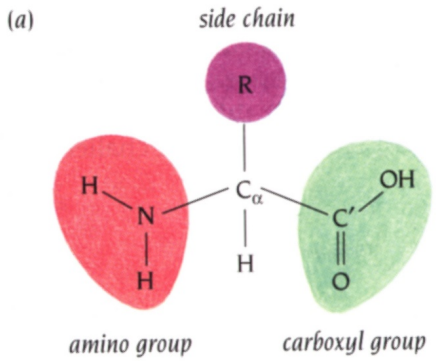
Protein Structure

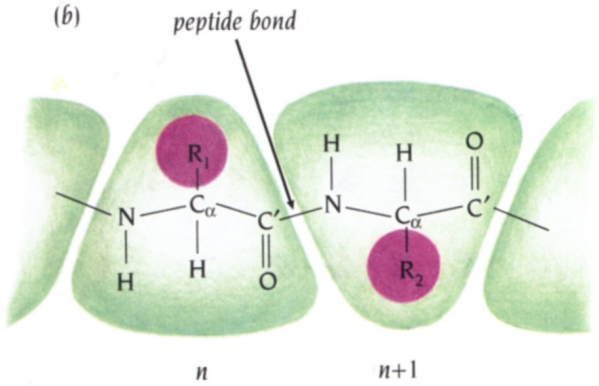
A quick reminder…

Geoff Barton

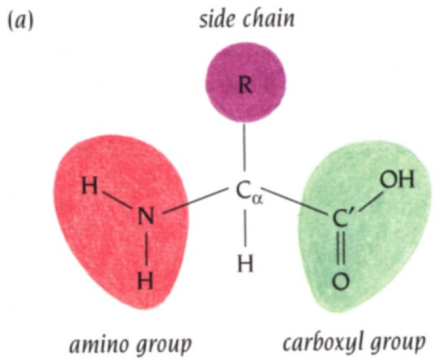
Amino acid	**Three letter code**	**One letter code**	**Mnemonic**
alanine	Ala	A	
arginine	Arg	R	**R**ginine
asparagine	Asn	N	asparagi**N**e
aspartic acid	Asp	D	aspar**D**ic
cysteine	Cys	C	
glutamic acid	Glu	E	glutamat**E**
glutamine	Gln	Q	**Q**tamine
glycine	Gly	G	
histidine	His	H	
isoleucine	Ile	I	
leucine	Leu	L	
lysine	Lys	K	**K** is the letter before **L**
methionine	Met	M	
phenylalanine	Phe	F	**F**enylalanine
proline	Pro	P	
serine	Ser	S	
threonine	Thr	T	
tryptophan	Trp/Try	W	t**W**o rings (W has two Vs)
tyrosine	Tyr	Y	t**Y**rosine
valine	Val	V	
asparagine or aspartic acid	Asx	B	A before G, **B** before **Z**
glutamine or glutamic acid	Glx	Z	A before G, **B** before **Z**
any amino acid	Unk	X	



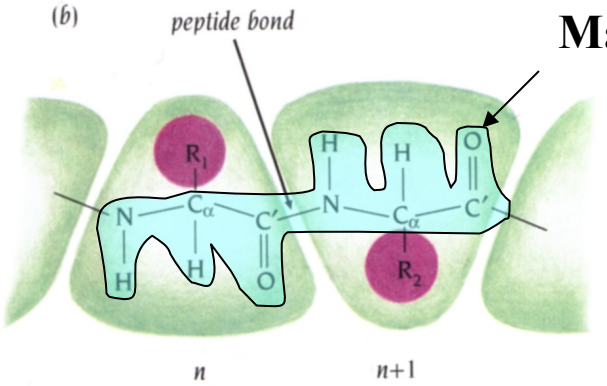
Covalent structure of an amino acid



Peptide units in protein chain.



Covalent structure of an amino acid

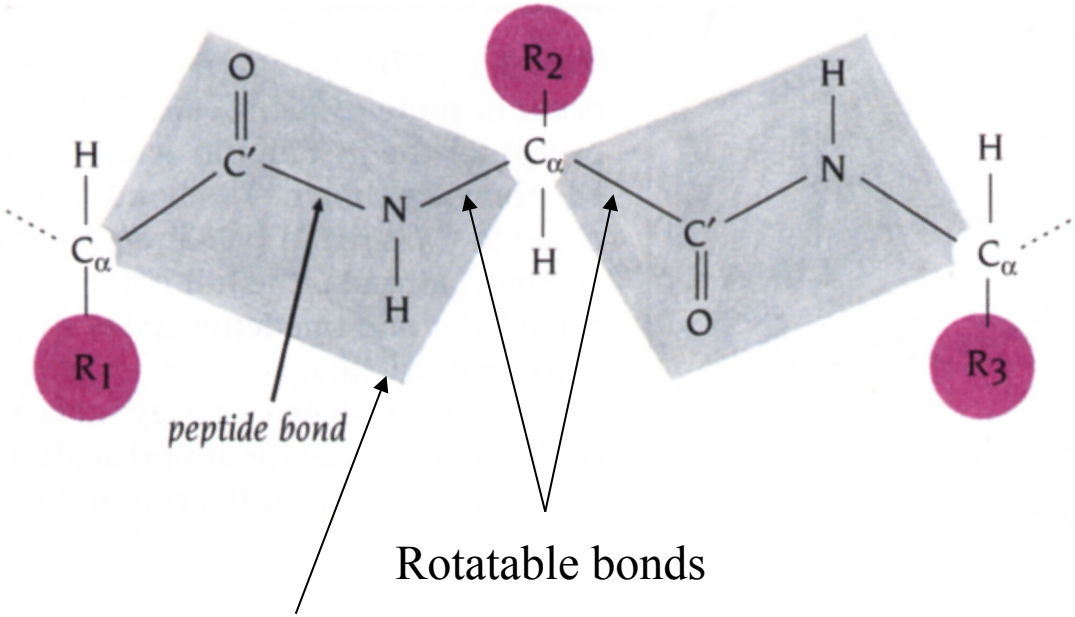


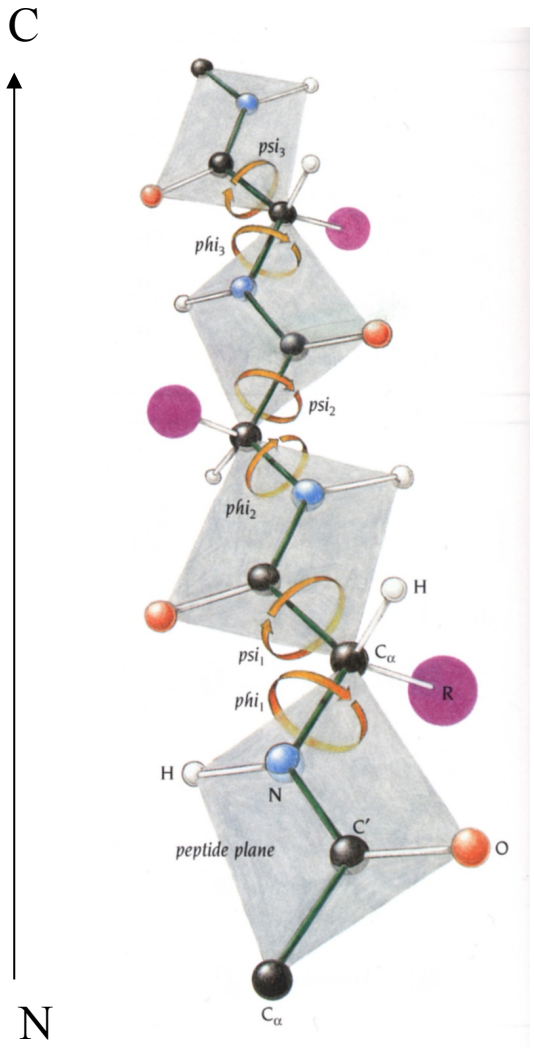
Mainchain atoms

Peptide units in protein chain.

