	551	103
104	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-TYR	104	552	104
105	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE	105	553	105
106	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-ALA	106	554	106
107	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-GLY	107	555	107
108	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-VAL	108	556	108
109	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-LEU	109	557	109
110	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-TRP	110	558	110
111	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-ILE	111	559	111
112	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PRO	112	560	112
113	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-SER	113	561	113
114	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-TYR	114	562	114
115	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE	115	563	115
116	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-ALA	116	564	116
117	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-GLY	117	565	117
118	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-VAL	118	566	118
119	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-LEU	119	567	119
120	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-TRP	120	568	120
121	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-ILE	121	569	121
122	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PRO	122	570	122
123	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-SER	123	571	123
124	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-TYR	124	572	124
125	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE	125	573	125
126	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-ALA	126	574	126
127	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-GLY	127	575	127
128	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-VAL	128	576	128
129	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-LEU	129	577	129
130	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-TRP	130	578	130
131	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-ILE	131	579	131
132	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PRO	132	580	132
133	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-SER	133	581	133
134	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-TYR	134	582	134
135	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE	135	583	135
136	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-ALA	136	584	136
137	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-GLY	137	585	137
138	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-VAL	138	586	138
139	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-LEU	139	587	139
140	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-TRP	140	588	140
141	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-ILE	141	589	141
142	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PRO	142	590	142
143	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-SER	143	591	143
144	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-TYR	144	592	144
145	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE	145	593	145
146	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-ALA	146	594	146
147	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-GLY	147	595	147
148	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-VAL	148	596	148
149	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-LEU	149	597	149
150	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-TRP	150	598	150
151	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-ILE	151	599	151
152	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PRO	152	600	152
153	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-SER	153	601	153
154	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-TYR	154	602	154
155	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE	155	603	155
156	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-ALA	156	604	156
157	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-GLY	157	605	157
158	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-VAL	158	606	158
159	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-LEU	159	607	159
160	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-TRP	160	608	160
161	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-ILE	161	609	161
162	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PRO	162	610	162
163	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-SER	163	611	163
164	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-TYR	164	612	164
165	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE	165	613	165
166	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-ALA	166	614	166
167	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-GLY	167	615	167
168	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-VAL	168	616	168
169	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-LEU	169	617	169
170	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-TRP	170	618	170
171	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-ILE	171	619	171
172	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PRO	172	620	172
173	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-SER	173	621	173
174	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-TYR	174	622	174
175	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE	175	623	175
176	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-ALA	176	624	176
177	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-GLY	177	625	177
178	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-VAL	178	626	178
179	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-LEU	179	627	179
180	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-TRP	180	628	180
181	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-ILE	181	629	181
182	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PRO	182	630	182
183	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-SER	183	631	183
184	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-TYR	184	632	184
185	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE	185	633	185
186	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-ALA	186	634	186
187	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-GLY	187	635	187
188	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-VAL	188	636	188
189	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-LEU	189	637	189
190	PHENYL-PHE-PHE-PHE-PHE-PHE-PHE-PHE-PHE-TRP	190	638	190
1				

First biological database

- Atlas of protein sequence and structure (1965)
- Margaret Oakley Dayhoff (Columbia University, US)
- First release: 65 protein sequences
- Regular updates



7 7 7 8 8 8 8 8 9 9 9 9 9 10 10

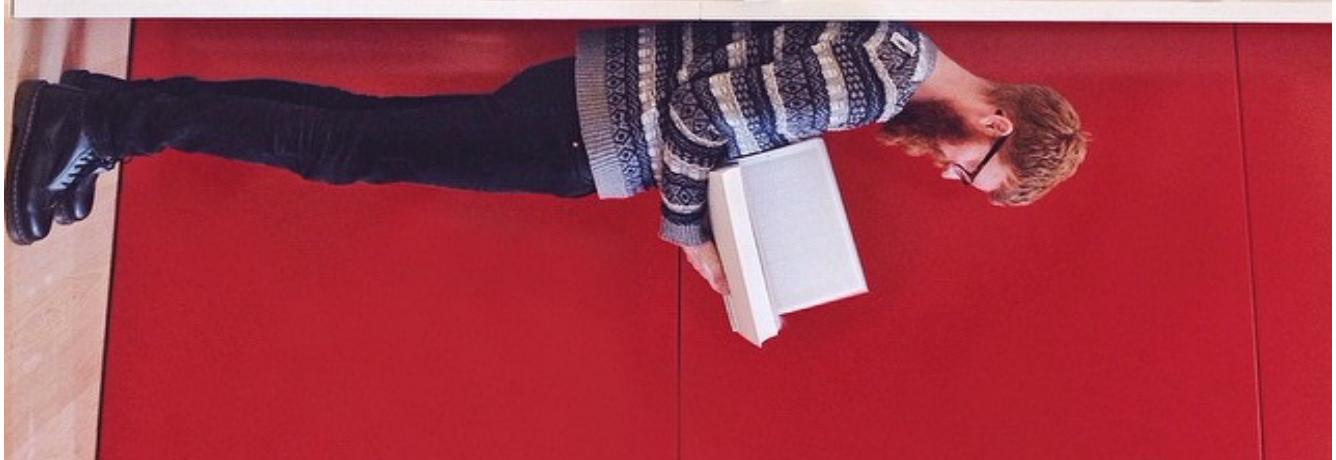
10 10 10 11 11 11 11 12 12 12 13 13



13 14 14 14 15 15 15 16 16 17 17 17

18 18 18 19 19 20 20 21 21 22 22 X X X

X X Y Y



Print out of the human
genome

Wellcome Collection
London, UK

First digital biological database

- EMBL Data Library (1982)
- European Molecular Biology Laboratory (EMBL), Heidelberg, Germany
- First release: 568 entries, ~500,000 bp
- Distributed on magnetic tapes
- From 1987 via EMBL File Server
- Predecessor of EMBL Bank/ENA



Biological databases

- I. A database can be thought as a large table where row represents record and columns represents fields.
- II. The organization of records allows for querying on the data to retrieve information from the databases
- III. An ideal biological database has fields as shown below

Accession number	Name	Length	Sequence	Taxonomy	Reference
NR235462.1	MTGA	268	ACTTGCA...	E.coli	A.Kelly et. al
NR235463.1	HKY	350	TGAGTAA...	E.coli	J.Jone et. al
NR235464.1	THY	289	TGACGT...	S.Aurius	K.Moy et. al

Importance of Biological Databases

- Means to handle and share large volumes of biological data.
- Biological databases represent an invaluable resource to support biological research. In some cases no need to sequence.
- Link knowledge to sequenced data – [Cross referencing](#), most sequences databases are cross linked to load of biological information: gene/protein – information, Structural information, pathways and biological processes literature.

Types of Biological Databases

1. **Primary databases:** hold raw sequenced data
 - GenBank
 - EMBL (European Molecular Biology Laboratory)
 - DDBJ (DNA Data Bank of Japan)
 - PDB (Protein Data Bank)
2. **Secondary databases:** they are curated and annotated
 - Swiss-Prot – detailed annotation if proteins
 - RefSeq – non-redundant, curated sequenced data
3. **Specialized databases:** these focus on data of specific research interest
 - VectorBase
 - OMIM: catalog of human genes, genetic disorders

Where to look for Biological databases

- ✓ Search Engines (google)
- ✓ Journals related to bioinformatics
 - Nucleic Acid research NAR online database issue
- ✓ Websites like; www.expasy.ch

