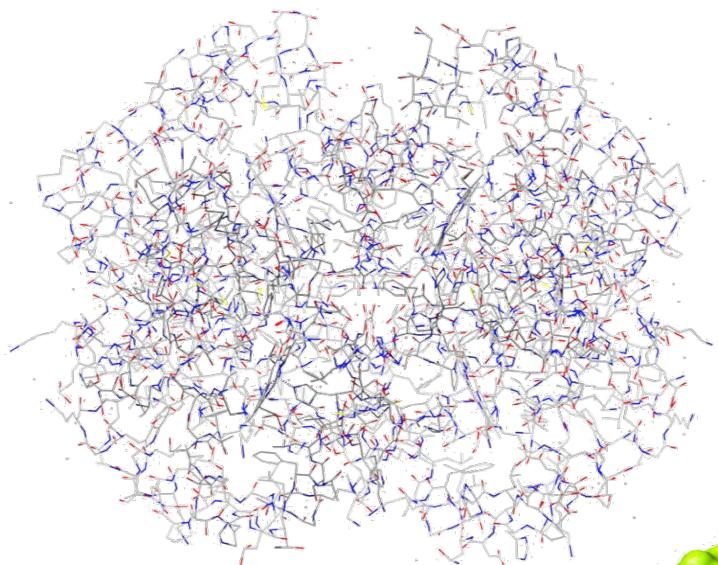
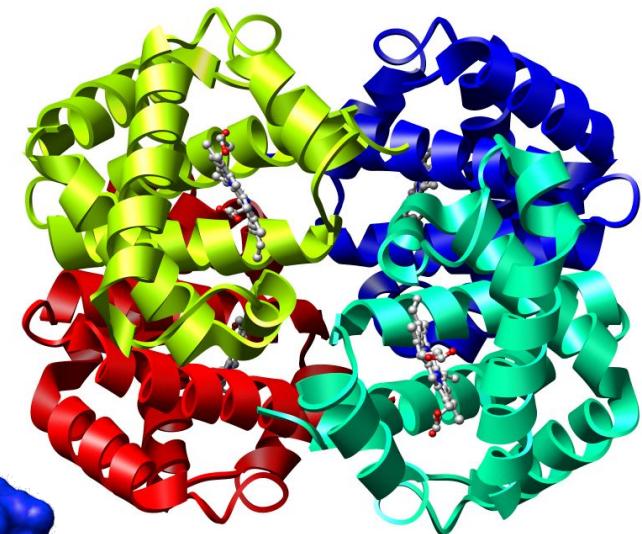


Visualization Conventions



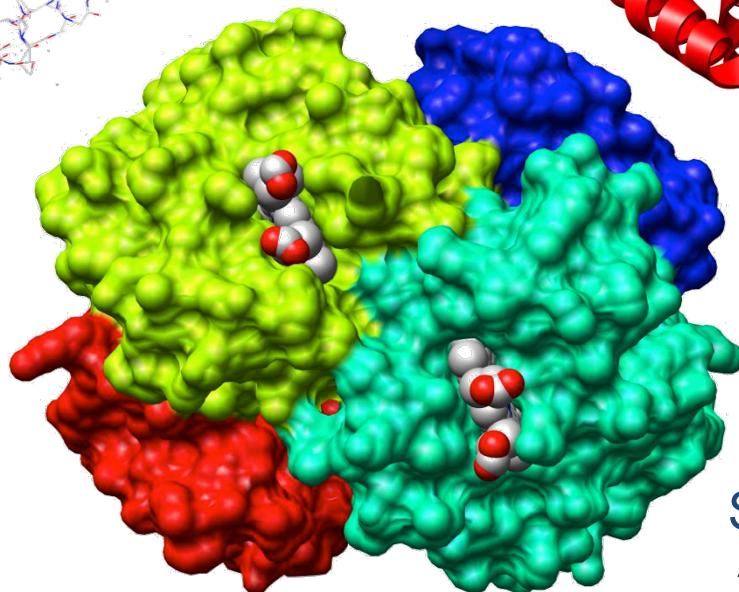
Wireframe

All atoms seen



Ribbons

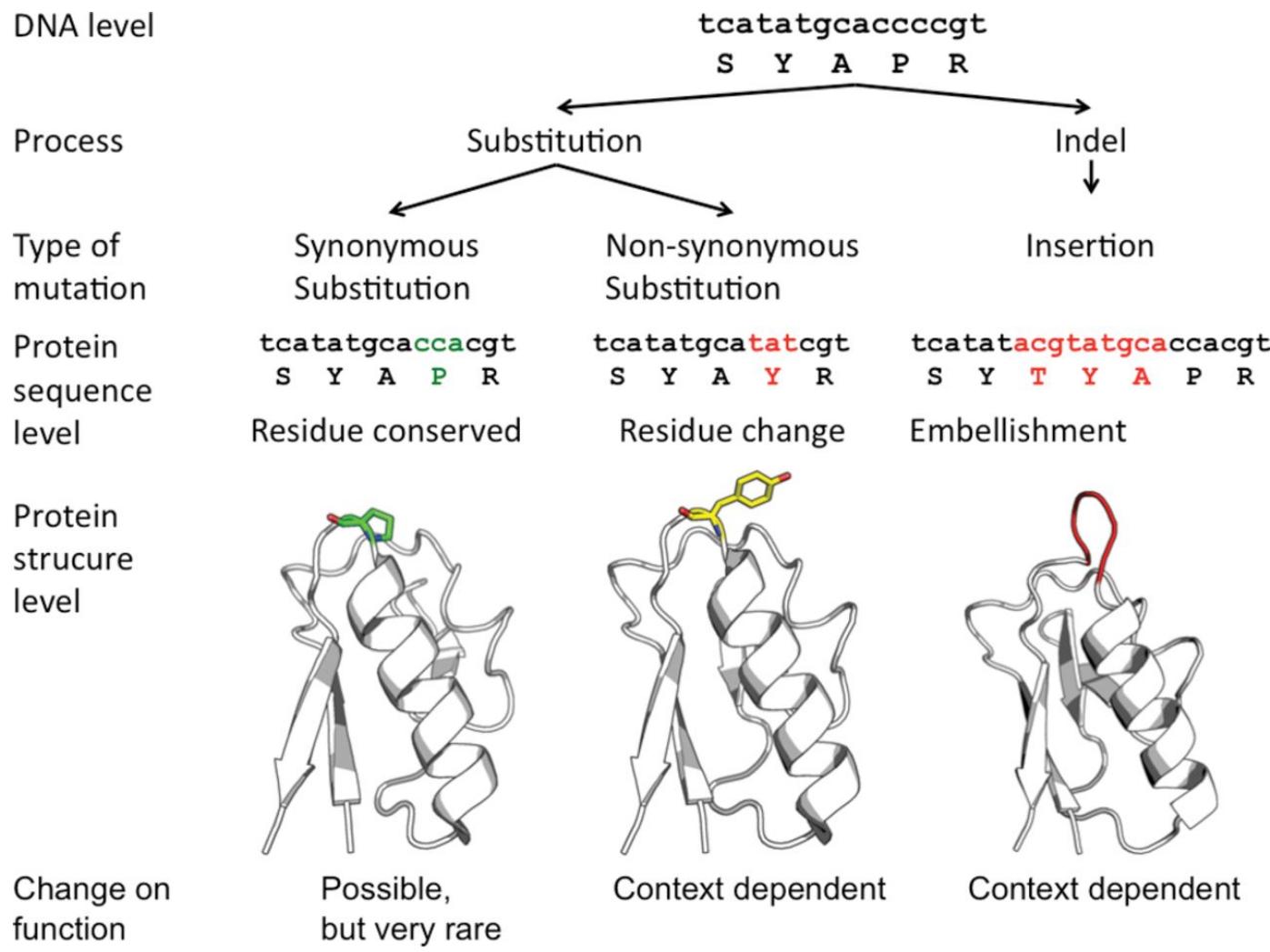
Only backbone C-alpha atoms seen



Spacefill

All atoms seen

Possible effects of mutations on proteins Various mutational processes can affect proteins.