Backbone torsion angles

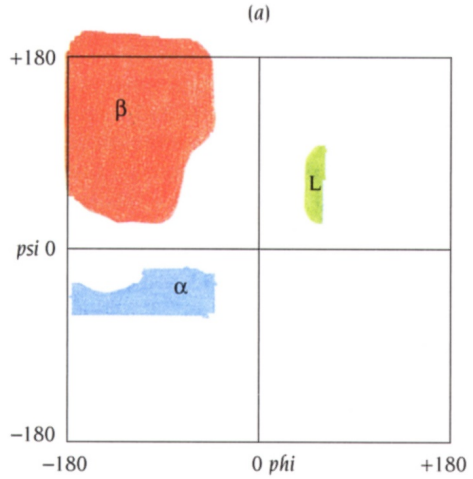
UNIVERSITY OF DUNDEE





The two rotatable mainchain bonds per peptide are called *phi* and *psi*

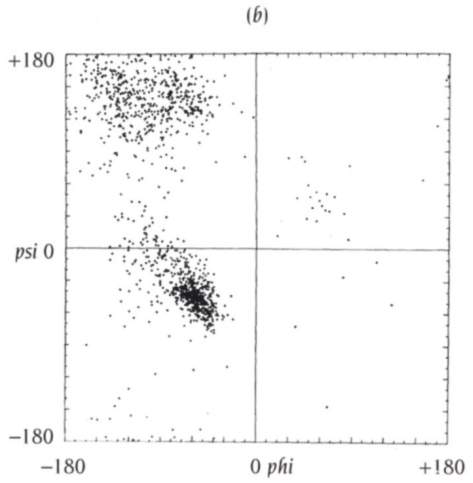
Not all combinations of phi and psi are equally favoured

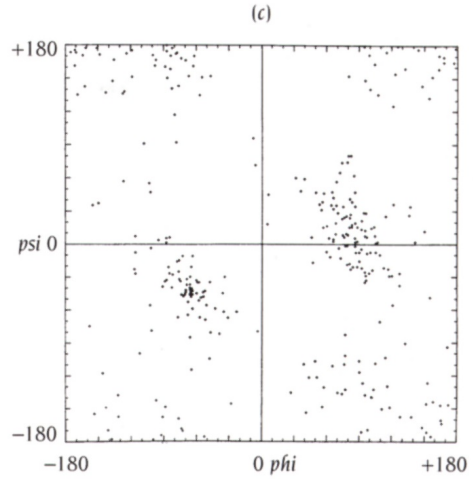


Plot *phi* against *psi* to identify preferred and disallowed backbone conformations.

(a) Allowed regions
(b) Plot for all amino acids except Glycine
(c) Plot for Glycine

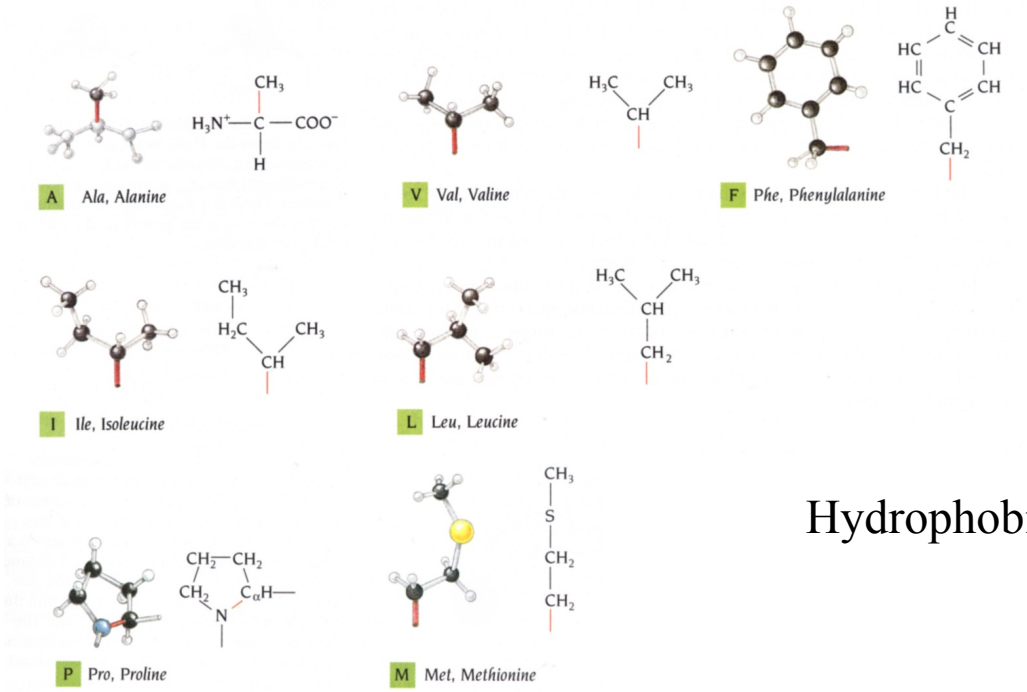
Known as a Ramachandran plot.



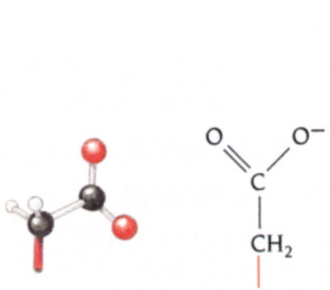


UNIVERSITY OF DUNDEE

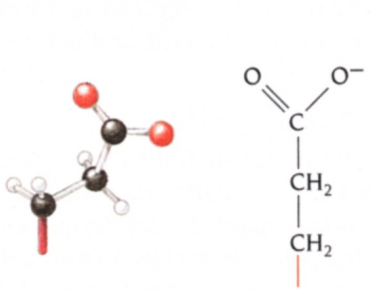
Standard text-book classification of amino acid physico-chemical properties