Bioinformatics Databases

AATDB, AceDb, ACUTS, ADB, AFDB, AGIS, AMSdb, ARR, AsDb, BBDB, BCGD, Beanref, Biolimage, BioMagResBank, BIOMDB, BLOCKS, BovGBASE, BOVMAP, BSORF, BTkbase, CANSITE, CarbBank, CARBHYD, CATH, CAZY, CCDC, CD4OLbase, CGAP, ChickGBASE, Colibri, COPE, CottonDB, CSNDB, CUTG, CyanoBase, dbCFc, dbEST, dbSTS, DDBJ, DGP, DictyDb, Picty_cDb, DIP, DOGS, DOMO, DPD, DPInteract, ECDC, ECGC, EC02DBASE, EcoCyc, EcoGene, EMBL, EMD db, ENZYME, EPD, EpoDb, ESTHER, FlyBase, FlyView, GCRDB, GDB, GENATLAS, Genbank, GeneCards, Genlilesne, GenLink, GENOTK, GenProtEC, GIFTS, GPCRDB, GRAP, GRBase, gRNAsdb, GRR, GSDB, HAEMB, HAMSTERS, HEART-2DPAGE, HEXAdb, HGMD, HIDB, HIDC, HIVdb, HotMolecBase, HOVERGEN, HPDB, HSC-2DPAGE, ICN, ICTVDB, IL2RGbase, IMGT, Kabat, KDNA, KEGG, KloTho, LGIC, MAD, MaizeDb, MDB, Medline, Mendel, MEROPS, MGDB, MGI, MHCPEP5, Micado, MitoDat, MITOMAP, MJDB, MmtDB, Mol-R-U\$, MPDB, MRR, MutBase, MycDB, NDB, NRSub, O-lycBase, OMIA, OMIM, OPD, ORDB, OWL, PAHdb, PatBase, PDB, PDD, Pfam, PhosphoBase, PigBASE, PIR, PKR, PMD, PPDB, PRESAGE, PRINTS, ProDom, Prolysis, PROSITE, PROTOMAP, RatMAP, RDP, REBASE, RGP, SBASE, SCOP, SeqAnaiRef, SGD, SGP, SheepMap, Soybase, SPAD, SRNA db, SRPDB, STACK, StyGene, Sub2D, SubList, SWISS-2DPAGE, SWISS-3DIMAGE, SWISS-MODEL Repository, SWISS-PROT, TelDB, TGN, tmRDB, TOPS, TRANSFAC, TRR, UniGene, URNADB, V BASE, VDRR, VectorDB, WDCM, WIT, WormPep, etc ..!!!

Major Biological databases

The most popular bioinformatics databases focus on:

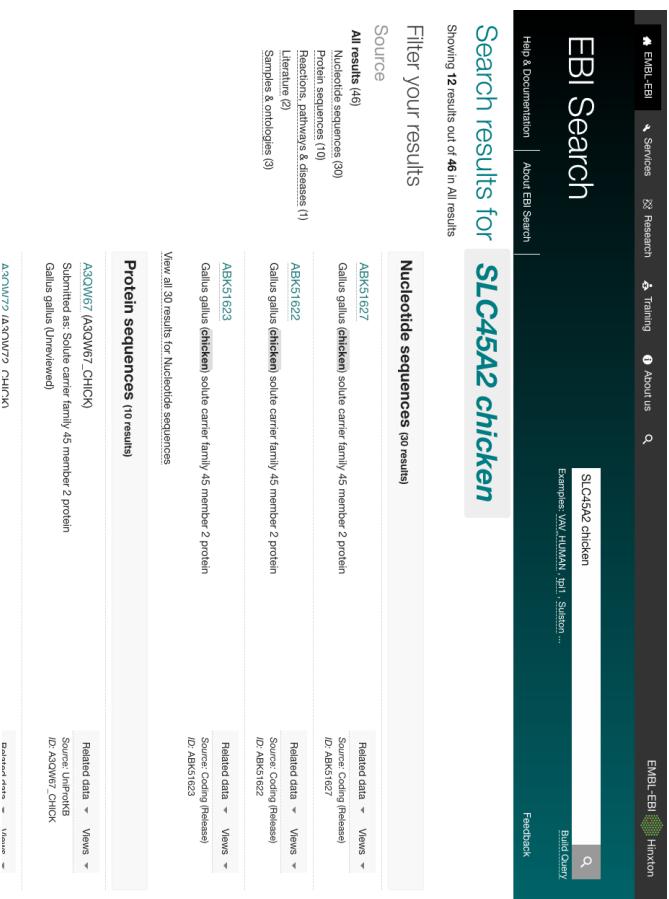
- Biomolecular sequence (e.g. GenBank, UniProt)
- Biomolecular structure (e.g PDB)
- Vertebrate genomes (e.g. Ensemble)
- Small molecules (e.g. PubChem)
- Biomedical literature (e.g. PubMed)

Other popular “boutique” databases

- Classifying protein families, domains and motifs (e.g. PFAM, PROSITE)
- Metabolic pathways (e.g. Kegg)
- Specific organisms (e.g. WormBase, FlyBase)
- Specific proteins of biomedical importance (e.g KinaseDB)

Key Resources (Searching multiple databases)

Both EBI and NCBI provide the option to search all their database resources in one go:



The screenshot shows the EBI Search interface. The search term "SLC45A2 chicken" is entered in the search bar. Below the search bar, it says "Showing 12 results out of 46 in All results". The results are filtered by "Nucleotide sequences (30 results)". The first result is "ABK51627 Gallus gallus (chicken) solute carrier family 45 member 2 protein". Below the result, there are links for "Related data", "Views", and "Source: Coding (Release)".

Search results for SLC45A2 chicken

Showing 12 results out of 46 in All results

Filter your results

Nucleotide sequences (30 results)

All results (46)

Nucleotide sequences (30)

Protein sequences (10)

Reactions, pathways & diseases (1)

Literature (2)

Samples & ontologies (3)

ABK51627 Gallus gallus (chicken) solute carrier family 45 member 2 protein

ABK51622 Gallus gallus (chicken) solute carrier family 45 member 2 protein

ABK51623 Gallus gallus (chicken) solute carrier family 45 member 2 protein

A3QW67 (ASQW67_CHICK) Submitted as: Solute carrier family 45 member 2 protein

A3QW67 (ASQW67_CHICK) Submitted as: Solute carrier family 45 member 2 protein

Feedback

Build Query

Help

Sign in to NCBI

EMBL-EBI Hinxton

NCBI Resources How To

Results found in 28 databases for "SLC45A2 chicken"

Literature

Books	0	books and reports
MeSH	0	ontology used for PubMed indexing
NLM Catalog	547	books, journals and more in the NLM Collections
PubMed	2	scientific & medical abstracts/citations
PubMed Central	48	full-text journal articles

Health

ClinVar	0	human variations of clinical significance
dbGap	219	genotype/pheno-type interaction studies
GTR	0	genetic testing registry
MedGen	1	medical genetics literature and links
OMIM	0	online mendelian inheritance in man
PubMed Health	0	clinical effectiveness, disease and drug reports

Genomes

Assembly	17	genome assembly information
BioProject	0	biological projects providing data to NCBI
BioSample	18,528	descriptions of biological source materials
Clone	10	genomic and cDNA clones
dbVar	6	genomic structural variation studies
Genome	1	genome sequencing projects by organism
GSS	201,802	genome survey sequences
Nucleotide	51	DNA and RNA sequences
Probe	6	sequence-based probes and primers
SNP	232	short genetic variations
STRA	14,055	high-throughput DNA and RNA sequence read archive
Taxonomy	1	taxonomic classification and nomenclature catalog

Proteins

Conserved Domains	61	conserved protein domains
Protein	14	protein sequences
Protein Clusters	13	sequence similarity-based protein clusters
Structure	1,764	experimentally-determined biomolecular structures