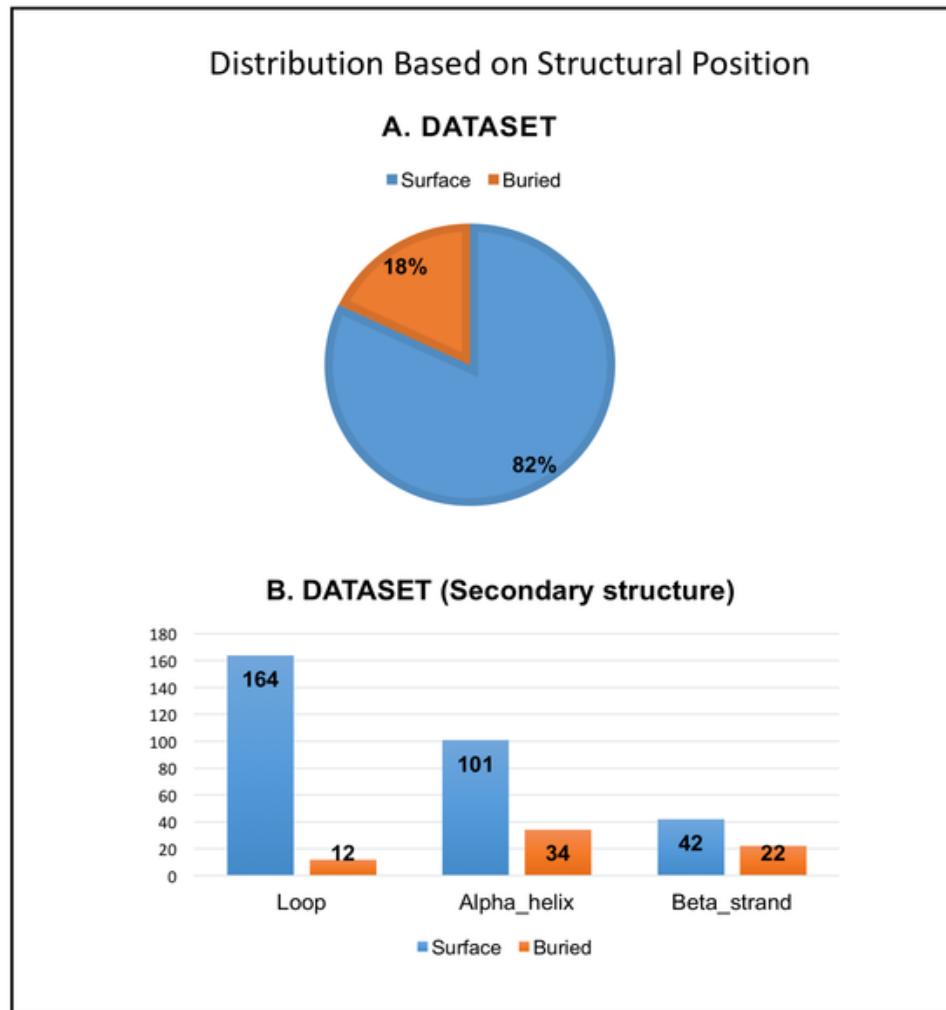
Romain A. Studer et al. Biochem. J. 2013;449:581-594

Possible effects of nonsynonymous mutations on proteins

- no impact
 - amino acids with similar biochemical properties
- potential effect, weak or strong
 - effect on folding, binding and enzyme catalysis
- damage the protein structure by affecting the stability.
- Indels block of sites is inserted or deleted
 - effect depends on context

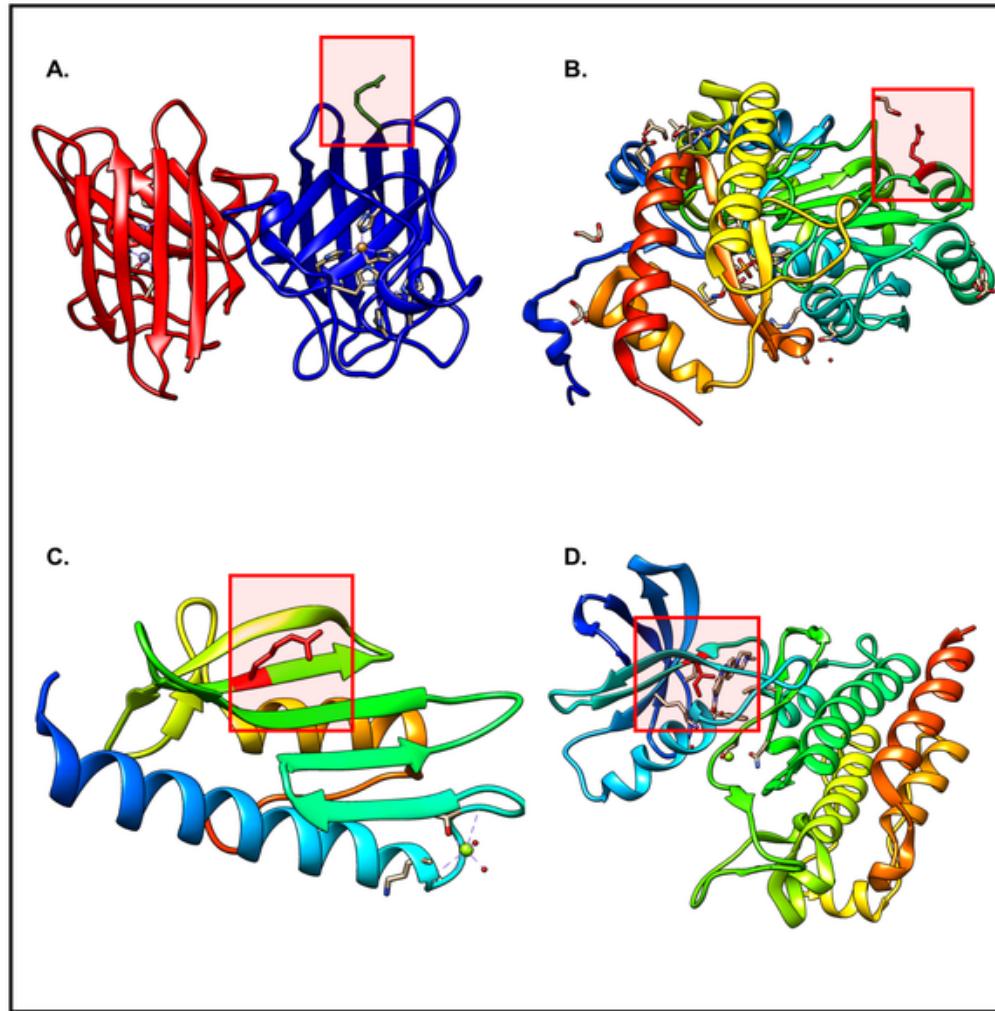
The probability of fixation of these mutations in the genome depends on different factors, such as the population size or the beneficial effect of the mutation on the organismal fitness.

Distribution of SNVs based on structural position.



Bhattacharya R, Rose PW, Burley SK, Prlić A (2017) Impact of genetic variation on three dimensional structure and function of proteins. PLOS ONE 12(3): e0171355. <https://doi.org/10.1371/journal.pone.0171355>
<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0171355>

Fig 2. SNV consequences map to various locations within protein structures.



Bhattacharya R, Rose PW, Burley SK, Prlić A (2017) Impact of genetic variation on three dimensional structure and function of proteins. PLOS ONE 12(3): e0171355. <https://doi.org/10.1371/journal.pone.0171355>
<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0171355>

Table 3. Examples for each SNV related effect category.

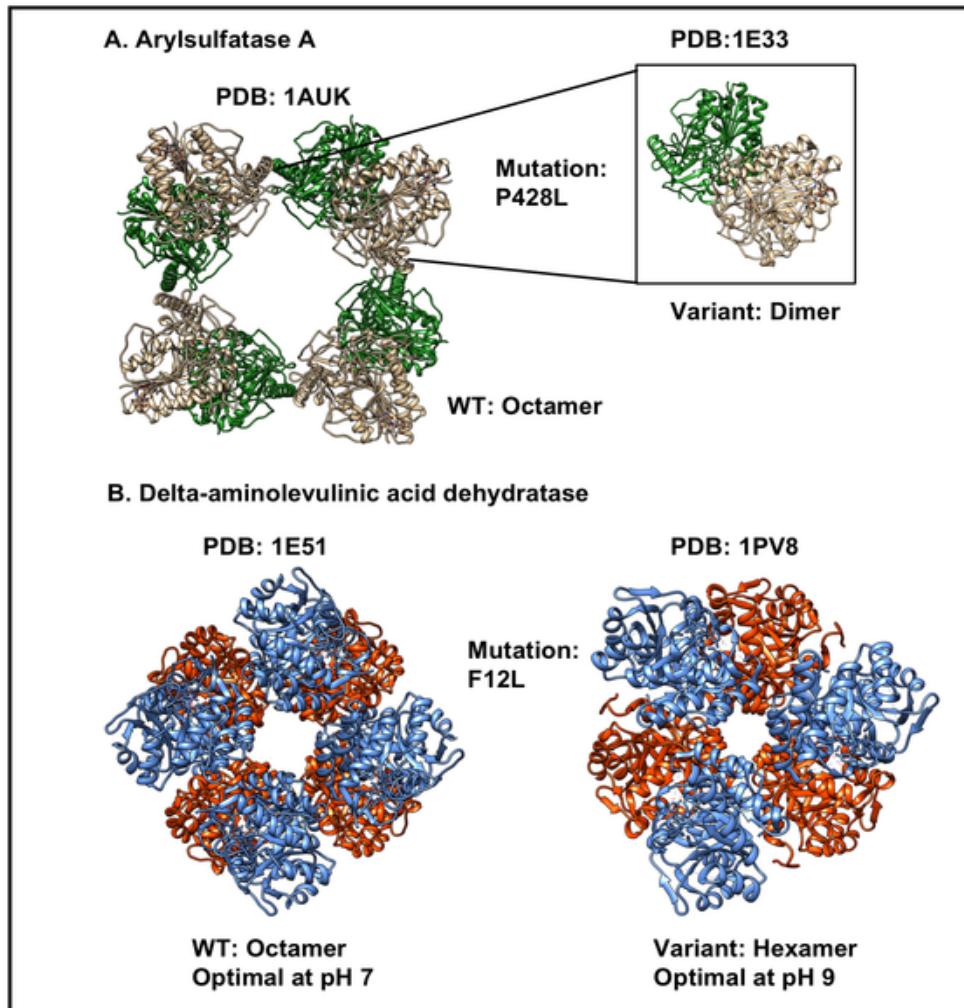
Activity	rs137852646	Glycyl-tRNA synthetase	2PMF	2ZT5	G526R	Loss of activity	Charcot-Marie-Tooth disease	[50]
Aggregation	rs121912442	Cu, Zn superoxide dismutase [HSOD]	1N19	4FF9	A4V	Destabilization of protein and formation of aggregates.	Lou Gehrig's disease	[51]
Stability	rs74315351	DJ-1	2RK4	1P5F	M26I	Leads to decrease thermal stability and inactivation.	Rare forms of familial Parkinsonism	[52,54]
Binding	rs104894227	HRAS	2QUZ	2CE2	K117R	Increases the rate of nucleotide dissociation and results in constitutive activation of HRAS.	Costello Syndrome	[55]
Assembly	rs1141718	Manganese superoxide dismutase	1VAR	1MSD	I58T	The packing defects due to the mutation disrupt the dimer-tetramer equilibrium and favor the dimer over tetramer in solution.	Amyotrophic Lateral Sclerosis	[56]
Rearrangement	rs61749389	von Willebrand factor	1IJK	1OAK	I546V	The mutation causes a "Gain of Function" effect and produces a phenotype in which regulation is lost	von Willebrand disease	[57]