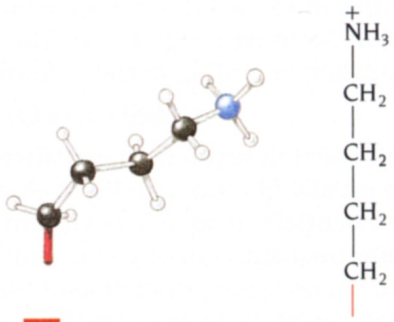
Hydrophobic Amino Acids



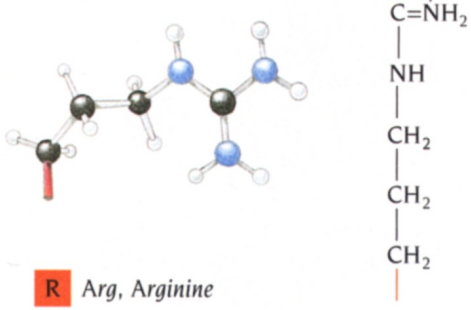
D Asp, Aspartic acid



E Glu, Glutamic acid



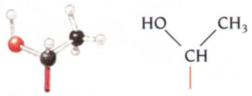
K Lys, Lysine



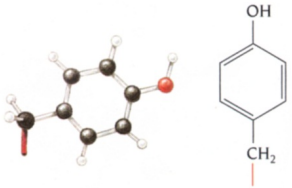
R Arg, Arginine

Charged Amino Acids

S Ser, Serine

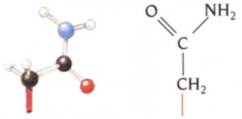


T Thr, Threonine

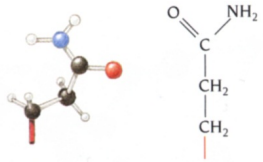


Y Tyr, Tyrosine

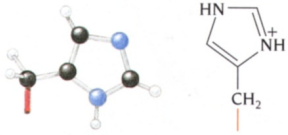
C Cys, Cysteine