Chemicals

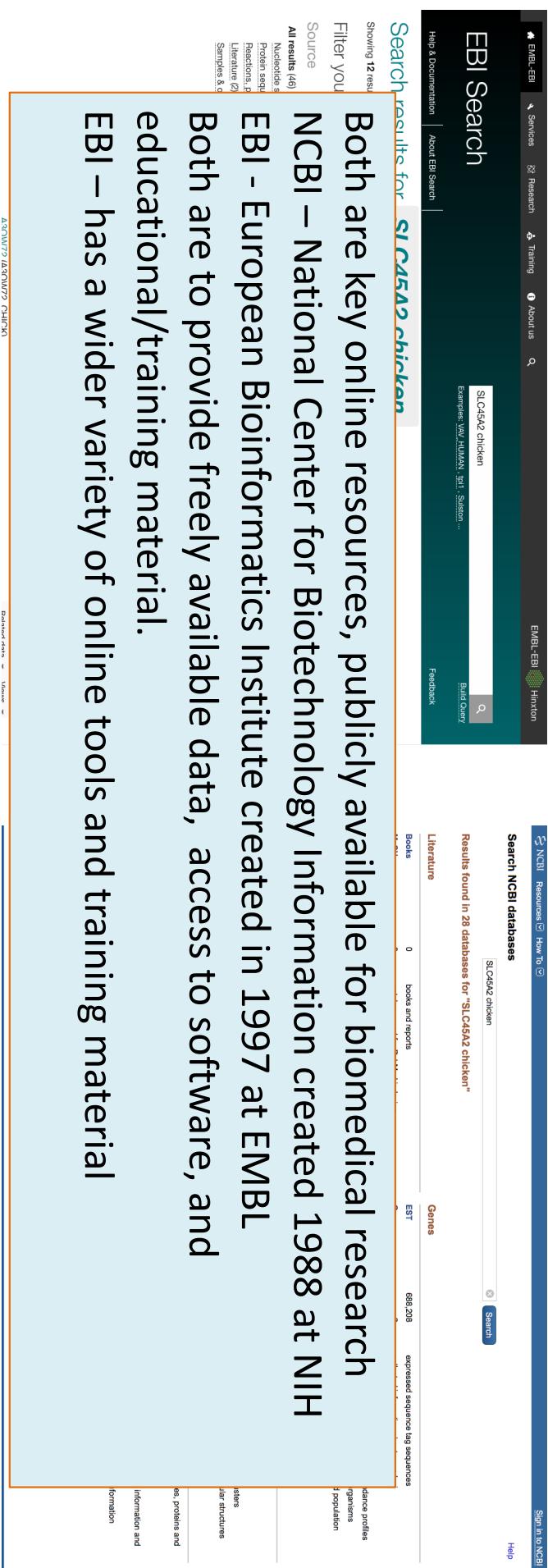
BioSystems	66	molecular pathways with links to genes, proteins and chemicals
PubChem BioAssay	1,656	bioactive screening studies
PubChem Compound	43	chemical information with structures, information and links

<http://www.ebi.ac.uk/ebisearch>

<http://ncbi.nlm.nih.gov>

Key Resources (Searching multiple databases)

Both EBI and NCBI provide the option to search all their database resources in one go:



The screenshot shows the EBI Search interface. At the top, there are links for Services, Research, Training, About us, and a search bar. Below the search bar, it says "Showing 12 results" and "Filter you". A sidebar on the left lists "Source" categories: All results (46), Nucleotides, Protein seqs, Reactions, P, Literature (2), Samples & c. The main area displays search results for "SLC45A2 chicken". It shows 28 databases found, with 0 books and reports. The results are categorized under "Genes" (698,208 entries) and "expressed sequence tag sequences" (1,121 entries). Below the results, there are links for "dance profiles", "organisms", "population", "sites", "jar structures", "es", "proteins and", "information and", and "formation".

Search results for **SLC45A2 chicken**

Showing 12 results

Filter you

Source

All results (46)

Nucleotides

Protein seqs

Reactions, P

Literature (2)

Samples & c

Results found in 28 databases for "SLC45A2 chicken"

Books 0 books and reports

Genes 698,208 expressed sequence tag sequences

Dance profiles
Organisms
Population
Sites
Jar structures
Es, proteins and
Information and
Formation

<http://www.ebi.ac.uk/ebisearch>

<http://ncbi.nlm.nih.gov>

UniProt: Protein sequence database

UniProt is a comprehensive, high quality resource of protein sequence and functional information.

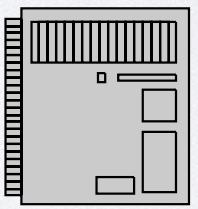
1. **UniProtKB** (Knowledgebase)
Contains Swiss-Prot and TrEMBL (hand and automatically curated respectively)
2. **UniREF** (Reference Clusters)
Filtered version of UniProtKB at various levels of sequence identity. UniRef90 has sequences with a maximum of 90% sequence identity to each other
3. **Uniparc** (Archive)
with database cross-references to source
4. **UniMES** (Metagenomic and Environmental Sequences)

UniProt



UniProtKB/TrEMBL

Redundant, automatically annotated - unreviewed



UniProtKB/Swiss-Prot

Non-redundant, high-quality manual annotation - reviewed



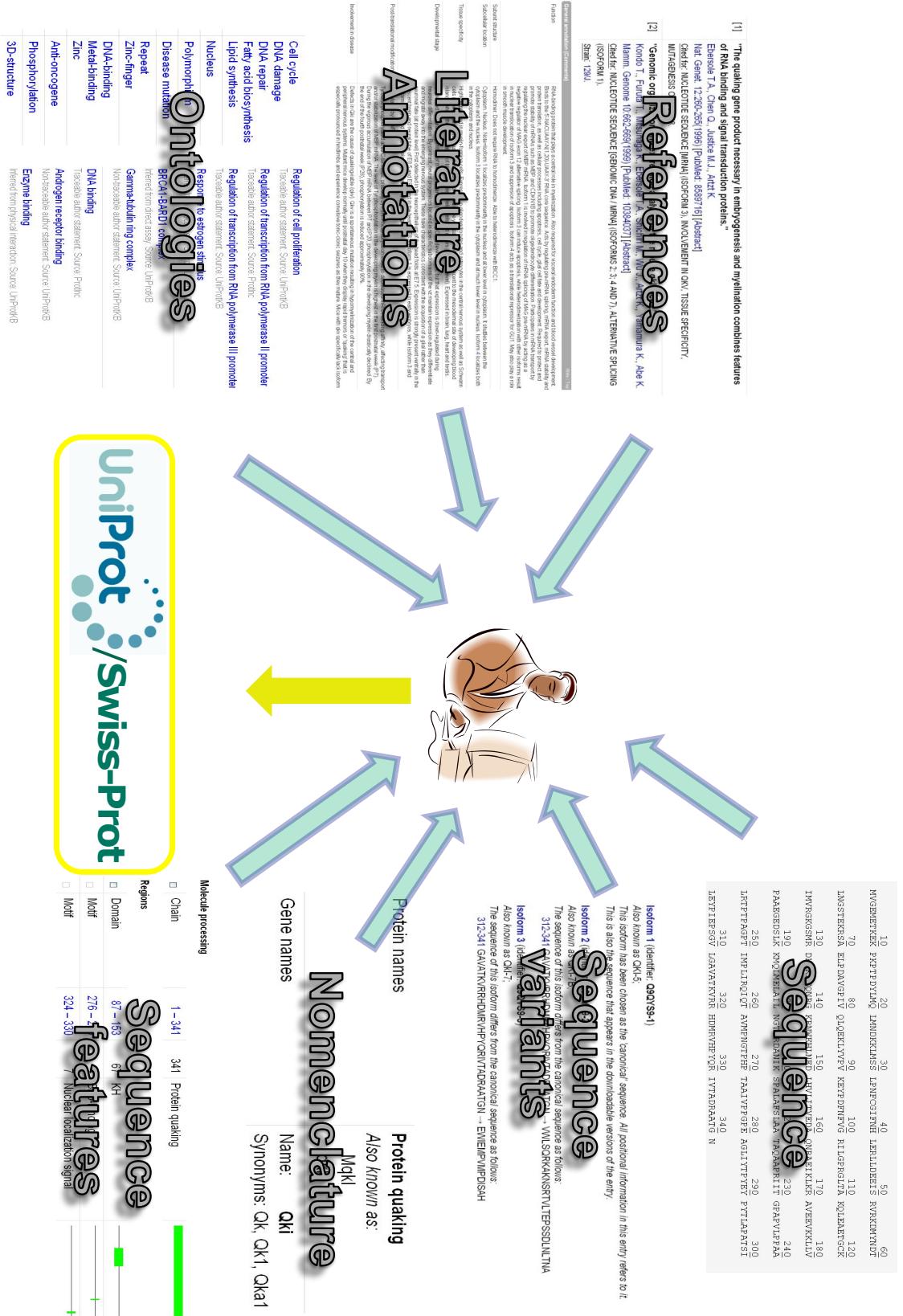
★ Unreviewed, UniProtKB/TrEMBL **Q9N0H9** (Q9N0H9_EQUAS)