<https://doi.org/10.1371/journal.pone.0171355.t003>

Bhattacharya R, Rose PW, Burley SK, Prlić A (2017) Impact of genetic variation on three dimensional structure and function of proteins. PLOS ONE 12(3): e0171355. <https://doi.org/10.1371/journal.pone.0171355>

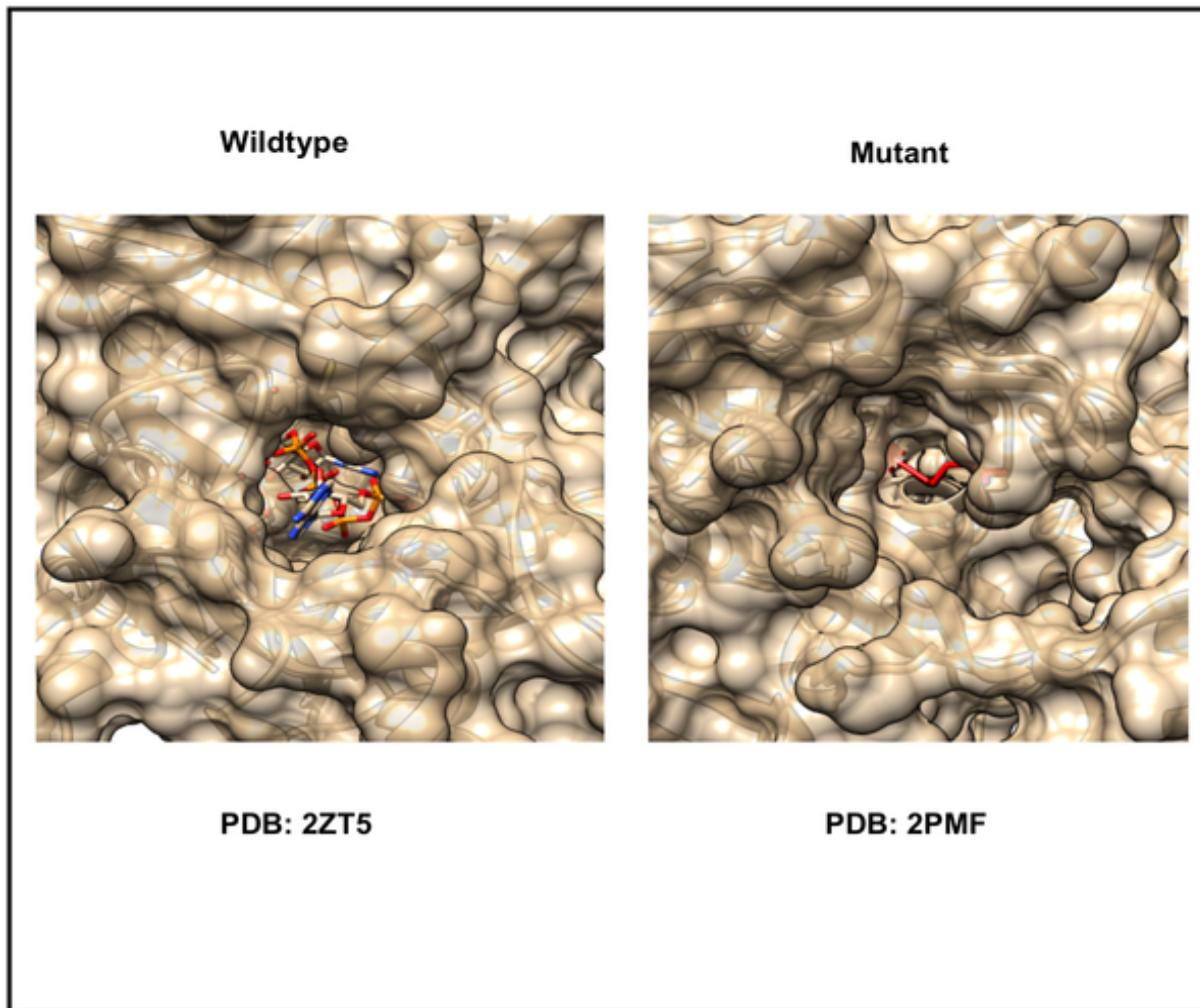
<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0171355>

SNVs that affect both protein structure and function.



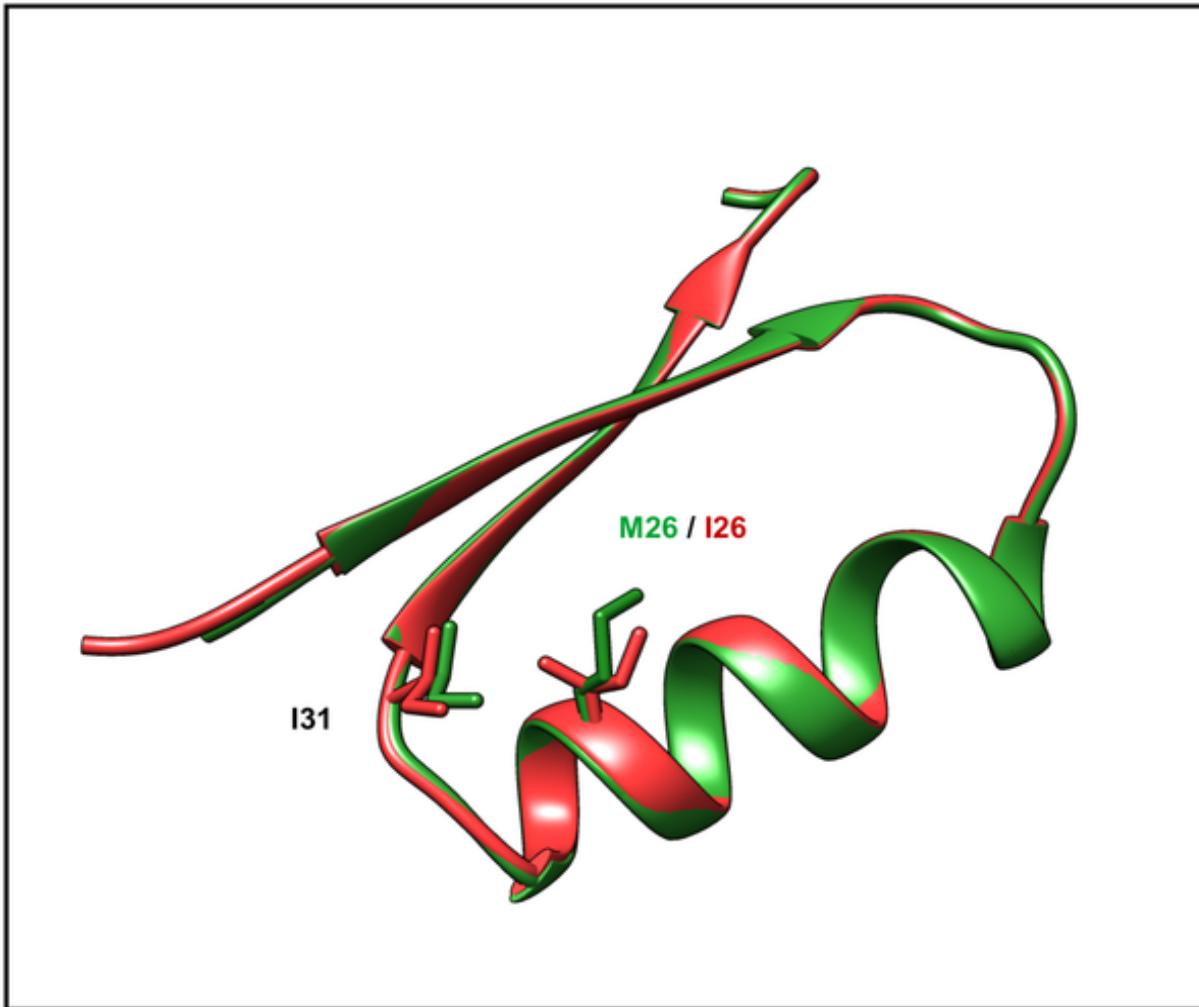
Bhattacharya R, Rose PW, Burley SK, Prlić A (2017) Impact of genetic variation on three dimensional structure and function of proteins. PLOS ONE 12(3): e0171355. <https://doi.org/10.1371/journal.pone.0171355>

SNV related change that affects enzymatic activity.