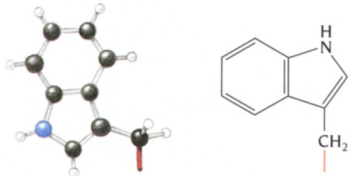
N Asn, Asparagine



Q Gln, Glutamine

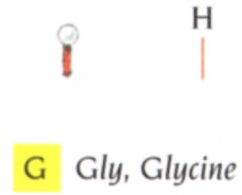


H His, Histidine



W Trp, Tryptophan

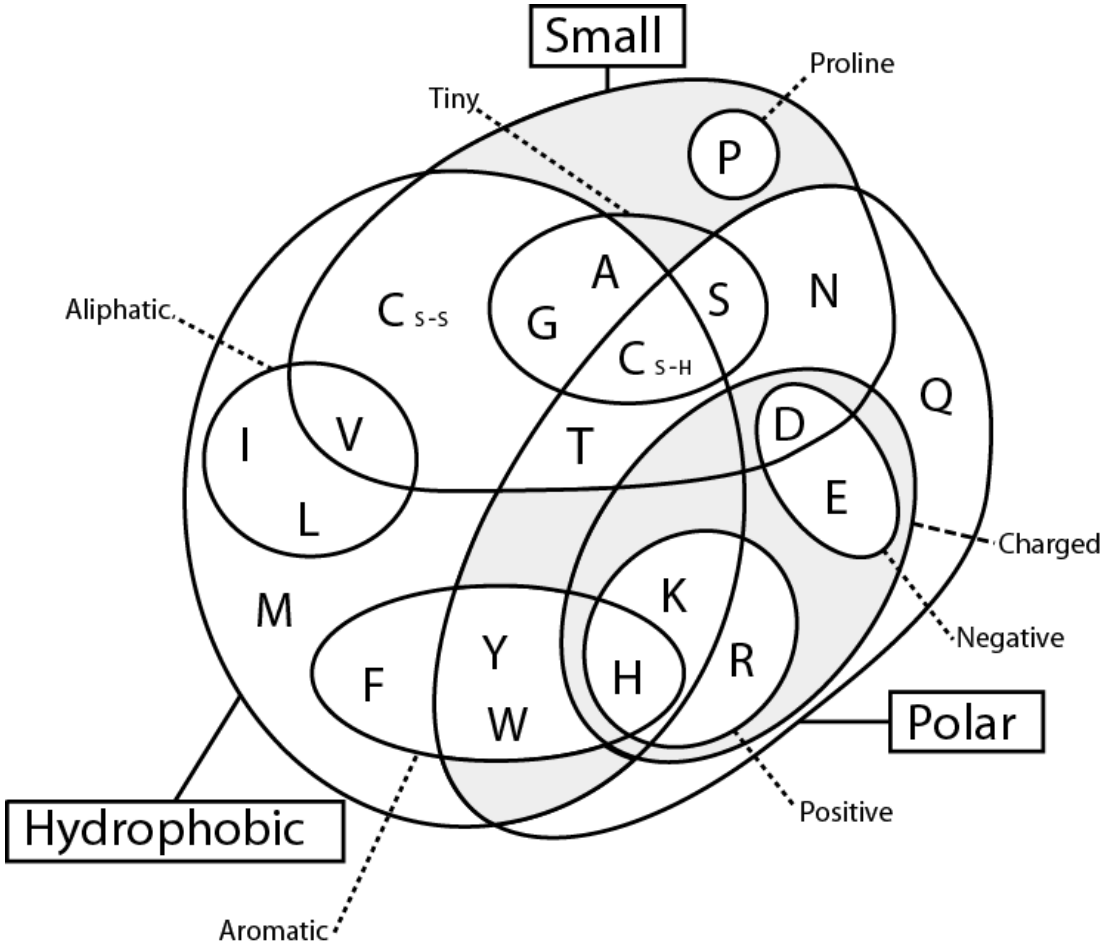
Polar Amino Acids

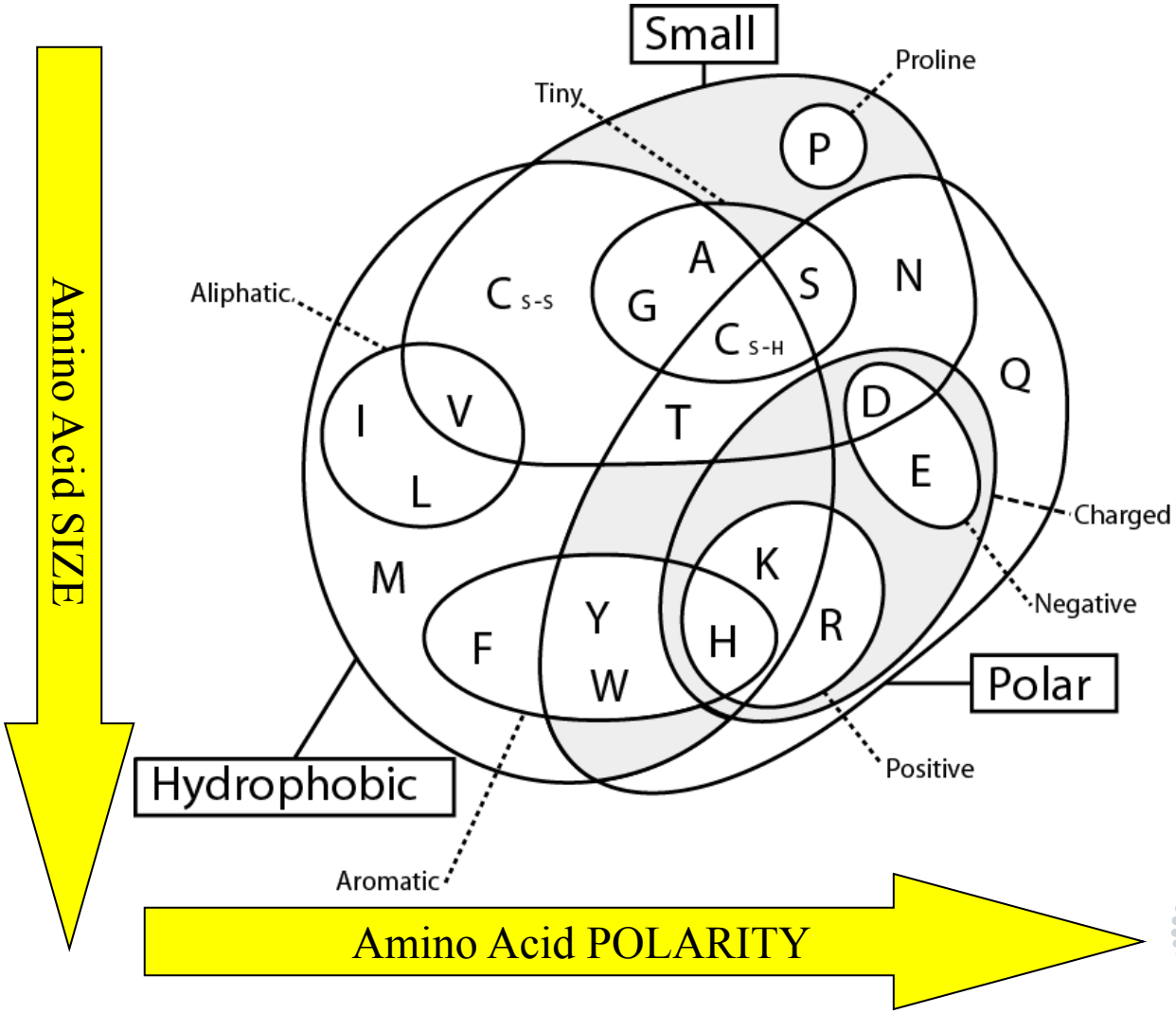


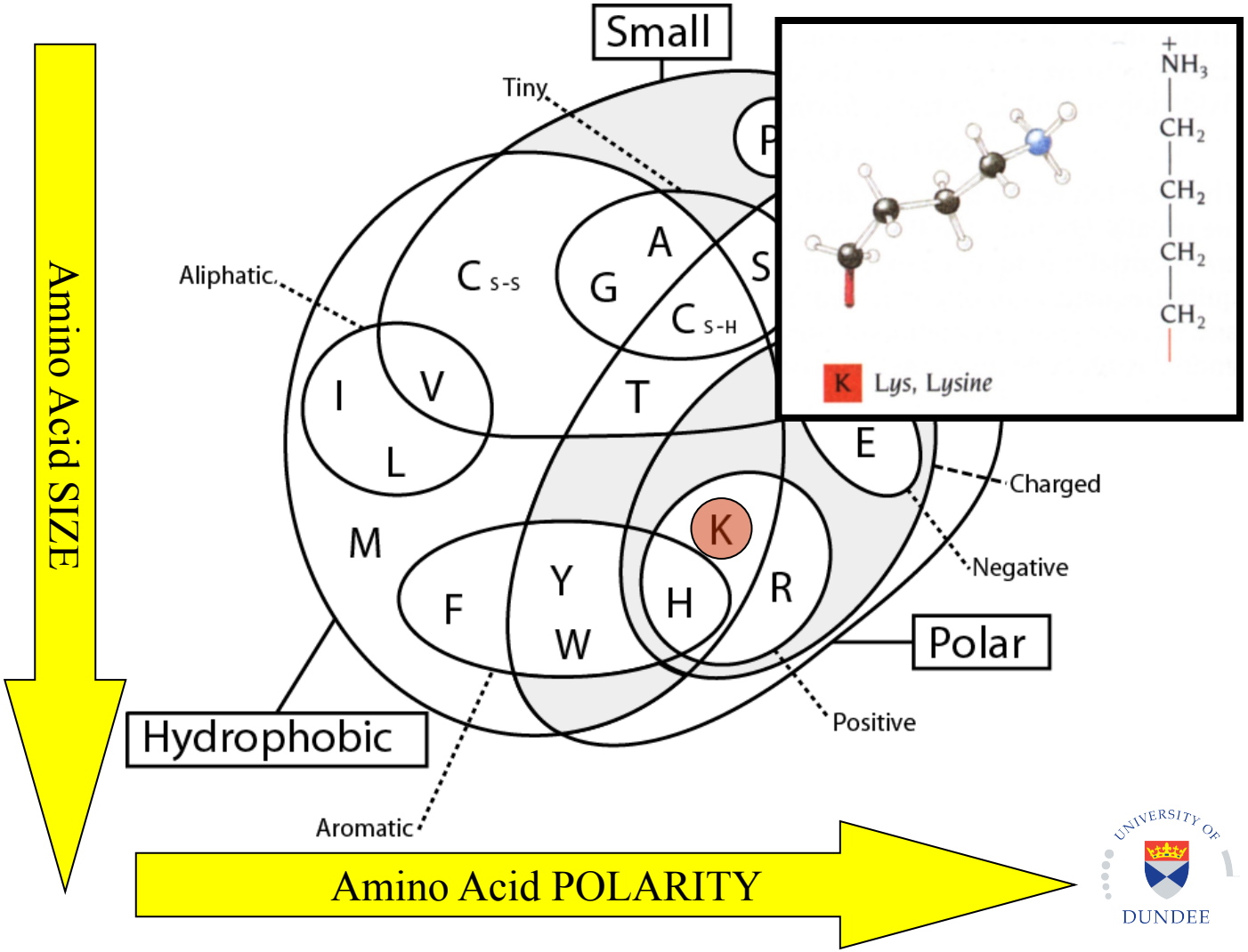
Glycine has no sidechain

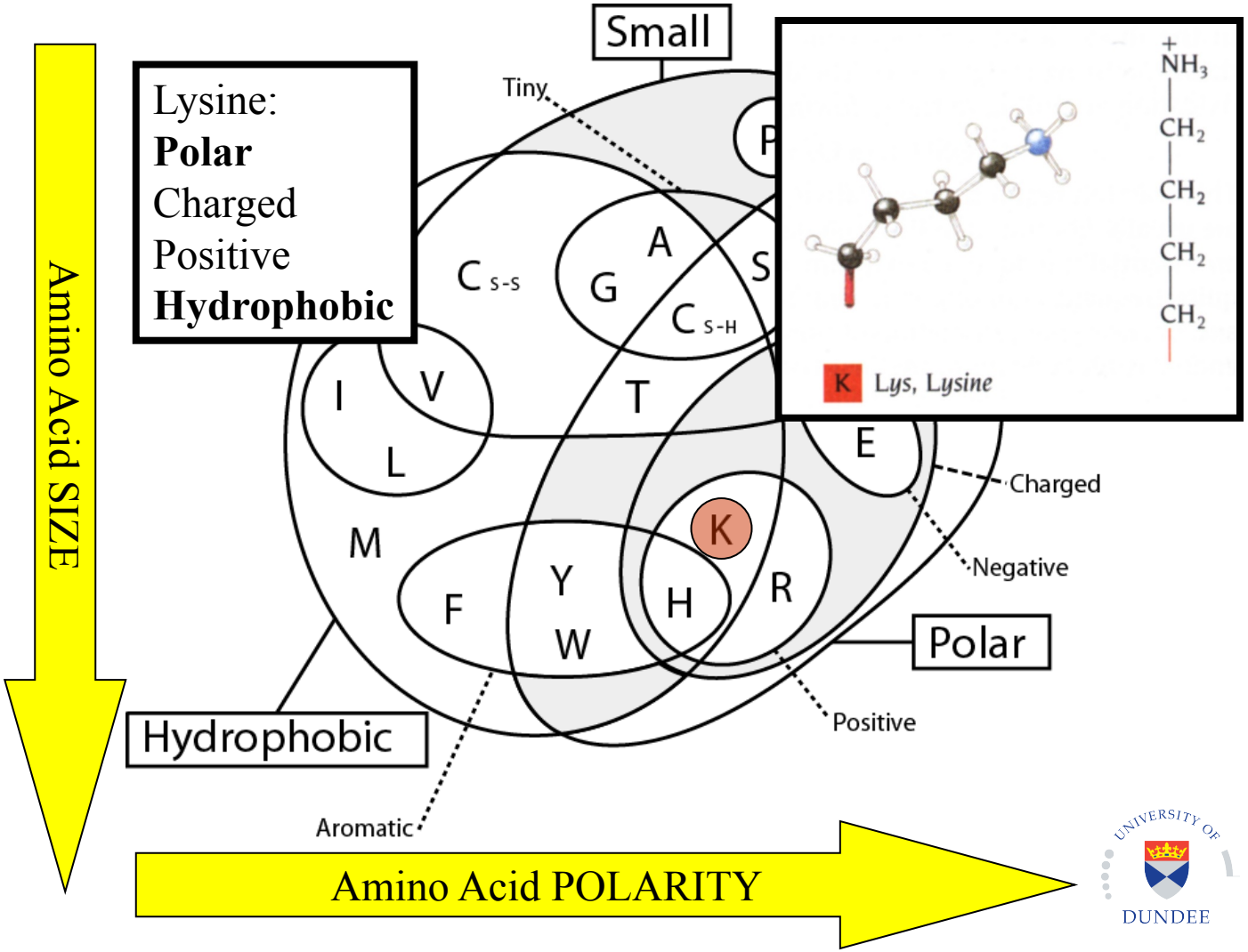
Classification of properties is simplistic

- Amino acids actually exhibit multiple properties
- This is better represented as a Venn diagram