★ Reviewed, UniProtKB/Swiss-Prot **P38398** (BRCA1_HUMAN)

Indicators of which part of UniProt an entry belongs to include the color of the stars and the ID

What information is added to a UniProt/Swiss-Prot entry



What is GenBank

GenBank is NCBI's primary nucleotide only sequence database

- Archival in nature- *reflects the state of knowledge at the time of submission*
- Subjective – *reflects the submitter point of view*
- Redundant – *can have many copies of the same nucleotide sequence*

GenBank is actually three collaborating international databases from the US, Japan and Europe

- GenBank – (US)
- DNA Database of Japan (DDBJ)
- European Nucleotide Archive (ENA)

The underlying raw DNA sequence are identical just different views and ways to navigate through the data

Accession Numbers

- ✓ Stable ways of identifying GenBank Entries.
- ✓ No biological meaning.
- ✓ Originally an uppercase letter followed by 5 digits **U00002**
- ✓ Now two uppercase letters followed by six digits **BC037153**
- ✓ Version of entry added later as a decimal
BC037153.1

Accession Numbers

X02775

GenBank genomic DNA sequence

NT_030059

Genomic contig

N91759.1

An expresses sequence tag

RNA

NM_006744

RefSeq DNA sequence (from a transcripts)

NP_007635

RefSeq protein

AAC02945

GenBank protein

Q28369

UniProtKB/SwissProt protein

1KT7

Protein Data Bank structure record

PMID:12205585

PubMed IDs identify articles at NCBI

Literature

Sequence formats

- ✓ This is the required arrangement of characters symbols and keywords in a sequence record.
- ✓ There many different sequence formats for the purpose of database integration and organizing sequenced data
- ✓ When considering a format for retrieval:
 - **What is easy to parse**
 - **What format do the tools need**
 - **What information is needed**

FASTA format

- Used by fasta tools
- Comment line “>” then sequence data

```
>gi|8547324|gb|AF271385.1| Fasciola hepatica cathepsin L mRNA, complete cds
GGCACAACAAATGAGATTGTAATCCTAACCTACTCGTCGGAGTGTGGCCTCAAATGACGATTGT
GGCATCAATGGAAGGAAATTACAATAAGAATACAATGGAGCTGACGATGCCACAGGAAATATTTG
GGAACAAATGTGAAACATATCCAAGAACACACCTGCCACGATCTCGGTCTCGTCACCTACAGTTG
GGATTGAACCAATTCAACCGATATGACATTCGAGGAATTCAAAGCCAAATATCTAACAGAAATGCCACCGG
CGTCTGAGTTACTCTCACACGGTACACGGTACACGGTACACGGTACACGGTACACGGTACACGGTACACGGT
GCGTGAATCCGGTTATGTGACGGAGGTGAAAGATCAGGGAGGCTGTGGTTCTGGCTTCTCAACA
ACAGGTGCTATGGAAGGACAGTATGAAAACCAAGAACACTAGTATTCTTGAAACAAACACTGG
TCGATTGTAGCCGTGATTGGCAATTATGGTTGTAATGGTGGACTAATGGAATGCATACGAATTATT
GAAACGATTGGATTGGAACCGAGCTTCTTACAGGGCTGTTGGAAGGACAGTGTGGATACAAC
GAGCAGTGGAGTTGCCAAAGTGACTGGCTACTACGGTACATTCTGGAGATGGAGTAGAATTGCAA
ATCTAGTCGGCGAAGGACCTGCTGCCGTGCTTGGATGTGGAGTCAGACTCATGATGTACAGGAG
TGGTATTTCAGAGCCAACTGTTCACCGGATCGTTGAACCATGGAGTGTGGCTGTGGTTATGGA
ATACAGGATGGTACTGACTACTGGATGTGAAACACAGTGGGAACGTGGGGTGGAGGACGGTTACA
TTCGAATGGTTAGGAAAGAGGTAAACATGTGTTGAATTGCTTCTGCCAGTGTCCCCGATGGTGGCACA
ATTCCGTGA|
```

GenBank format

- Flat file format used by the GenBank
- Has annotation, author, version etc

Cucumber mosaic virus RNA2 segment, complete sequence

GenBank: AF033667.1

[FASTA](#) [Graphics](#)

[Go to:](#)

LOCUS AF033667 3047 bp RNA linear VRL 20-NOV-1997
DEFINITION Cucumber mosaic virus RNA2 segment, complete sequence.
ACCESSION AF033667
VERSION AF033667.1
KEYWORDS .
SOURCE Cucumber mosaic virus (cucumber mosaic cucumovirus)
ORGANISM [Cucumber mosaic virus](#)
Viruses; ssRNA viruses; ssRNA positive-strand viruses, no DNA
stage; Bromoviridae; Cucumovirus.
REFERENCE 1 (bases 1 to 3047)
AUTHORS Kim,S.H. and Park,Y.I.
TITLE Direct Submission
JOURNAL Submitted (09-NOV-1997) Graduate School of Biotechnology, Korea
University, Anam-dong 5-ga. Sungbook-ku, Seoul 136-701, Korea
FEATURES Location/Qualifiers
source
1..3047

```
/organism="Cucumber mosaic virus"
/mol_type="genomic RNA"
/strain="As"
/db_xref="taxon:12305"
```

FEATURES Location/Qualifiers

source

1..3047

/organism="Cucumber mosaic virus"

/mol_type="genomic RNA"

/strain="As"

/db_xref="taxon:12305"

/chromosome="RNA2"

gene

79..2655

ORIGIN

1 gtttattttac aagagcgtac ggttcaatcc ctgcctcccc tggaaaactc cctagacttt
61 ctagttctt tctttcttat gggtttccc gcccccgctt ttccacttagc caatctttg
121 aacgggttgtt acgggtgtca cactcccgag gaagtggaa ggggtgcgtc tgacaaacgc
181 gacggaggctg ctggggcttg togttaattac aggccttac ccgtctgtgg tgcacggag
241 agtgtcccgag aggtatgcga ttccctccga acttcgttgcg gagctccctc tgaagagggt
301 tctgttagagt ttgttactta ttgttgcgtaa gattaaaccttgg aaaatctgtt tgatggactc
361 ttgtgtcgctt ttgttagacgt ggtcaaaccc atgcgtatcg gacaactgtt gtgcctcg

GNEWAEGSF"

FEATURES

source

Location/Qualifiers
1..3047

/organism="Cucumber mosaic virus"

/mol_type="genomic RNA"

/strain="As"

/db_xref="taxon:12305"

/chromosome="RNA2"

gene
79..2655

/gene="2a"
79..2655

CDS
.....
gene
79..2655

/function="RNA-dependent RNA polymerase"
/note="2a protein"

/codon_start=1

/product="RNA replicase"

/protein_id="AAB86650.1"

/translation="MAFPAPAFSLANLINGSYGVDTPEEVERVRSEQRDEAAACRNY
RPLPAVDVSESVFEDAHSRLTPDGAPSEEVSVFVTYGAEDYLEKSDDELIVAFETMV
KPMRIGQIWCPCAPNKSFISSIAMARALLAPRTSHRTMKCFEDLVAAITYTKSDFYD