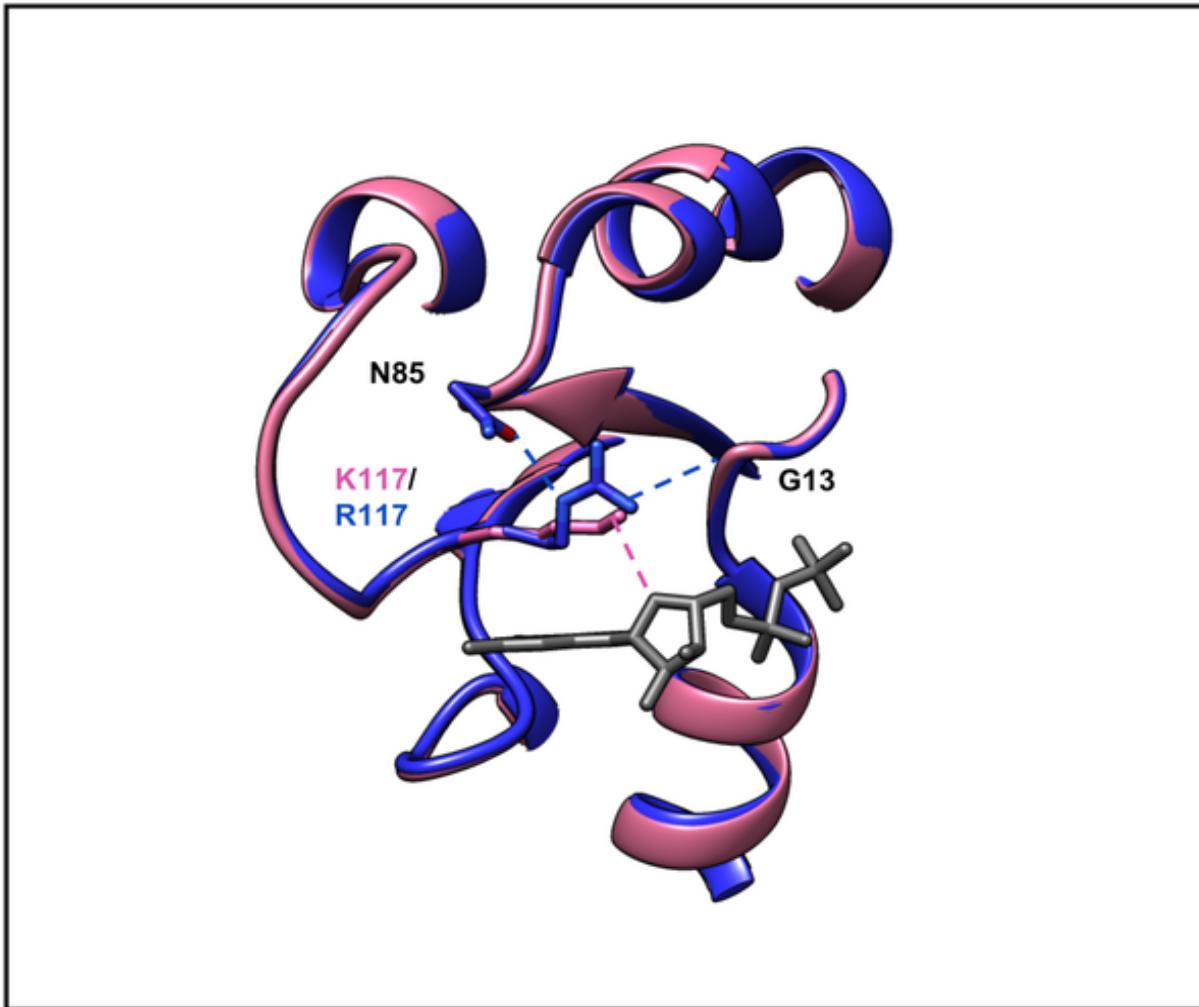
Bhattacharya R, Rose PW, Burley SK, Prlić A (2017) Impact of genetic variation on three dimensional structure and function of proteins. PLOS ONE 12(3): e0171355. <https://doi.org/10.1371/journal.pone.0171355>
<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0171355>

SNV that affects protein structure stability.



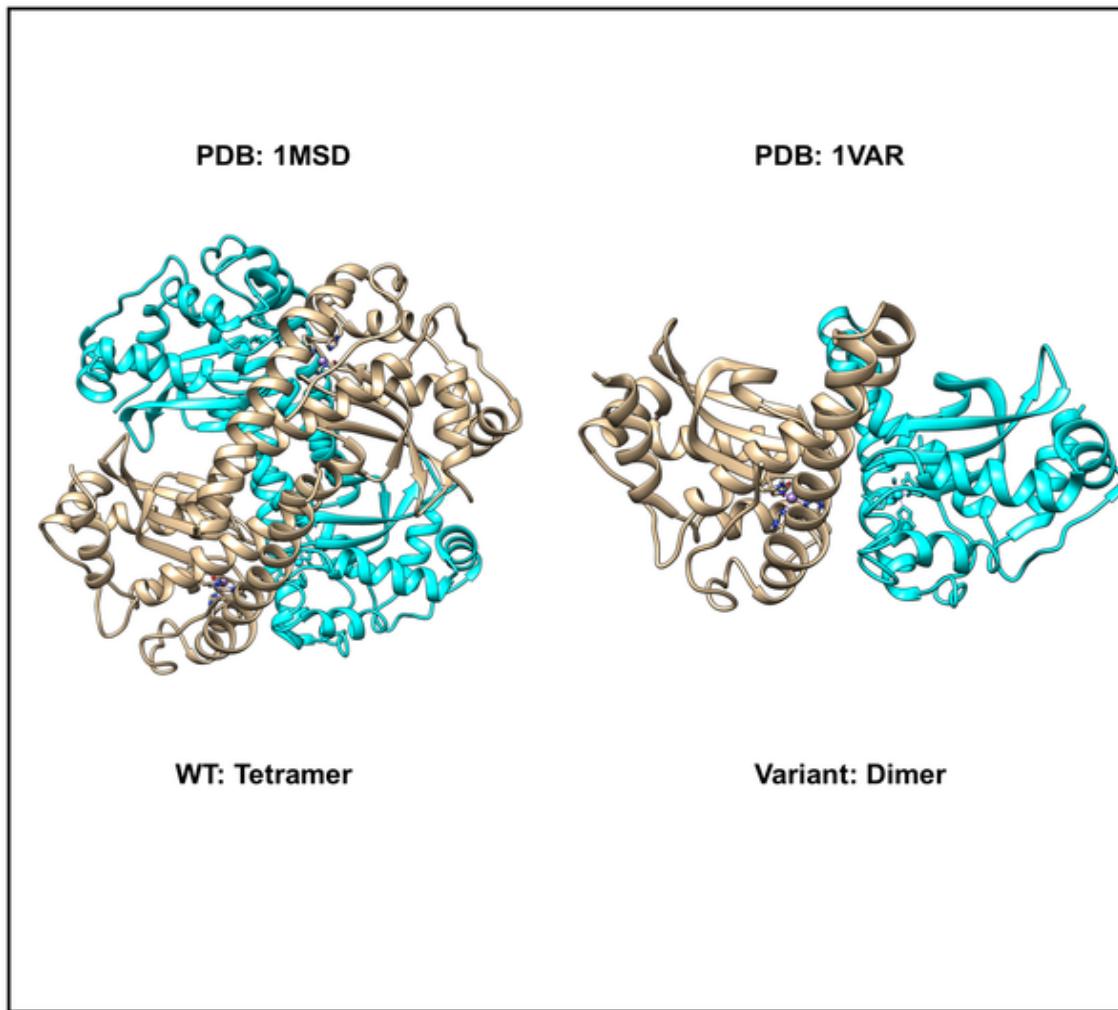
Bhattacharya R, Rose PW, Burley SK, Prlić A (2017) Impact of genetic variation on three dimensional structure and function of proteins. PLOS ONE 12(3): e0171355. <https://doi.org/10.1371/journal.pone.0171355>
<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0171355>

Close-up view of the nucleotide-binding region of Lys117Arg.