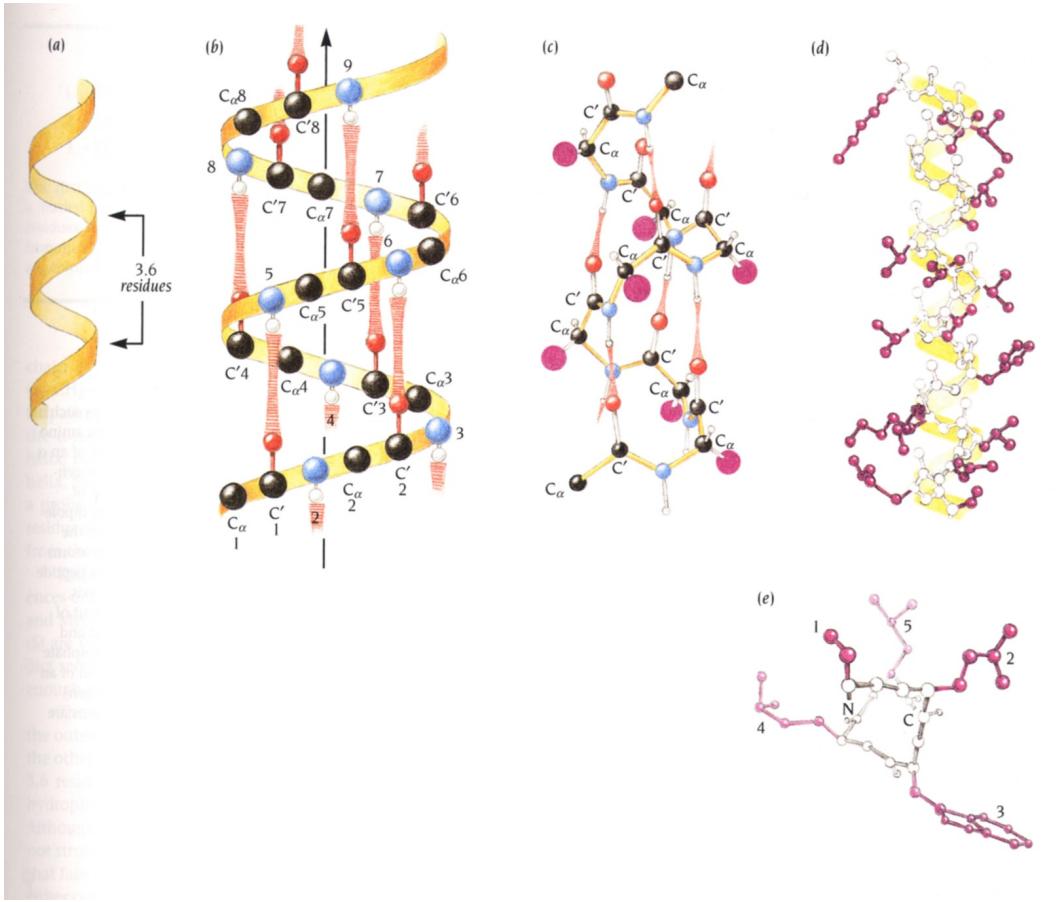








Seondary Structure



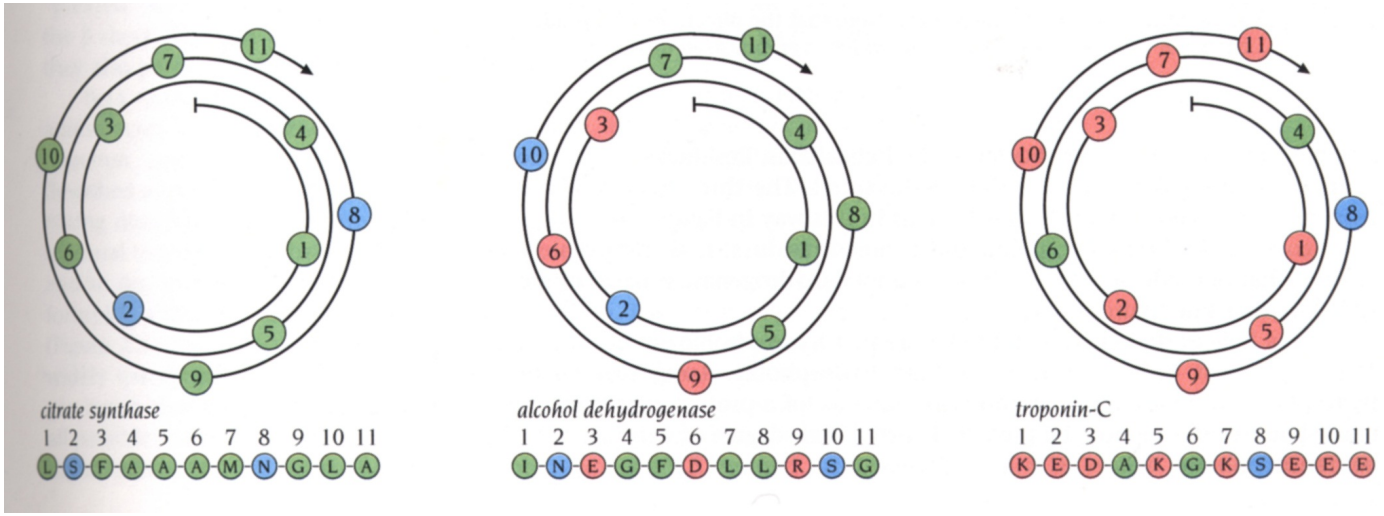
α-helix

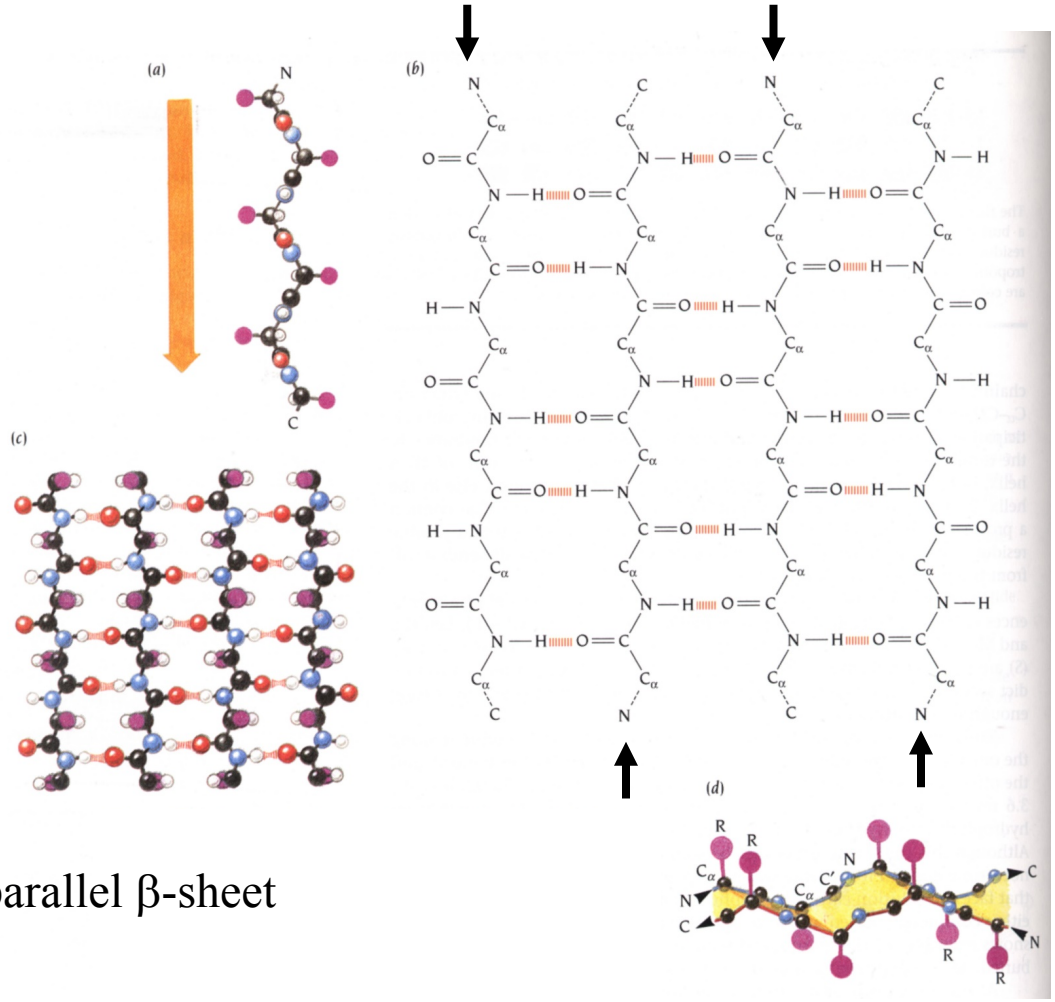
H-bonds stabilise helix structure

Other helix structures also occur with different pitch.

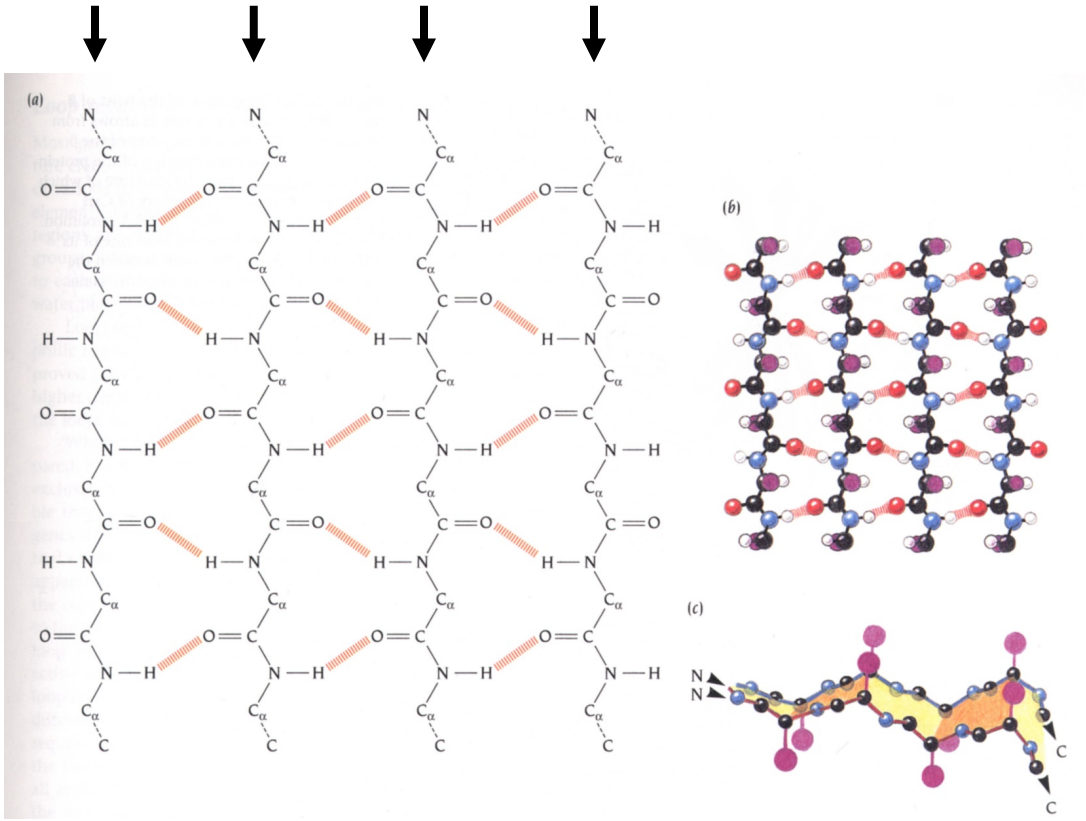
1. Buried helix; 2. part exposed helix; 3. exposed helix

Helical wheel plots to show location of hydrophobic amino acids on face of helix.