- Features section contains annotations including the conceptual translation of the nucleotide sequence.
- The actual sequence starts with the word origin

SYKELYYSDRROCELINSFSCVELRIERSSTSKORERKDIERRRNDKRRRPTGSYGG
GEEMETKVQSESTGTRSQKSOREGAFKQSQAVPLTILSSRWFGTDTDRDVPCEHHGGTV
RV"

gene
2414..2749
/gene="2b"
CDS
.....
gene
2414..2749
/gene="2b"
/function="host-specific virus accumulation"
/note="2b protein"
/codon_start=1
/product="long-distance movement protein"
/protein_id="AAB86651.1"
/translation="MELNEGAMTNVELQLARMWEAKQRSSHKQNRRERRGHKSPER
ARSNLRLFRFLPFYQVDGSELIMYHHVNMGVLISESEAPCFTLPAEEDHDFDDWFA
GNEWAEGSF"

ORIGIN
1 gtttattttac aagagcgtac ggttcaatcc ctgcctcccc tggaaaactc cctagacttt
61 cttagttctt tctttcttat gggtttccc gcccccgctt ttctacttagc caatctttg
121 aacggtagtt acgggtcgaa cactcccgag gaagtggAAC ggggtgcgtc tgacacgc
181 gacggaggctg ctggggcttg togttaattac aggccttac ccgtgtgg tgcacggag
241 agtgtcccgag aggtatgcga ttccctccga acttcgttgcg gagctccctc tgaagagggtg
301 tctgttagagt ttgttactta ttgtgtgtgaa gattaaatcttg aaaaatctgtg tcatgagctc
361 ttgtgtcgctt ttgttagatgtt ggtcaaaccc atgcgtatcg gacaactgtg gtgcctgcg

Sequence formats

- Sequence are stored in databases or in files as **simple text** (ASCII text)
- Microsoft Word format is **not** a sequence format
- Save sequence files as **text .txt file !!!**
- Use **text editors** like note-pad, text-pad to open such files

DATABASE VIGNETTE

Perhaps you have heard of BRCA1 gene that is Breast cancer type one and you want to find out more about it how would you do this

- Google?
- Library?
- Bioinformatics databases at NCBI and EBI

<https://www.ncbi.nlm.nih.gov/>



NCBI Home

Resource List (A-Z)

All Resources

Chemicals & Bioassays

Data & Software

DNA & RNA

Domains & Structures

Genes & Expression

Genetics & Medicine

Genomes & Maps

Homology

Literature

Proteins

Sequence Analysis

Develop

Analyze

Research

NCBI Announcements

Taxonomy
Training & Tutorials
Variation

Use NCBI APIs and code
libraries to build
applications



The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

[About the NCBI](#) | [Mission](#) | [Organization](#) | [NCBI News](#) | [Blog](#)

Popular Resources

PubMed

Bookshelf

PubMed Central

PubMed Health

BLAST

Nucleotide

Genome

SNP

Gene

Protein

PubChem

In this minute NCBI staff will show you:
April 26th NCBI Minute: Medical Genetics
Summaries on the NCBI Bookshelf - a

May 10th NCBI Minute: How to Locate
and Use Human Genomes and
Annotations from NCBI

01 May 2017

Search NCBI databases

BRACA1

Search

Results found in 8 databases for "BRACA1"

Literature			
		Genes	
		EST	Gene
Books	1	books and reports	0 expressed sequence tag sequences
MeSH	0	ontology used for PubMed indexing	2 collected information about gene loci
NLM Catalog	0	books, journals and more in the NLM Collections	0 functional genomics studies
PubMed	4	scientific & medical abstracts/citations	0 gene expression and molecular abundance profiles
PubMed Central	59	full-text journal articles	0 homologous gene sets for selected organisms
Health			
ClinVar	0	human variations of clinical significance	0 sequence sets from phylogenetic and population studies
dbGaP	0	genotype/phenotype interaction studies	3 clusters of expressed transcripts
GTR	0	genetic testing registry	
MedGen	0	medical genetics literature and links	
OMIM	1	online mendelian inheritance in man	
PubMed Health	0	clinical effectiveness, disease and drug reports	
Genomes			
Structure	0	structures	
Proteins			
Conserved Domains	0	conserved protein domains	
Protein	3	protein sequences	
Protein Clusters	0	sequence similarity-based protein clusters	
		experimentally-determined biomolecular structures	

Gene

Gene

Create RSS Create alert Advanced

Search

Help

Gene sources

Genomic

Categories

Alternatively spliced

Annotated genes

Protein-coding

Send to: ▾

Filters: [Manage Filters](#)

Results by taxon

[Top Organisms](#) [Tree][Homo sapiens \(1\)](#)

Thalassiosira pseudonana

CCMP1335 (1)

Sequence content

CCDS

Ensembl

RefSeq

RefSeqGene

Status

Current

Chromosome locations

more...

clear

Name/Gene ID	Description	Location	Aliases	MIM
<input type="checkbox"/> BRIP1 ID: 83990	BRCA1 interacting protein C-terminal helicase 1 [<i>Homo sapiens</i> (human)]	Chromosome 17, NC_000017.11 (61679186..61864120, complement)	BACH1, FANCJ, OF	605882
<input type="checkbox"/> BRAC2 ID: 7448960	breast cancer 2 early onset [<i>Thalassiosira pseudonana</i> CCMP1335]	Chromosome 7, NC_012070.1 (427471..428441)	THAPS_263089	

Find related data

Database: [Select](#)[Find items](#)

Search details

BRACA1[All Fields] AND alive[prop]

[Clear all](#)[Show additional filters](#)

Full Report ▾

Showing Current items.

Send to: ▾

Hide sidebar >>

BRIP1 BRCA1 interacting protein C-terminal helicase 1 [*Homo sapiens* (human)]

Gene ID: 83990, updated on 1-May-2017

Summary



?

Table of contents

[Summary](#)[Genomic context](#)[Genomic regions, transcripts, and products](#)[Bibliography](#)[Phenotypes](#)[Variation](#)[Pathways from BioSystems](#)[Interactions](#)[General gene information](#)[Markers, Clone Names, Homology, Gene Ontology](#)[General protein information](#)[NCBI Reference Sequences \(RefSeq\)](#)[Related sequences](#)[Additional links](#)[Locus-specific Databases](#)**Orthologs**[mouse](#) [all](#)

Official Symbol BRIP1 [provided by HGNC](#)

Primary source HGNC:HGNC:20473 [provided by HGNC](#)

See related Ensembl:ENSG00000136492 MIM:605882; Vega:OTTHUMMG00000179233

Gene type protein coding

RefSeq status REVIEWED

Organism [Homo sapiens](#)

Lineage Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo

Also known as OF; BACH1; FANCJ

Summary The protein encoded by this gene is a member of the RecQL DEAH helicase family and interacts with the BRCT repeats of breast cancer, type 1 (BRCA1). The bound complex is important in the normal double-strand break repair function of breast cancer, type 1 (BRCA1). This gene may be a target of germline cancer-inducing mutations. [provided by RefSeq, Jul 2008]

Full Report ▾

Showing Current items.

BRIP1 **BRCA1** interacting protein C-terminal helicase 1

Gene ID: 83990, updated on 1-May-2017

What chromosome location and What gene are in the vicinity?