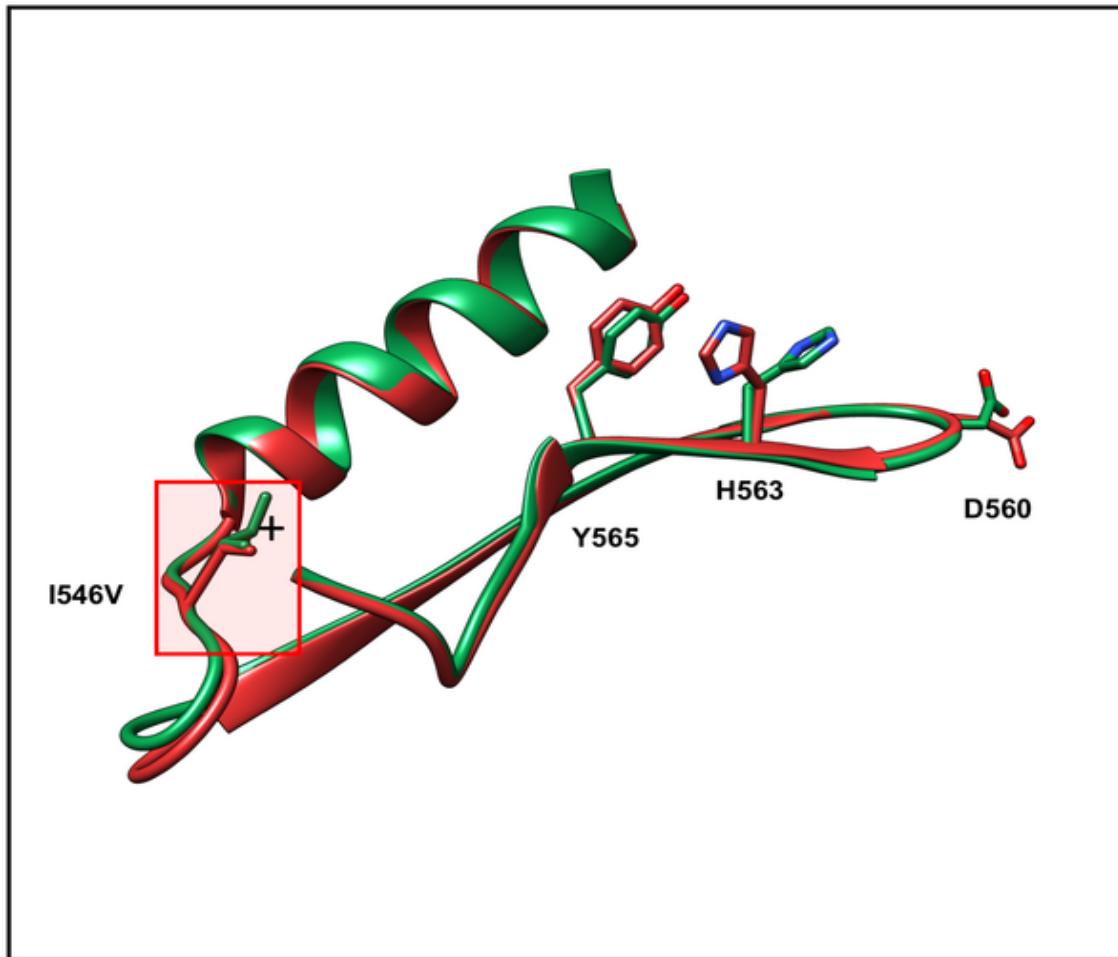
Bhattacharya R, Rose PW, Burley SK, Prlić A (2017) Impact of genetic variation on three dimensional structure and function of proteins. PLOS ONE 12(3): e0171355. <https://doi.org/10.1371/journal.pone.0171355>
<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0171355>

In manganese superoxide dismutase, a SNV can affect protein assembly.



Bhattacharya R, Rose PW, Burley SK, Prlić A (2017) Impact of genetic variation on three dimensional structure and function of proteins. PLOS ONE 12(3): e0171355. <https://doi.org/10.1371/journal.pone.0171355>
<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0171355>

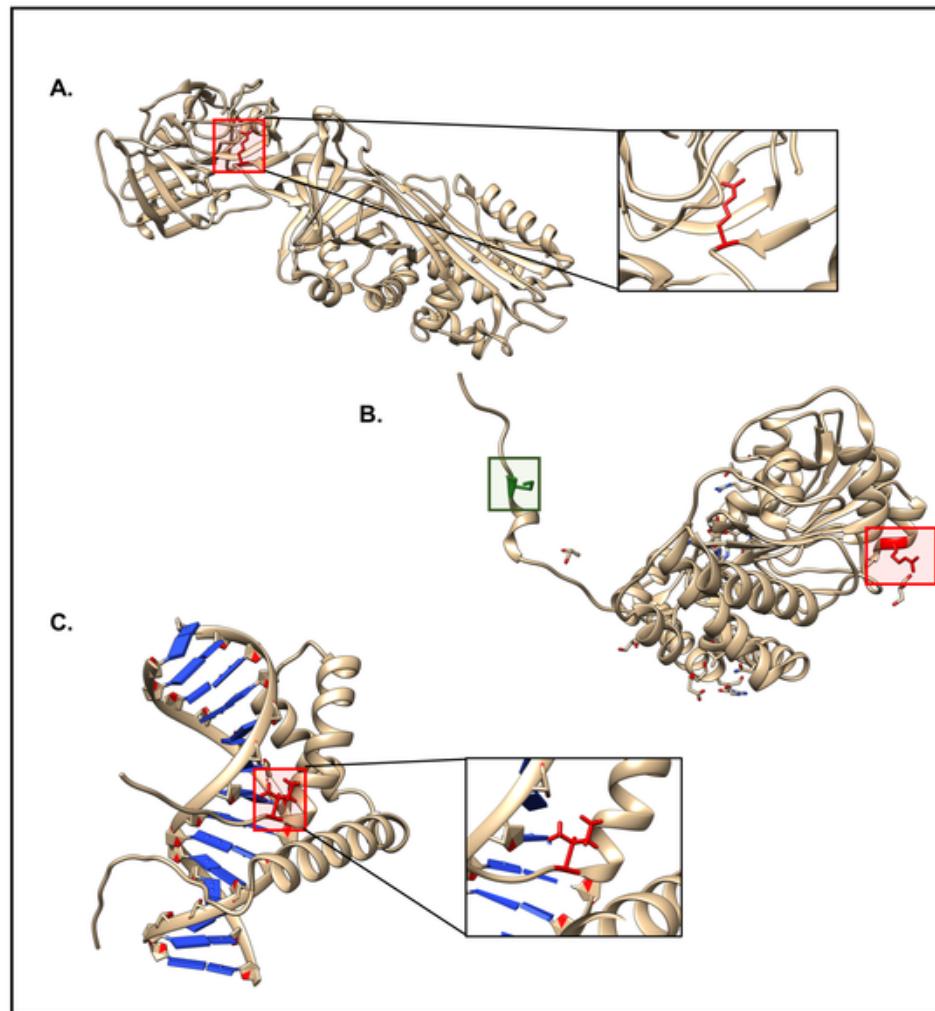
von Willebrand factor (wild-type: green PDBID 1OAK; I546V mutant PDB: 1IJK) with the location of I546V mutation highlighted.



Bhattacharya R, Rose PW, Burley SK, Prlić A (2017) Impact of genetic variation on three dimensional structure and function of proteins. PLOS ONE 12(3): e0171355. <https://doi.org/10.1371/journal.pone.0171355>

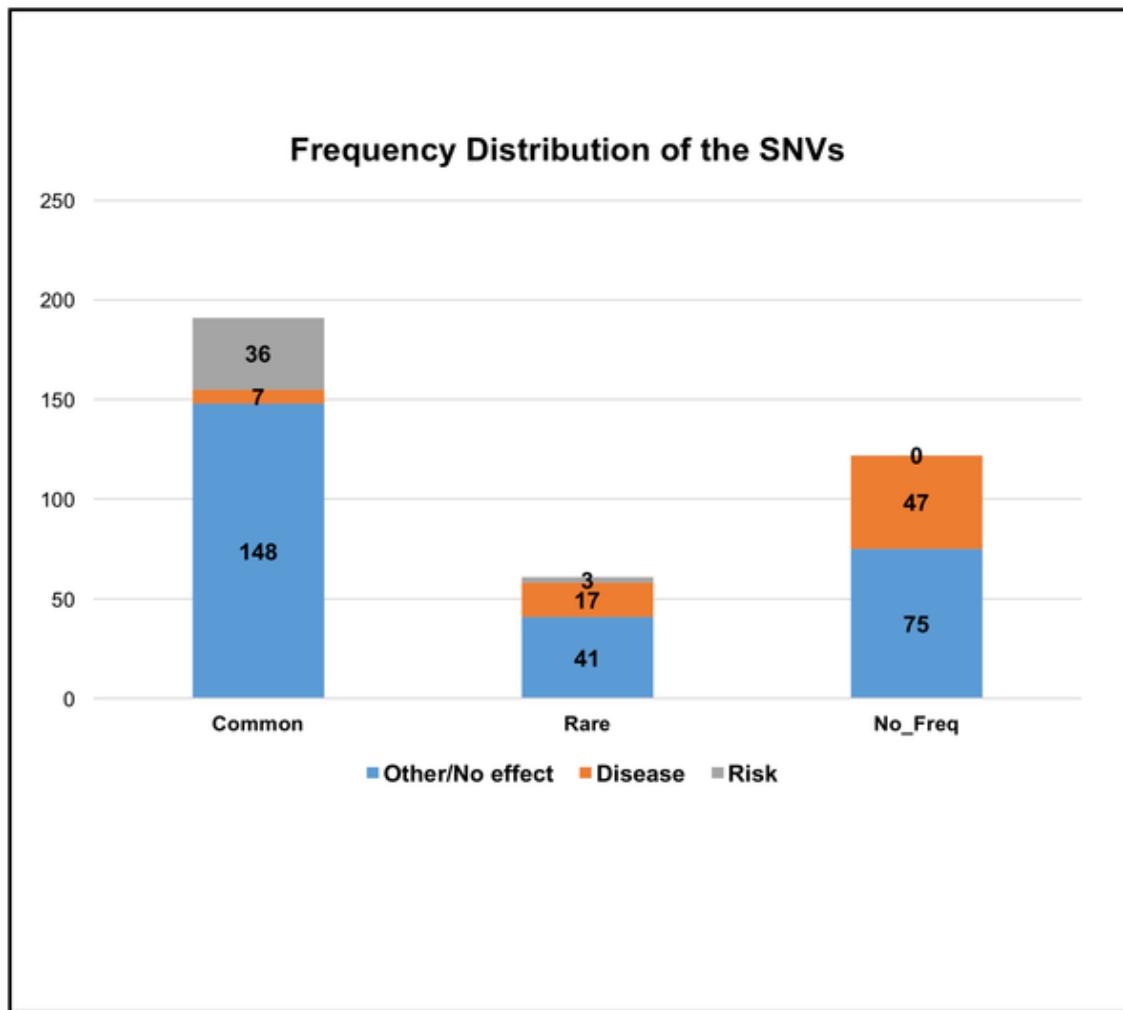
<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0171355>

special cases



Bhattacharya R, Rose PW, Burley SK, Prlić A (2017) Impact of genetic variation on three dimensional structure and function of proteins. PLOS ONE 12(3): e0171355. <https://doi.org/10.1371/journal.pone.0171355>
<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0171355>

Frequency distribution of the SNVs.