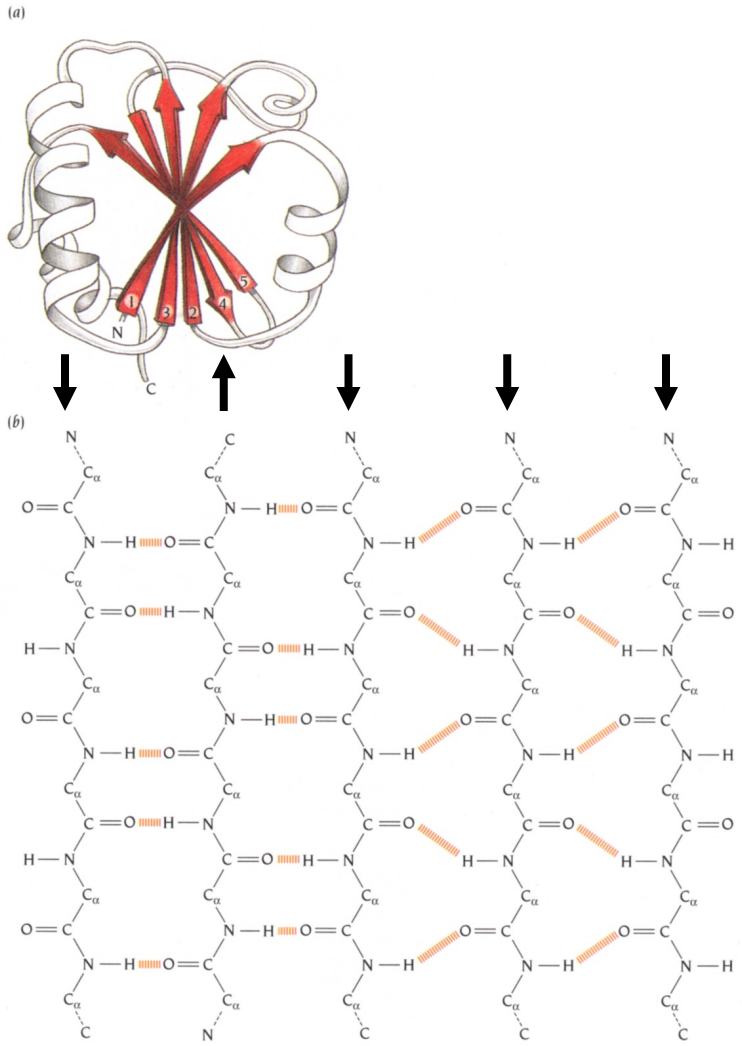




Antiparallel β-sheet



parallel β-sheet

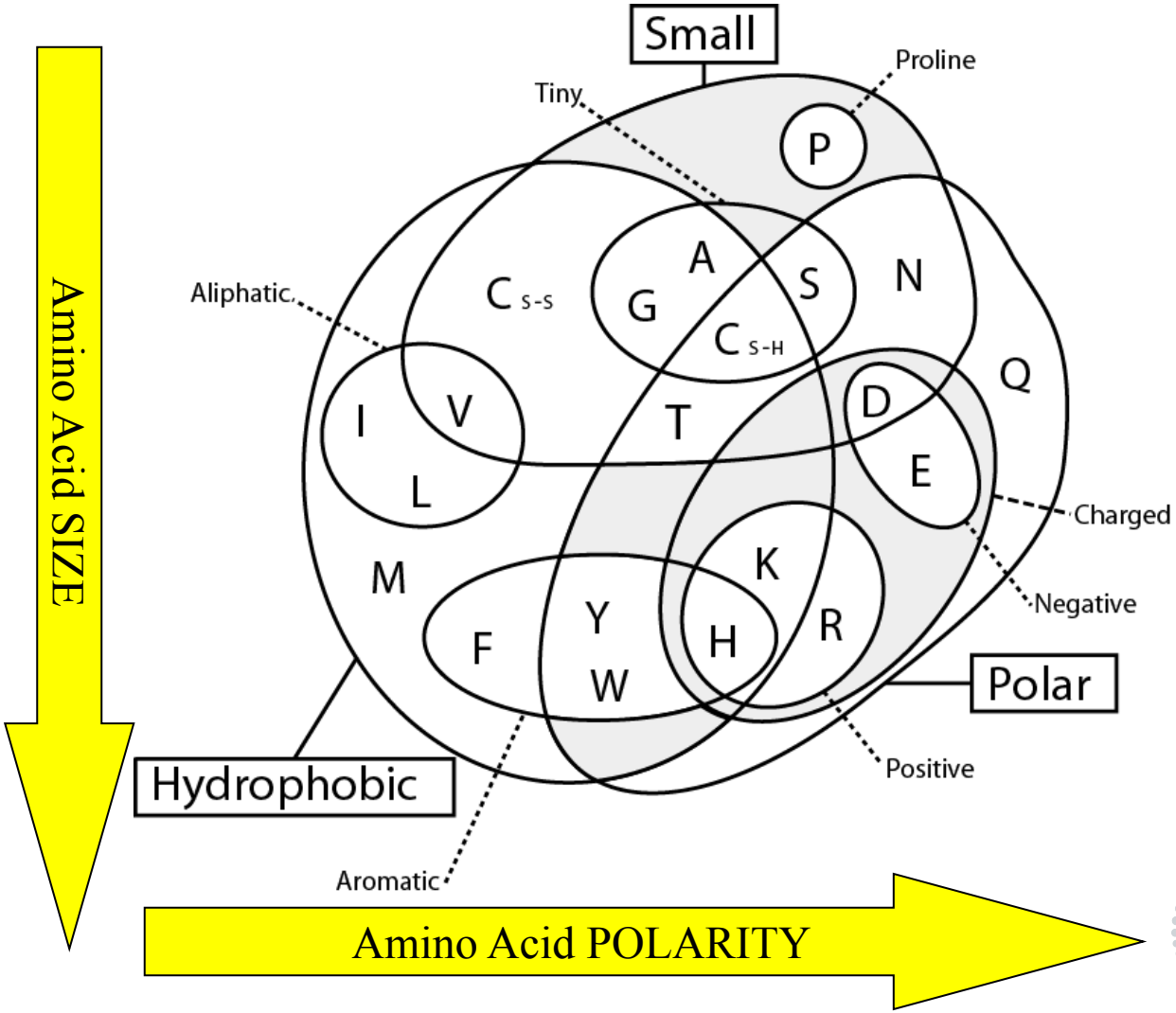


Thioredoxin – protein with mixed sheet.

Glycine and Proline

Amino acids with special effects on protein three-dimensional structure

UNIVERSITY OF DUNDEE



Glycine

- Special properties
- “tiny” due to lack of sidechain
- Can occupy more of the Ramachandran plot due to lack of steric clashes from side-chain.

- Has special role in tight turns in protein structure

UNIVERSITY OF DUNDEE

Proline

- Proline – strictly is not an amino acid due to cyclic structure.
- Cannot form main-chain hydrogen bonds.
- Has only one rotatable mainchain bond.
- Tends to disrupt a helix
- Forms “bulge” in beta sheet

UNIVERSITY OF DUNDEE

Where to find protein structure files

The World-Wide Protein Data Bank (PDB)

PDB started in 1971 and is the oldest continuously supported databank in molecular biology

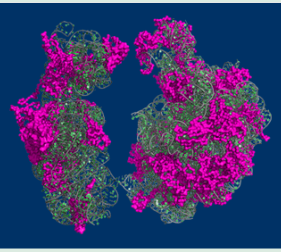
WORLDWIDE PDB PROTEIN DATA BANK

VALIDATION | DEPOSITION | DATA DICTIONARIES | DOCUMENTATION | TASK FORCES | STATISTICS | ABOUT

Since 1971, the Protein Data Bank archive (PDB) has served as the single repository of information about the 3D structures of proteins, nucleic acids, and complex assemblies.

The Worldwide PDB (wwPDB) organization manages the PDB archive and ensures that the PDB is freely and publicly available to the global community.

Learn more about PDB **HISTORY** and **FUTURE**.



Validate Structure
or View validation reports

Deposit Structure
All Deposition Resources

Download Archive

wwPDB Members

Each site offers tools for searching, visualizing, and analyzing PDB data:

PDBj

› **Protein Data Bank Japan**

Supports browsing in multiple languages such as Japanese, Chinese, and Korean; SeSAW identifies functionally or evolutionarily conserved motifs by locating and annotating sequence and structural similarities, tools for bioinformaticians, and more.

PDBe

› **Protein Data Bank in Europe**

Rich information about all PDB entries, multiple search and browse facilities, advanced services including PDBePISA, PDBeFold and PDBeMotif, advanced visualisation and validation of NMR and EM structures, tools for bioinformaticians.

BMRB

› **Biological Magnetic Resonance Bank**

Collects NMR data from any experiment and captures assigned chemical shifts, coupling constants, and peak lists for a variety of macromolecules; contains derived annotations such as hydrogen exchange rates, pKa values, and relaxation parameters.