Send to: ▾

Hide sidebar >>

[Table of contents](#)
[Summary](#)[Genomic context](#)
[Bibliography](#)[Summary](#)[Phenotypes](#)
[Variation](#)
[Pathways from BioSystems](#)
[Interactions](#)
[General gene information](#)
[Markers, Clone Names, Homology, Gene Ontology](#)

Official Symbol BRIP1 [provided by HGNC](#)
Primary source HGNC:HGNC:20473
See related Ensembl:ENSG0000136492 MIM:605882; Vega:OTTHUMMG00000179233

Gene type protein coding
RefSeq status REVIEWED

Organism Homo sapiens

Lineage Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo

Also known as

Summary

The protein encoded by this gene is a member of the RecQL DEAH helicase family and interacts with the BRCT repeats of breast cancer, type 1 (BRCA1). The bound complex is important in the normal double-strand break repair function of breast cancer, type 1 (BRCA1). This gene may be a target of germline cancer-inducing mutations. [provided by RefSeq, Jul 2008]

Orthologs

[mouse](#) [all](#)

Locus-specific Databases

► Genomic context

Location: 17q23.2

Exon count: 25

See BRIP1 in [Genome Data Viewer Map Viewer](#)



Genome Browsers

[Genome Data Viewer](#)

[Map Viewer](#)

[Variation Viewer \(GRCh37.p13\)](#)

[Variation Viewer \(GRCh38\)](#)

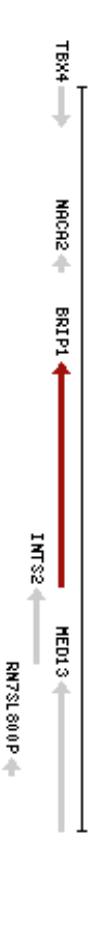
[1000 Genomes Browser \(GRCh37.p13\)](#)

[Ensembl](#)

[UCSC](#)

Annotation release	Status	Assembly	Chr	Location
108	current	GRCh38.p7 (GCF_000001405.33)	17	NC_000017.11 (61679186..61864120, complement)
105	previous assembly	GRCh37.p13 (GCF_000001405.25)	17	NC_000017.10 (59756547..59940920, complement)

Chromosome 17 - NC_000017.11



Related Information

[Order cDNA clone](#)

[3D structures](#)

[BioAssay by Target \(List\)](#)



[BioAssay by Target \(Summary\)](#)

[BioAssay, by Gene target](#)

[BioAssays, RNAi Target, Active](#)

[Go to reference sequence details](#)

Genomic Sequence: [NC_000017.11 Chromosome 17 Reference GRCh38.p7 Primary Assembly](#)



[Full Report](#) ▾

1 Showing Current items.

BRIP1 BRCA1 inter

Gene ID: 83990, updated on

[Summary](#)

What 'molecular functions',

'biological processes' and 'cellular component' information is available?

to: ▾

[Hide sidebar >>](#)[Table of contents](#)[Summary](#)[Genomic context](#)[Genomic regions, transcripts, and products](#)[Bibliography](#)[Phenotypes](#)[Variation](#)[Pathways from BioSystems](#)[Interactions](#)

General gene information

[Markers, Clone Names, Homology, Gene Ontology](#)[General protein information](#)[NCBI Reference Sequences \(RefSeq\)](#)[Related sequences](#)[Additional links](#)[Locus-specific Databases](#)[Orthologs](#)[mouse](#) [all](#)

Official Symbol	BRIP1 <small>provided by HGNC</small>
Primary source	HGNC:HGNC:20473
See related	Ensembl:ENSG00000136492 MIM:605882; Vega:OTTHUMMG00000179233
Gene type	protein coding
RefSeq status	REVIEWED
Organism	<u>Homo sapiens</u>
Lineage	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo
Also known as	OF; BACH1; FANCJ
Summary	The protein encoded by this gene is a member of the RecQL DEAH helicase family and interacts with the BRCT repeats of breast cancer, type 1 (BRCA1). The bound complex is important in the normal double-strand break repair function of breast cancer, type 1 (BRCA1). This gene may be a target of germline cancer-inducing mutations. [provided by RefSeq, Jul 2008]

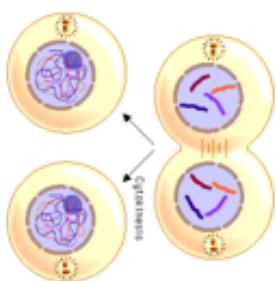
Function	Evidence Code	Pubs
4 iron, 4 sulfur cluster binding	IEA	
ATP binding	IEA	
ATP-dependent DNA helicase activity	NAS	PubMed
DNA binding	NAS	PubMed
chromatin binding	IEA	
metal ion binding	IEA	
protein binding	IPI	PubMed

Process	Evidence Code	Pubs
DNA damage checkpoint	NAS	PubMed
DNA duplex unwinding	IEA	
DNA replication	TAS	
DNA synthesis involved in DNA repair	TAS	
cellular response to angiotensin	IEA	
cellular response to hypoxia	IEA	
cellular response to vitamin	IEA	
chiasma assembly	IEA	
double-strand break repair	NAS	PubMed
double-strand break repair involved in meiotic recombination	IEA	
meiotic DNA double-strand break processing involved in reciprocal meiotic recombination	IEA	
negative regulation of cell proliferation	IEA	
negative regulation of gene expression	IEA	
regulation of signal transduction by p53 class mediator	TAS	
regulation of transcription from RNA polymerase II promoter	IDA	PubMed

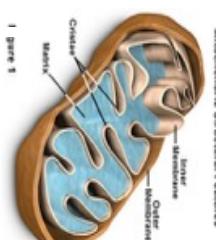
Gene Ontologies (GO)

- Ontologies / controlled vocabularies provide consistent descriptions of gene and gene products

Biological processes: a series of molecular events of functions e.g. “glycolysis”, mitosis (what?)

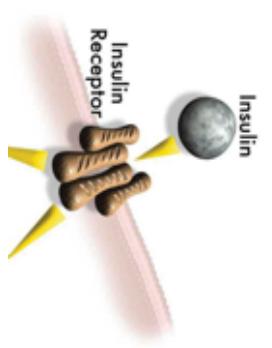


Cellular component: the part of the cell: e.g. “mitochondrion” (where?)



Molecular function: activities performed by individual

gene products at the molecular level e.g. “kinase activity, insulin binding”(how?)



UniProtKB - Q9BX63 (FANCJ_HUMAN)

Display

[Entry](#)
[Publications](#)
[Feature viewer](#)
[Feature table](#)
[None](#)

[BLAST](#) [Align](#) [Format](#) [Add to basket](#) [History](#)
Protein **BRIP1**
Gene **Fanconi anemia group J protein**
Organism *Homo sapiens (Human)*
Status Reviewed - Annotation score: - Experimental evidence at protein levelⁱ

[Feedback](#) [Help video](#) [Other tutorials and videos](#)

[Basket](#) ▾

Functionⁱ

- Names & Taxonomy
- Subcell. location
- Pathol./Biotech
- PTM / Processing
- Expression
- Interaction
- Structure
- Family & Domains

DNA-dependent ATPase and 5' to 3' DNA helicase required for the maintenance of chromosomal stability. Acts late in the Fanconi anemia pathway, after FANCCD2 ubiquitination. Involved in the repair of DNA double-strand breaks by homologous recombination in a manner that depends on its association with BRCA1. 4 Publications ▾

Catalytic activityⁱ
 $\text{ATP} + \text{H}_2\text{O} = \text{ADP} + \text{phosphate}$.

Cofactorⁱ
[4Fe-4S] cluster 1 Publication ▾
Note: Binds 1 [4Fe-4S] cluster. 1 Publication ▾

UniProt will detail much more for example which position of the protein is responsible for Metal binding

Sites

Feature key	Position(s)	Description	Actions	Graphical view	Length
Metal binding ⁱ	283	Iron-sulfur (4Fe-4S) By similarity			1
Metal binding ⁱ	298	Iron-Sulfur (4Fe-4S) By similarity			1
Metal binding ⁱ	310	Iron-sulfur (4Fe-4S) By similarity			1
Metal binding ⁱ	350	Iron-sulfur (4Fe-4S) By similarity			1

Display

Entry

Publications

Feature viewer

Feature table

None

Disease susceptibility is associated with variations affecting disease description: A common malignancy originating from breast epithelial tissue. Breast neoplasms can be distinguished by their histologic pattern. Invasive ductal carcinoma is by far the most common type. Breast cancer is etiologically and genetically heterogeneous. Important genetic factors have been indicated by familial occurrence and bilateral involvement. Mutations at more than one locus can be involved in different families or even in the same case.