Bhattacharya R, Rose PW, Burley SK, Prlić A (2017) Impact of genetic variation on three dimensional structure and function of proteins. PLOS ONE 12(3): e0171355. <https://doi.org/10.1371/journal.pone.0171355>
<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0171355>

Vocabulary

SNV: Single-nucleotide variants are a substitution of one DNA base pair for another and may fall within genes (either protein-coding or functional RNA genes) in gene regulatory regions or in intergenic regions.

Synonymous substitution: they encode the same amino acid due to redundancy/degeneracy in the genetic code and so have no effect on the protein product of a gene

Nonsynonymous substitution (NSV): they change a single amino acid in the protein

Not all nonsynonomous substitutions are “damaging”

Damaging functional effect: affecting the biochemical activity or regulatory control of a protein

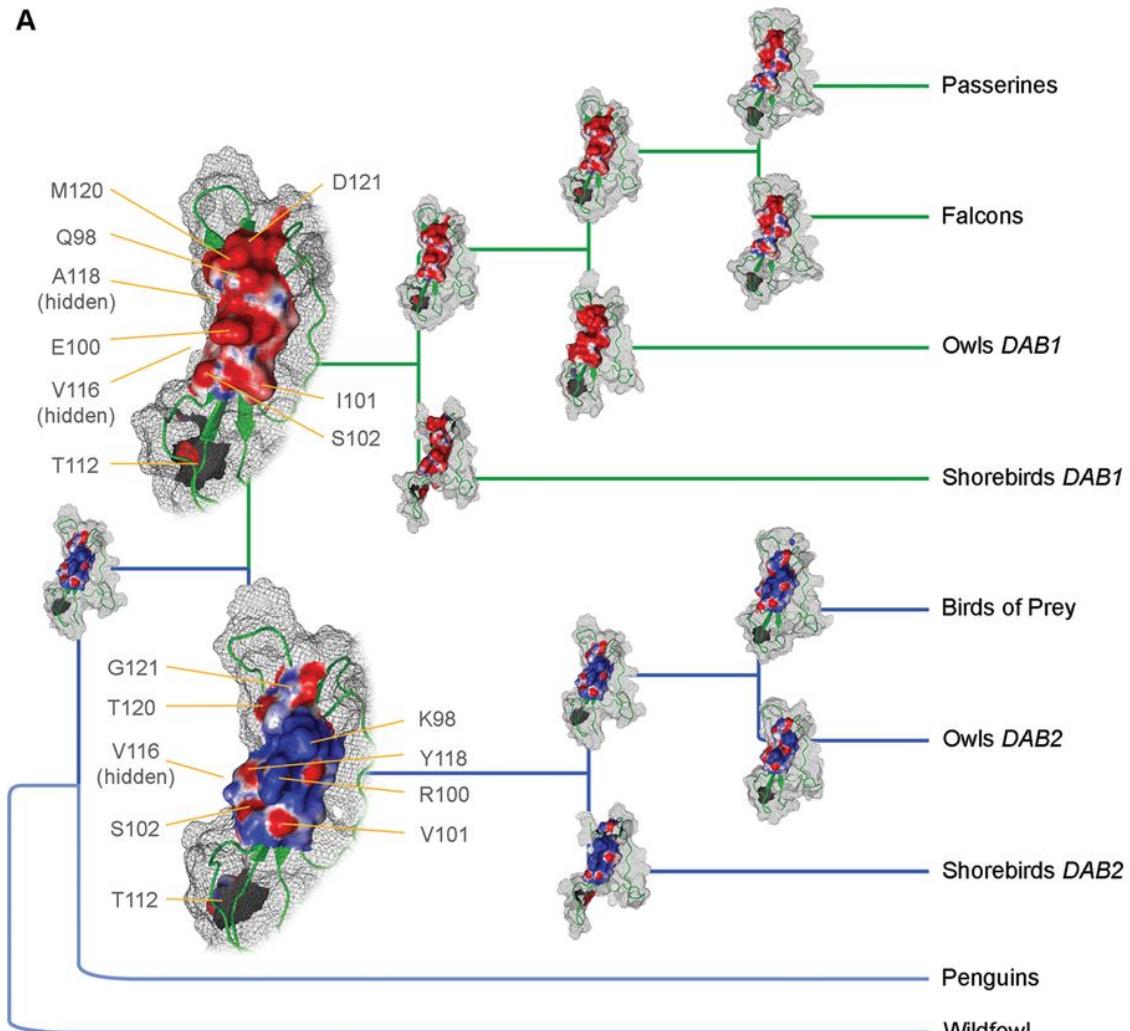
- decrease or increase activity
- affect binding
- affect stability
- affect folding

Not all damaging NSV are deleterious

Deleterious effect: result in phenotype at organism level that is subjected to natural negative selection

Evolution of dimerization binding interface of MHC in birds Evolutionary history of MHC II protein in birds.

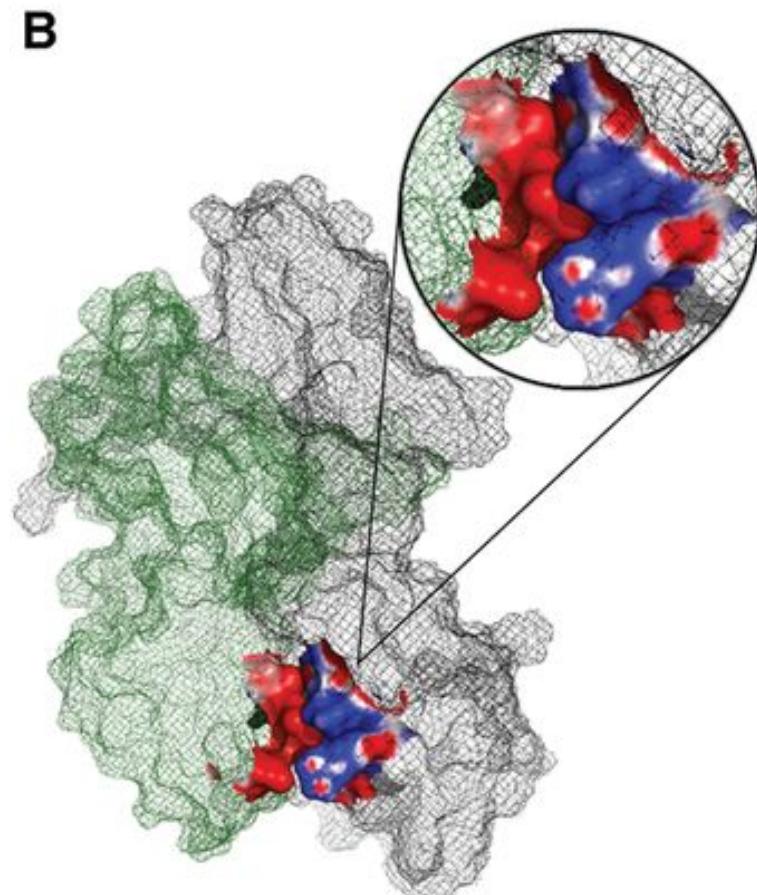
A



The clustering of sites in three-dimensional space allows for the visualization of the electrostatic patches and their differences between the two branches of the trees (blue and green).
Basic residues – blue
Acidic residues - red

Romain A. Studer et al. Biochem. J. 2013;449:581-594

Evolution of dimerization binding interface of MHC in birds Evolutionary history of MHC II protein in birds.