RCSB PDB

› **Research Collaboratory for Structural Bioinformatics Protein Data Bank**

Simple and advanced searching for macromolecules and ligands, tabular reports, specialized visualization tools, sequence-structure comparisons, RCSB PDB Mobile, Molecule of the Month and other educational resources at PDB-101, and more.

wwPDB Resources

Data Dictionaries

› **Macromolecular Dictionary (PDBx/mmCIF)**
› **Small Molecule Dictionary (CCD)**
› **Peptide-like antibiotic and inhibitor molecules (BIRD)**

Annotation

› **Procedures and policies**
› **Improvements for consistency and accuracy**

Community Input: Task Forces and Working Groups

› **Validation Task Forces (X-ray, NMR, 3DEM)**
› **Small Angle Scattering Task Force**
› **PDB/mmCIF Working Group**
› **Hybrid/Integrative Methods Task Force**
› **Ligand Validation Workshop**

PDB Data Growth & Usage Statistics

› **Depositions: by data center, by year, and by depositor location**
› **Downloads: by year for all entries**

Workshops & Symposia

› **Summaries and presentations from past meetings and events**

Information for Journals

› **Policies, procedures, coordination with publishers, and preferred *Instructions to Authors***

Cite wwPDB:

Nature Structural Biology **10**, 980 (2003)
doi: 10.1038/nsb1203-980
More publications

News & Announcements

February 05, 2016

› **Coming soon: Updated X-ray Validation Reports for Archived PDB Structures**

The wwPDB partners are pleased to announce that validation reports for all X-ray crystal structures deposited in the PDB archive will be updated in March.

The updates include:

- New percentile statistics reflecting the state of the PDB archive on December 30th 2015
- Updated versions for component software packages:
  - CCP4 V6.5 (Refmac 5.8.0135)
  - Mogul 2015 (CSD archive as536be)
- Improvements in the way that Mogul analysis of ligand geometry is carried out
- Clearer graphical elements for representing quality of macromolecular chains
- Improvements to make the report text clearer
- Updated **user guide and FAQs**

Read more

January 28, 2016

› **Maintenance Scheduled for wwPDB Deposition & Annotation System Saturday January 30**

From 12:00 noon to 2 pm UTC on Saturday, January 30, the wwPDB Deposition & Annotation System will be down for maintenance. New and in-progress data depositions will be accessible after this brief window.

Read more

January 11, 2016

› **wwPDB Deposition & Annotation System Now Available for NMR and 3DEM Structures**

The wwPDB partners are pleased to announce the launch of a new Deposition & Annotation system that supports structures determined using 3DEM, NMR, and X-ray, neutron and electron crystallography. New entries from all methods can be submitted online (**RCSB PDB** | **PDBe** | **PDBj**).

Read more

All News

News & Announcements

Members:

Download Archive

RCSB PDB ftp | **PDBe ftp** | **PDBj ftp**
Data Download Instructions

Archive Snapshots

RCSB PDB | **PDBj**

Cite wwPDB:

Nature Structural Biology **10**, 980 (2003)
doi: 10.1038/nsb1203-980
More publications

wwPDB:

Experimentally determined three-dimensional structures of proteins, nucleic acids and complex assemblies

Partnership between:

Europe, USA and Japan

EMBL-EBI | Services | Research | Training | About us

Protein Data Bank in Europe

Bringing Structure to Biology

Search | EMsearch

Examples: hemoglobin, BRCA1_HUMAN

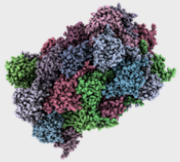
PDBe home | Deposition | PDBe services | PDBe training | Documentation | About PDBe | Share | Feedback

PDBe is the European resource for the collection, organisation and dissemination of data on biological macromolecular structures. Read more about PDBe.

Featured structure

Cellular recycling top 2015 PDB structure

2nd January 2016



Just as old plastic bottles are ground up at the recycling plant to be used to make new ones, so proteins which are damaged or no longer required are sent to one of several different recycling plants in the cell. In 2015, more structures of one of these recycling plants were made public by the wwPDB than any other molecule.

Read more...

Previous featured structures

News

What does the new wwPDB deposition system mean for you?

13 January, 2016

New wwPDB deposition system for NMR and 3DEM

11 January, 2016

Looking back at PDBe in 2015

7 January, 2016

Successful workshop on "3D Segmentations and Transformations"

15 December, 2015

More news

Events

CCP-EM Icknield Workshop on Model Building and Refinement for High Resolution EM Maps

Didcot, UK
2 Mar 2016 to 4 Mar 2016

ELIXIR All Hands 2016

Barcelona, Spain
9 Mar 2016 to 10 Mar 2016

VIZBI 2016

Heidelberg, Germany
9 Mar 2016 to 11 Mar 2016

More events

Publications

Start2Fold: a database of hydrogen/deuterium exchange data on protein folding and stability.

Pancsa R, Varadi M, Tompa P, Vranken WF.
Nucleic Acids Res Volume 44 (2016) p.d429-34

Just a Flexible Linker? The Structural and Dynamic Properties of CBP-ID4 Revealed by NMR Spectroscopy.

Piai A, Calçada EO, Tarenzi T, Grande AD, Varadi M, Tompa P, Felli IC, Pierattelli R.
Biophys J Volume 110 (2016) p.372-381

Popular

- EMsearch
- PDBeFold
- PDBePISA
- Sequence search
- PDBe REST API
- EM resources
- NMR resources
- EMPIAR
- News
- Events
- Training
- Contact us

Latest archive statistics

As of 10 February 2016 the PDB contains 115918 entries (latest PDB entries, chemistry, biology) and EMDB contains 3451 entries (latest map releases, latest header releases, latest updates).

Connect with us

- Facebook
- YouTube
- Twitter
- RSS

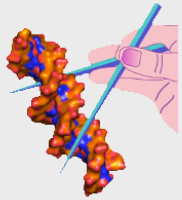
Tweets

Follow

Protein Data Bank @PDBeurope 5h
#HappyValentinesDay from PDBe. This heart-shaped RNA publ @J_A_C_S is from @SzostakLabPDBe.org/4u34
pic.twitter.com/SBEddB6Hkq

Tweet to @PDBeurope

Quips



bZip transcription factors

Interactively explore this remarkable protein in our latest Quips article.

More Quips articles

PDBe site (pdbe.org)

Very modern interface with sophisticated search tools

Easy to refine your search

EMBL-EBI

Services | Research | Training | About us

Protein Data Bank in Europe

Bringing Structure to Biology

Search

Examples: hemoglobin, BRCA1_HUMAN

Search EMDB

Share | Feedback

Search results

Reset | "TIM Barrel"

View basket (0)

Refine query:

Save search | Download | Per page: 10

Entries | Macromolecules | Compounds | Protein families

< 1 2 3 ... 428 > Entry 1 to 10 of 4275 Sort results

− Macromolecules (100+)

- Orotidine 5'-phosphate decarboxylase (141)
- Aldose reductase (116)
- Xylose isomerase (93)
- Triosephosphate isomerase (91)
- Beta-galactosidase (73)
- Tryptophan synthase alpha chain (70)
- Queuine tRNA-ribosyltransferase (66)
- 2-dehydro-3-deoxyphosphooctonate ald... (63)
- Pancreatic alpha-amylase (56)

− Molecule type (1)

- Protein (4275)

+ Interacting macromolecules (98)

+ Interacting compounds (100+)

− Species name (100+)

- Homo sapiens (474)
- Escherichia coli (249)
- Thermotoga maritima (99)
- Saccharomyces cerevisiae (87)
- Escherichia coli K-12 (82)
- Salmonella enterica subsp. enterica ser... (64)
- Bacillus subtilis (63)
- Methanothermobacter thermautotrophic... (62)
- Zymomonas mobilis (60)

+ Experimental methods (4)

− Assembly composition (5)

- protein structure (3685)
- polysaccharide/protein complex (370)
- protein/protein complex (226)
- RNA/protein complex (5)
- DNA/protein complex (4)

+ Assembly polymer count (12)

+ Homo / hetero assembly (2)

+ Journal (100+)

+ Resolution distribution

+ EC number (100+)

+ Cell component (100+)

+ Entry status (1)

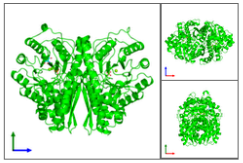
2akm Fluoride Inhibition of Enolase: Crystal Structure of the Inhibitory Complex

Qin J, Chai G, Brewer JM, Lovelace LL
Biochemistry (2006) [PMID: 16411755]

Source organism: *Homo sapiens*
Assembly composition: protein only structure
Interacting compounds: PO4 MG TRS

Add to basket | Download files

X-ray diffraction
1.92Å resolution
Released: 21 Mar 2006
Model geometry
Fit model/data



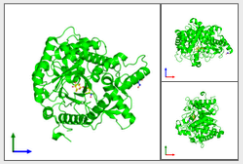
3ai0 Crystal structure of beta-glucosidase from termite Neotermes koshunensis in complex with para-nitrophenyl-beta-D-glucopyranoside

Jeng W-Y, Liu C-I, Wang AH-J
J. Struct. Biol. (2011) [PMID: 20682343]

Source organism: *Neotermes koshunensis*
Assembly composition: protein only structure
Interacting compounds: GOL PNW

Add to basket | Download files

X-ray diffraction
1.4Å resolution
Released: 18 Aug 2010
Model geometry
Fit model/data