See also OMIM:114480

Feature key	Position(s)	Description	Actions	Graphical view	Length
Natural variant ⁱ (MAR_020896)	47	P → A in BC; early onset; loss of ATPase and helicase activities.	2 Publications ▾	Corresponds to variant dbSNP:rs289903098	1
Natural variant ⁱ (MAR_020900)	299	M → I in BC; early onset; reduces helicase efficiency on longer substrates.	2 Publications ▾	Ensembl, ClinVar.	1

Fanconi anemia complementation group J (FANCIJ)

3 Publications ▾

Function

Names & Taxonomy

Subcell. location

Pathol/BioTech

PTM / Processing

Expression

Interaction

Structure

Family & Domains

Sequences (2)

Cross-references

Entry information

Miscellaneous

Similar proteins

What variants of this enzyme are involved in other human diseases

Pathology & Biotechⁱ

Involvement in diseaseⁱ

Breast cancer (BC) 2 Publications ▾

Disease description: A common malignancy originating from breast epithelial tissue. Breast neoplasms can be distinguished by their histologic pattern. Invasive ductal carcinoma is by far the most common type. Breast cancer is etiologically and genetically heterogeneous. Important genetic factors have been indicated by familial occurrence and bilateral involvement. Mutations at more than one locus can be involved in different families or even in the same case.

See also OMIM:114480

Feature key	Position(s)	Description	Actions	Graphical view	Length
Natural variant ⁱ (MAR_023700)	255	Q → H in FANCIJ.	1 Publication ▾		1
Natural variant ⁱ (MAR_023702)	349	A → P in FANCIJ; destabilizes iron-sulfur-binding and abolishes helicase activity.	2 Publications ▾		1
Natural variant ⁱ (MAR_023703)	647	W → C in FANCIJ; associated with C-647.	1 Publication ▾		1
Natural variant ⁱ (MAR_023704)	707	R → C in FANCIJ; associated with C-647.	1 Publication ▾		1

▲ Top

Display

Entry

Publications

Feature viewer

Feature table

None

Domainⁱ

Regionⁱ

Motifⁱ

Feature key

Position(s)

Actions

Graphical view

Length

11 – 442

Helicase ATP-binding

PROSITE-ProRule annotation

Add

BLAST

1 Publication

432

888 – 1063

Interaction with BRCA1

1 Publication

Add

BLAST

176

158 – 175

Nuclear localization signal

Sequence analysis

Add

BLAST

18

393 – 396

DEAH box

Add

BLAST

4

Miscellaneous databases
EvolutionaryTraceⁱ Q9BX63.

Family & Domainsⁱ

What is known about the protein family

Domains and Repeats					
Feature key	Position(s)	Description	Actions	Graphical view	Length
Domain ⁱ	11 – 442	Helicase ATP-binding	PROSITE-ProRule annotation	Add	432
Region ⁱ	888 – 1063	Interaction with BRCA1	1 Publication	Add	176
Motif ⁱ	Position(s)	Description	Actions	Graphical view	Length
Motif ⁱ	158 – 175	Nuclear localization signal	Sequence analysis	Add	18
Motif ⁱ	393 – 396	DEAH box	Add	BLAST	4
Domain ⁱ	Position(s)	Description	Actions	Graphical view	Length
Sequences (2)	4Fe-4S iron-sulfur-binding	is required for helicase activity.	1 Publication		
Cross-references					
Entry information					
Miscellaneous					
Similar proteins					
Phylogenomic databases					
eggNOG ⁱ	KOG1132. Eukaryota.				
	COG1199. LUCA.				

▲ Top

Display

Entry

Publications

Feature viewer

Feature table

None

Function

Names & Taxonomy

Subcellular location

Pathology & Biotech

PTM / Processing

Expression

Interaction

Structure

Family & Domains¹

Sequence similarities¹

Belongs to the RuBisCO small chain family.

Curated

Keywords - Domain¹

Transit peptide

Family and domain databases

Gene3Dⁱ 3.30.190.10. 1 hit.

InterProⁱ View protein in InterPro

IPR024681. RuBisCO_sc.

IPR000894. RuBisCO_sc_dom.

IPR024680. RuBisCO_ssu_N.

Pfamⁱ View protein in Pfam

PF12338. RbcS. 1 hit.

PF00101. RuBisCO_small. 1 hit.

PFAM is one of the best protein databases

PRINTSⁱ PR00152. RUBISCOSMALL.

SMARTⁱ View protein in SMART

SM00961. RuBisCO_small. 1 hit.

SUPFAMⁱ SSF55239. SSF55239. 1 hit.



You can find out more about the
**Protein family and its species
distribution and residue wise
conservation e.t.c...?**

Family: RbcS (PF12338)

Summary

Domain organisation

Clan

Alignments

HMM logo

Trees

Curation & model

Species

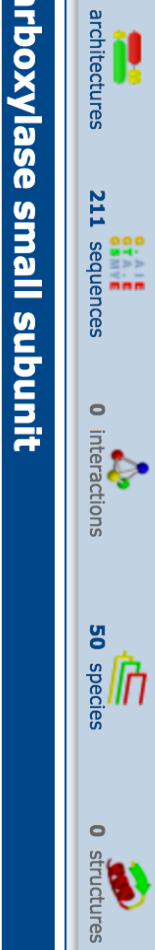
Interactions
Structures

Literature references

Jump to...

enter ID/acc

Go



Summary: Ribulose-1,5-bisphosphate carboxylase small subunit

Pfam includes annotations and additional family information from a range of different sources. These sources can be accessed via the tabs below.

This tab holds the annotation information that is stored in the Pfam database. As we move to using Wikipedia as our main source of annotation, the contents of this tab will be gradually replaced by the Wikipedia tab.

Ribulose-1,5-bisphosphate carboxylase small subunit

This domain family is found in eukaryotes, and is approximately 40 amino acids in length. The family is found in association with PF00101. There is a conserved APF sequence motif. There are two completely conserved residues (L and P) that may be functionally important. This family is the small subunit of ribulose-1,5-bisphosphate.

1. Hutchinson KW, Harvie PD, Singer PB, Brunner AF, Greenwood MS., Plant Mol Biol. 1990;14:281-284.: Nucleotide sequence of the small subunit of ribulose-1,5-bisphosphate carboxylase from the conifer Larix laricina. PUBMED:2101694 EPMC:2101694

Comments or questions on the site? Send a mail to pfam-help@ebi.ac.uk.

European Molecular Biology Laboratory

Family: RbcS (PF12338)

You can find out more about the
species distribution



Summary

Domain organisation

Clan

Alignments

HMM logo

Trees

Curation & model

Species

Interactions

Structures

Jump to...