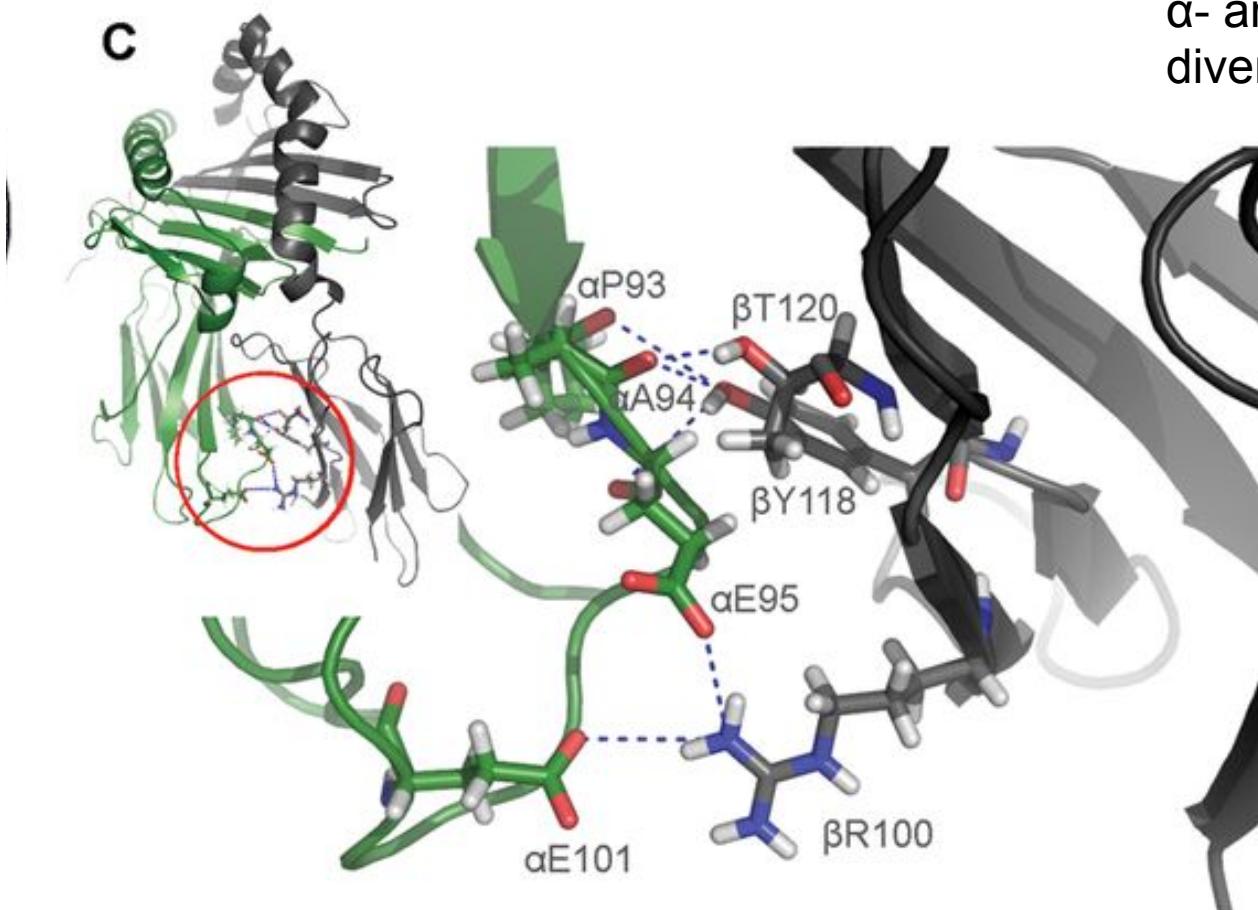


The structure of the chicken MHC II $\alpha\beta$ -heterodimer. The divergent protein domain is indicated in colour.

Romain A. Studer et al. Biochem. J. 2013;449:581-594

Evolution of dimerization binding interface of MHC in birds Evolutionary history of MHC II protein in birds.

Charge–charge interactions between the chicken MHC II α - and β -chains in the divergent protein domain.



Romain A. Studer et al. Biochem. J. 2013;449:581-594

Effect of variations on protein structure

Uniprot

UniProt

- UniProt is a collaboration between the [European Bioinformatics Institute \(EBI\)](#), the [Swiss Institute of Bioinformatics \(SIB\)](#) and the [Protein Information Resource \(PIR\)](#).
- The mission of [UniProt](#) is to provide the scientific community with a comprehensive, high-quality and freely accessible resource of protein sequence and functional information.

UniProt

- **SwissProt** high quality annotation, non-redundant & cross-referenced to many other databases.
- **TrEMBL** - computer translation of the genetic information from the EMBL Nucleotide Sequence Database → many proteins are poorly annotated since only automatic annotation is generated

UniProt

- Connected to other databases
 - Pfam , Prosite, EC, GO, PdbSum, PDB
- Each sequence has a unique 6 letter **accession ID**
- **Download** sequence in FASTA format

UniProt: <http://www.uniprot.org/>

Type accession ID etc



The mission of UniProt is to provide the scientific community with a comprehensive, high-quality and freely accessible resource of protein sequence and functional information.

UniProtKB
UniProt Knowledgebase

Swiss-Prot (558,125)
Manually annotated and reviewed.

TrEMBL (124,797,108)
Automatically annotated and not reviewed.

UniRef
Sequence clusters

UniParc
Sequence archive

Proteomes

Supporting data

Literature citations
Cross-ref. databases

Taxonomy
Diseases

Subcellular locations
Keywords

News

BLOG TWITTER FACEBOOK RSS

[Forthcoming changes](#)
Planned changes for UniProt

[UniProt release 2018_08](#)
Human brain development: slow and steady wins the race | 'Enzyme regulation' becomes 'Activity regulation'
| New advanced search interface...

[UniProt release 2018_07](#)
Ubiquitin ligation: new insight into mechanistic diversity | Cross-references to UniLectin | Change of UniProt

[News archive](#)

Getting started

[Text search](#)

Our basic text search allows you to search all the resources available



UniProt data

[Download latest release](#)

Get the UniProt data

Protein spotlight



[Best Left Unsaid](#)

September 2018

There are times in life when things

UniProt

BLAST Align Retrieve/ID mapping Peptide search

Help Cont

UniProtKB results

Basket

UniProtKB consists of two sections:

Reviewed (Swiss-Prot) - Manually annotated

Records with information extracted from literature and curator-evaluated computational analysis.

The UniProt Knowledgebase (UniProtKB) is the central hub for the collection of functional information on proteins, with accurate, consistent and rich annotation. In addition to capturing the core data mandatory for each UniProtKB entry (mainly, the amino acid sequence, protein name or description, taxonomic data and citation information), as much annotation information as possible is added.

Unreviewed (TrEMBL) - Computationally analyzed

Records that await full manual annotation.

? Help  UniProtKB help video  Other tutorials and videos  Download

Filter byⁱ

	Entry	Entry name		Protein names	Gene names	Organism	Length
 Reviewed (440)	P00519	ABL1_HUMAN		Tyrosine-protein kinase ABL1	ABL1 ABL, JTK7	Homo sapiens (Human)	1,130
 Unreviewed (6,248)	P00520	ABL1_MOUSE		Tyrosine-protein kinase ABL1	Abl1 Abl	Mus musculus (Mouse)	1,123
Popular organisms	P42684	ABL2_HUMAN		Tyrosine-protein kinase ABL2	ABL2 ABLL, ARG	Homo sapiens (Human)	1,182
Human (192)	Q4JIM5	ABL2_MOUSE		Tyrosine-protein kinase ABL2	Abl2 Arg	Mus musculus (Mouse)	1,182
Mouse (129)	P00522	ABL_DROME		Tyrosine-protein kinase Abl	Abl ABL-1, Dash, CG4032	Drosophila melanogaster (Fruit fly)	1,620
Rat (52)	P03949	ABL1_CAEEL		Tyrosine-protein kinase abl-1	abl-1 M79.1	Caenorhabditis elegans	1,224

Display 25

UniProt

General data: name, origin, EC (enzymatic reaction)...

UniProtKB - P00519 (ABL1_HUMAN)

Display

Entry

Publications

Feature viewer

Feature table

None

Function

Names & Taxonomy

Subcellular location

Pathology & Biotech

PTM / Processing

Expression

Interaction

Structure

Protein | Tyrosine-protein kinase ABL1

Gene | ABL1

Organism | *Homo sapiens (Human)*

Status | Reviewed - Annotation score: ●●●●● - Experimental evidence at protein levelⁱ

Functionⁱ

Non-receptor tyrosine-protein kinase that plays a role in many key processes linked to cell growth and survival such as cytoskeleton remodeling in response to extracellular stimuli, cell motility and adhesion, receptor endocytosis, autophagy, DNA damage response and apoptosis. Coordinates actin remodeling through tyrosine phosphorylation of proteins controlling cytoskeleton dynamics like WASF3 (involved in branch formation); ANXA1 (involved in membrane anchoring); DBN1, DBNL, CTTN, RAPH1 and ENAH (involved in signaling); or MAPT and PXN (microtubule-binding proteins). Phosphorylation of WASF3 is critical for the stimulation of lamellipodia formation and cell migration. Involved in the regulation of cell adhesion and motility through phosphorylation of key regulators of these processes such as BCAR1, CRK, CRKL, DOK1, EFS or NEDD9. Phosphorylates multiple receptor tyrosine kinases and more particularly promotes endocytosis of EGFR, facilitates the formation of neuromuscular synapses through MUSK, inhibits PDGFRB-mediated chemotaxis and modulates the endocytosis of activated B-cell receptor complexes. Other substrates which are involved in endocytosis regulation are the caveolin (CAV1) and RIN1. Moreover, ABL1 regulates the CBL family of ubiquitin ligases that drive receptor down-regulation and actin remodeling. Phosphorylation of CBL leads to increased EGFR stability. Involved in late-stage autophagy by regulating positively the trafficking and function of lysosomal components. ABL1

UniProt

Functional data, known sites, GO annotations

Display

Entry

Publications

Feature viewer

Feature table

None

Function

Names & Taxonomy

Subcellular location

Pathology & Biotech

PTM / Processing

Expression

Interaction

Structure

Family & Domains

Sequences (2+)

Similar proteins

Cross-references

Entry information

Sites

Feature key	Position(s)	Description	Actions	Graphical view	Len
Binding site ⁱ	271	ATP			
Active site ⁱ	363	Proton acceptor PROSITE-ProRule annotation			

Regions

Feature key	Position(s)	Description	Actions	Graphical view	Len
Nucleotide binding ⁱ	248 – 256	ATP			
Nucleotide binding ⁱ	316 – 322	ATP			

GO - Molecular functionⁱ

- actin filament binding Source: Ensembl
- actin monomer binding Source: UniProtKB
- ATP binding Source: UniProtKB
- DNA binding Source: UniProtKB
- ephrin receptor binding Source: ARUK-UCL
- kinase activity Source: UniProtKB
- magnesium ion binding Source: UniProtKB
- manganese ion binding Source: UniProtKB
- mitogen-activated protein kinase binding Source: BHF-UCL
- neuropilin binding Source: BHF-UCL
- nicotinate-nucleotide adenylyltransferase activity Source: UniProtKB
- non-membrane spanning protein tyrosine kinase activity Source: UniProtKB
- phosphotyrosine residue binding Source: CAFA
- proline-rich region binding Source: UniProtKB
- protein C-terminus binding Source: UniProtKB
- protein kinase activity Source: MGI

UniProt

Names, Taxonomy

Display

Entry

Publications

Feature viewer

Feature table

None

Function

Names & Taxonomy

Subcellular location

Pathology & Biotech

PTM / Processing

Expression

Interaction

Structure

Family & Domains

Sequences (2+)

Similar proteins

Cross-references

MoonDBⁱ P00519 Predicted

Names & Taxonomyⁱ

Protein namesⁱ

Recommended name:

Tyrosine-protein kinase ABL1 (EC:2.7.10.2  2 Publications 

Alternative name(s):

- Abelson murine leukemia viral oncogene homolog 1
- Abelson tyrosine-protein kinase 1
- Proto-oncogene c-Abl
- p150

Gene namesⁱ

Name:ABL1

Synonyms:ABL, JTK7

Organismⁱ

Homo sapiens (Human)

Taxonomic identifierⁱ

9606 [NCBI]

Taxonomic lineageⁱ

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

Proteomesⁱ

UP000005640 Componentⁱ: Chromosome 9

Organism-specific databases

EuPathDBⁱ

HostDB:ENSG0000097007.17

HGNCⁱ

HGNC:76 ABL1

MIMⁱ

189980 gene

neXtProtⁱ

NX_P00519

Subcellular locationⁱ

UniProt

Variants

Display

Entry

Publications

Feature viewer

Feature table

None

Function

Names & Taxonomy

Subcellular location

Pathology & Biotech

PTM / Processing

Expression

Interaction

Structure

Family & Domains

Sequences (2+)

Similar proteins

Cross-references

Entry information

See also OMIM:b1/b02

Feature key	Position(s)	Description	Actions	Graphical view	Length
Natural variant ⁱ (VAR_079482)	226	Y → C in CHDSKM; increases kinase activity; no effect on protein levels. 1 Publication Corresponds to variant dbSNP:rs1060499547			
Natural variant ⁱ (VAR_079483)	337	A → T in CHDSKM; increases kinase activity; no effect on protein levels. 1 Publication Corresponds to variant dbSNP:rs1060499548			

Mutagenesis

Feature key	Position(s)	Description	Actions	Graphical view	Length
Mutagenesis ⁱ	735	T → A: Abolishes phosphorylation. Loss of binding YWHAS and YWHAZ. Localizes to the nucleus. No effect on kinase activity.	1 Publication		

Sites

Feature key	Position(s)	Description	Actions	Graphical view	Length
Site ⁱ	26 – 27	Breakpoint for translocation to form BCR-ABL oncogene			

Keywords - Diseaseⁱ

Disease mutation, Proto-oncogene

Organism-specific databases

DisGeNET ⁱ	25
MalaCards ⁱ	ABL1
MIM ⁱ	608232 phenotype 617602 phenotype
OpenTargets ⁱ	ENSG00000097007
Orphanet ⁱ	521 Chronic myeloid leukemia

UniProt

Sequence

Display

Sequences (2+)ⁱ

Entry

Publications

Feature viewer

Feature table

None

Function

Names & Taxonomy

Subcellular location

Pathology & Biotech

PTM / Processing

Expression

Interaction

Structure

Family & Domains

Sequences (2+)

Similar proteins

Cross-references

Entry information

Sequence statusⁱ: Complete.

This entry describes **2** isoformsⁱ produced by **alternative splicing**. [Align](#) [Add to basket](#)

This entry has 2 described isoforms and 1 potential isoform that is computationally mapped.ⁱ [Show all](#)

Isoform IA (identifier: P00519-1) [UniParc] [FASTA](#) [Add to basket](#)

This isoform has been chosen as the 'canonical' sequence. All positional information in this entry refers to it. This is also the sequence that appears in the downloadable versions of the entry.

[« Hide](#)

Download

10	20	30	40	50
MLEICLKLVG	CKSKKGGLSSS	SSCYLEEAQ	RPVASDFEPQ	GLSEAARWNS
60	70	80	90	100
KENLLAGPSE	NDPNLFVALY	DFVASGDN	SITKGEKL	RV LGYNHNGEW
110	120	130	140	150
EAQTKNGQGW	VPSNYITPVN	SLEKHWSYHG	PVSRNAAEYL	LSSGINGSFL
160	170	180	190	200
VRESESSPGQ	RSISLRYEGR	VYHYRINTAS	DGKLYVSSES	RFNTLAEVH
210	220	230	240	250
HHSTVADGLI	TTLHYPAPKR	NKPTVYGVSP	NYDKWEMERT	DITMKHKLG
260	270	280	290	300
GOYGEVYEGV	WKKYSLTVAV	KTLKEDTMEV	EEFLKEAAVM	KEIKHPNLVQ
310	320	330	340	350
LLGVCTREPP	FYIITEFMTY	GNLLDYLRREC	NRQEVN	AVVL LYMATQISSA
360	370	380	390	400
MEYLEKKNFI	HRDLAARNCL	VGENHLVKVA	DFGLSRLMTG	DYTAYAHAGAK

Length: 1,130

Mass (Da): 122,873

Last modified: January 24, 2006 - v4

Checksum:ⁱ 85FE6C1C0E483EA2

BLAST [GO](#)

Send to BLAST

UniProt

Structure

Display

Entry

Publications

Feature viewer

Feature table

None

- Function
- Names & Taxonomy
- Subcellular location
- Pathology & Biotech
- PTM / Processing
- Expression
- Interaction
- Structure
- Family & Domains
- Sequences (2+)
- Similar proteins
- Cross-references
- Entry information

Chemistry databases

BindingDBⁱ P00519

Structureⁱ



PDB Entry	Method	Resolution	Chain	Positions	Links
1AB2	NMR		A	120-220	PDBe RCSB P PDBj PDBsui
1ABL	Model		A	65-121	PDBe RCSB P PDBj PDBsui
1AWO	NMR		A	65-119	PDBe RCSB P PDBj PDBsui
1BBZ	X-ray	1.65 Å	A/C/E/G	64-121	PDBe RCSB P PDBj PDBsui
1JU5	NMR		C	62-122	PDBe RCSB P PDBj PDBsui

UniProt

Links to other databases

Display

Entry

Publications

Feature viewer

Feature table

None

✓ Function

✓ Names & Taxonomy

✓ Subcellular location

✓ Pathology & Biotech

✓ PTM / Processing

✓ Expression

✓ Interaction

✓ Structure

✓ Family & Domains

✓ Sequences (2+)

✓ Similar proteins

✓ Cross-references

✓ Entry information

Alternative sequence

Feature key	Position(s)	Description	Actions	Graphical view	Len
Alternative sequence ⁱ (VSP_004957)	1 – 26	MLEIC...SCYLE → MGQQPGKVLGDQRRPSLPAL HFIKGAGKKESSRHHGPHCN VFVEH in isoform IB.	1 Publication	Add BLAST	

Sequence databases

Select the link destinations:	<input checked="" type="radio"/> EMBL ⁱ <input type="radio"/> GenBank ⁱ <input type="radio"/> DDBJ ⁱ	M14752 mRNA Translation: AAA51561.1 X16416 mRNA Translation: CAA34438.1 U07563 Genomic DNA Translation: AAB60394.1 U07563, U07561 Genomic DNA Translation: AAB60393.1 DQ145721 Genomic DNA Translation: AAZ38718.1 AL359092 Genomic DNA No translation available. AL161733 Genomic DNA No translation available. CH471090 Genomic DNA Translation: EAW87948.1 BC117451 mRNA Translation: AAI17452.1 S69223 Genomic DNA Translation: AAD14034.1
CCDS ⁱ		CCDS35165.1 [P00519-2] CCDS35166.1 [P00519-1]
PIR ⁱ		S08519 TVHUA
RefSeq ⁱ		NP_005148.2, NM_005157.5 [P00519-1] NP_009297.2, NM_007313.2 [P00519-2]
UniGene ⁱ		Hs.431048

Genome annotation databases

Ensembl ⁱ	ENST00000318560; ENSP00000323315; ENSG00000097007 [P00519-1] ENST00000372348; ENSP00000361423; ENSG00000097007 [P00519-2]
GeneID ⁱ	25
KEGG ⁱ	hsa:25
Uniprot ⁱ	UC004hv4 human P00519-1

Uniprot

- comprehensive, high-quality and freely accessible resource of protein sequence and functional information.
- Watch videos for Uniprot
- <https://www.youtube.com/watch?v=x9GNm2DLP-U>
- <https://www.youtube.com/watch?v=JjdLtoaNpf4>

Uniprot activity

- GO to Uniprot <https://www.uniprot.org/>
- Search for your protein
 - Choose only reviewed, human proteins
 - What does “reviewed” mean?
 - Download the list of proteins (Uniprot IDs). This time pay attention: Make sure you focus on your protein of interest.
 - For the top entry (your protein of interest),
 - What is the UniProt ID?
 - note active site, binding site or other functionally important residues
 - Note natural variants (Feature viewer)
 - Are these natural variants close to the active site in **sequence**?
 - Find out more about your protein. What is its biological function? What does it do?
 - GO terms give you information about molecular function, cellular component, biological process. You can read more about them here:
<http://geneontology.org/>
 - Report the GO terms for your protein.