enter ID/acc



Species distribution

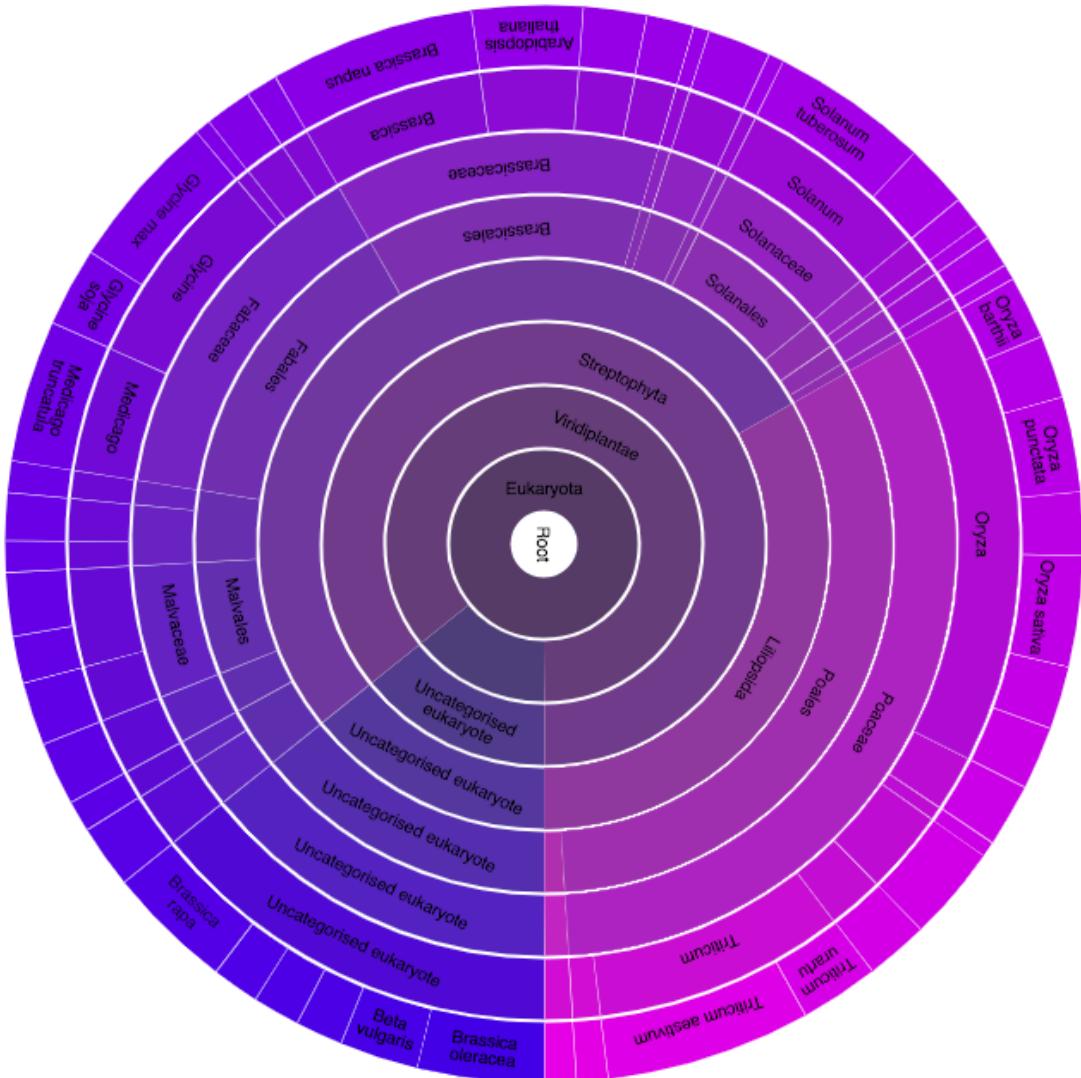
Sunburst

Tree

This visualisation provides a simple graphical representation of the distribution of this family across species. You can find the original interactive tree in the adjacent tab. [More...](#)

Sunburst controls Hide

Brassica oleracea var. oleracea



Weight segments by...

● number of sequences

○ number of species

Change the size of the sunburst

Small ▾ Large ▾

Colour assignments

Archaea	Orange
Bacteria	Green
Viruses	Magenta
Virooids	Red
Eukaryota	Purple
Other sequences	Yellow
Unclassified	Cyan
Undifferentiated sequence	Light Blue

Selections

Align selected sequences to HMM

Generate a FASTA-format file

[Clear selection](#)

Family: Kinesin (PF00225)

>Loading page components (1 remaining)...

Summary

Domain organisation

Clans

Alignments

HMM logo

Trees

Curation & models

Species

Interactions

Structures

Jump to... ↪

enter ID/acc

Go

Interactions

There are **6** interactions for this family. [More...](#)

[Tubulin](#)

[Tubulin_C](#)

[Tubulin](#)

[Kinesin](#)

[Kinesin](#)

[Tubulin](#)

[Kinesin](#)

[Kinesin](#)



Family: Kinesin (PF00225)

126 architectures 4150 sequences 6 interactions 248 species 114 structures

Summary

Domain organisation
Clans

Alignments

HMM logo

Trees

Curation & models

Species

Interactions

Structures

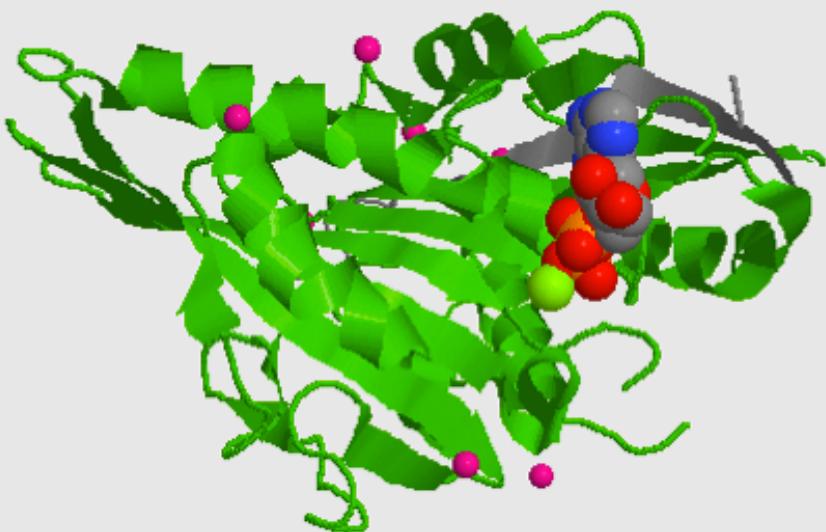
Structures

For those sequences which have a structure in the [Protein DataBank](#), we use the mapping between [UniProt](#), PDB and Pfam coordinate systems from the [PDBeG](#) group, to allow us to map Pfam domains onto UniProt sequences and three-dimensional protein structures. The table below shows the structures on which the Kinesin domain has been found.

UniProt entry	UniProt residues	PDB ID	PDB chain ID	PDB residues	View
A8BKD1_GIALA	11 - 335	2wq	A	11 - 335	Jmol AstexViewer SPICE
			B	11 - 335	Jmol AstexViewer SPICE
CENPE_HUMAN	12 - 329	1t5c	A	12 - 329	Jmol AstexViewer SPICE
			B	12 - 329	Jmol AstexViewer SPICE
		1f9t	A	392 - 723	Jmol AstexViewer SPICE
		1f9u	A	392 - 723	Jmol AstexViewer SPICE
		1f9v	A	392 - 723	Jmol AstexViewer SPICE
KAR3_YEAST	392 - 723	1f9w	A	392 - 723	Jmol AstexViewer SPICE
		B	392 - 723	Jmol AstexViewer SPICE	
		3kar	A	392 - 723	Jmol AstexViewer SPICE
K113B_HUMAN	11 - 352	3gbi	A	11 - 352	Jmol AstexViewer SPICE
			B	11 - 352	Jmol AstexViewer SPICE
			C	11 - 352	Jmol AstexViewer SPICE
		1ii6	A	24 - 359	Jmol AstexViewer SPICE
			B	24 - 359	Jmol AstexViewer SPICE
		1q0b	A	24 - 359	Jmol AstexViewer SPICE
			B	24 - 359	Jmol AstexViewer SPICE
		1x88	A	24 - 359	Jmol AstexViewer SPICE
			B	24 - 359	Jmol AstexViewer SPICE
			A	24 - 359	Jmol AstexViewer SPICE

Jump to... ↻

enter ID/acc [Go](#)



Jmol

PDB	UniProt		Pfam family		Colour	
Chain	Start	End	ID	Start	End	
A	49	368	KIF22_HUMAN	49	368	Kinesin (PF00225)

Conclusion

- Bioinformatics deals with the collection, arching, organizing and interpretaion of a wide range of biological data.
- There are a large number of primary, secondary and tertiary databases
- Use appropriate databases
- Know what kind of information to expect
- The NCBI and EBI are the major online bioinformatics service providers