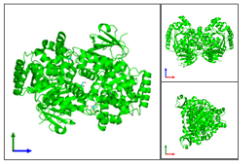
4dyk Crystal structure of an adenosine deaminase from pseudomonas aeruginosa pao1 (target nysgrc-200449) with bound zn

Vetting MW, Toro R, Bhosle R, Wasserman SR, Morisco LL, Sojitra S, Chamala S, Kar A, Lafleur J, Villigas G, Evans B, Hammonds J, Gizzi A, Zencheck WD, Hillerich B, Love J, Seidel RD, Bonanno JB, Raushel FM, Almo SC, New York Structural Genomics Research Consortium (NYSGRC)
To be published

Source organism: *Pseudomonas aeruginosa PAO1*
Assembly composition: protein only structure
Interacting compounds: ZN GOL MG

Add to basket | Download files

X-ray diffraction
2Å resolution
Released: 14 Mar 2012
Model geometry
Fit model/data



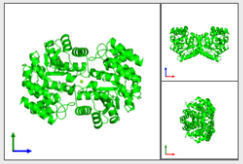
3b4u Crystal structure of dihydrodipicolinate synthase from Agrobacterium tumefaciens str. C58

Zhang R, Xu L, Gu J, Savchenko A, Edwards AM, Joachimiak A, Midwest Center for Structural Genomics (MCSG)
To be published

Source organism: *Agrobacterium fabrum str. C58*
Assembly composition: protein only structure
Interacting compound: MG

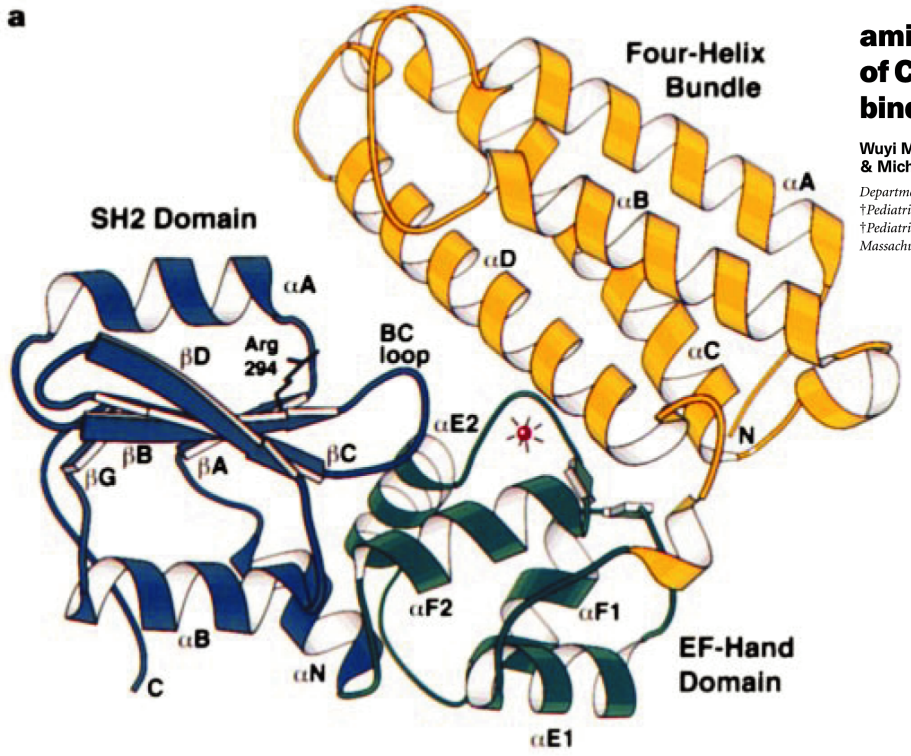
Add to basket | Download files

X-ray diffraction
1.2Å resolution
Released: 04 Dec 2007
Model geometry
Fit model/data



Protein Structure Classification

Organise proteins by “Domain”

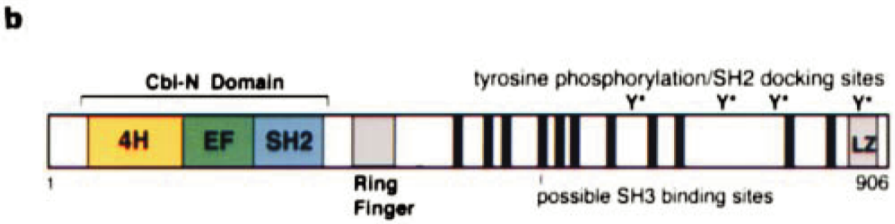


amino-terminal domain of Cbl complexed to its binding site on ZAP-70 kinase

Wuyi Meng*, Sansana Sawasdikosol†, Steven J. Burakoff† & Michael J. Eck*

*Departments of *Biological Chemistry and Molecular Pharmacology and †Pediatrics, Harvard Medical School, and Departments of *Cancer Biology and †Pediatric Oncology, Dana-Farber Cancer Institute, 44 Binney Street, Boston, Massachusetts 02115, USA*

Nature 398,84-90, 1999



PDB: 1b47

Protein Structure Classification

- Proteins can be grouped according to their similarity at different levels of the structural hierarchy
- This helps navigation of protein structure "space" and the assignment of possible function to a protein from its structure or sequence alone

Two major structure classification databases

- **SCOP:** **S**tructural **C**lassification of **P**roteins http://scop.mrc-lmb.cam.ac.uk/scop/
- **SCOPe:** Based on SCOP, but more up to date: http://scop.berkley.edu
- **CATH:** **C**lass **A**rchitecture **T**opology **H**omology http://www.cathdb.info/
- Both databases have similar hierarchies, but slightly different philosophies, so both are useful to look at when studying protein structure.
- The best way to learn about scop(e) and CATH is to browse them…

SCOP

- Classification is of domains. A domain is defined as an independent folding unit if a 3D structure is known of a similar domain on its own.
- SCOP hierarchy: Class, Fold, Superfamily, Family
- I'll illustrate with SCOPe

SCOPe: Structural Classification of Proteins — extended. Release 2.05 (updated 2016-01-14, stable release February 2015)

Browse | Stats & History | ASTRAL Subsets | Downloads | Related Resources | References | Help | About

Welcome to SCOPe!
SCOPe is a database developed at the Berkeley Lab and UC Berkeley to extend the development and maintenance of SCOP.
SCOP was conceived at the MRC Laboratory of Molecular Biology, and developed in collaboration with researchers in Berkeley.
Work on SCOP (version 1) concluded in June 2009 with the release of SCOP 1.75.
SCOPe classifies many newer structures through a combination of automation and manual curation, and corrects some errors in SCOP,
aiming to have the same accuracy as the hand-curated SCOP releases. **SCOPe** also incorporates and updates the ASTRAL database.
For prior releases, click on the Stats & History tab above. For more info, click on the About tab above.

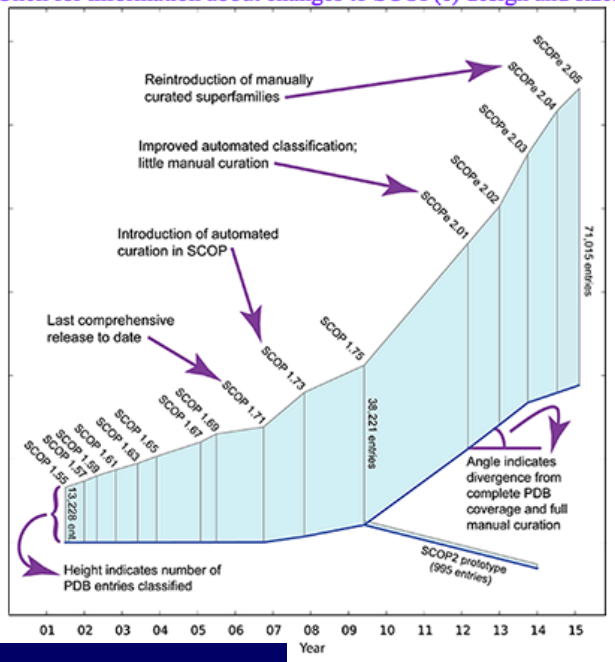
News:
2015-09-22: We recently published a paper about how **SCOP** and **SCOPe** have been used in recent studies [PDF].
2015-11-25: The **SCOPe** website now supports SSL.
2016-01-14: New PDB entries were added in a periodic update; for more info on these updates click on the Help tab above.

Search **SCOPe** (example): Search

Classes in SCOPe 2.05:

1. a: All alpha proteins [46456] (286 folds)
2. b: All beta proteins [48724] (176 folds)
3. c: Alpha and beta proteins (a/b) [51349] (148 folds)
4. d: Alpha and beta proteins (a+b) [53931] (381 folds)
5. e: Multi-domain proteins (alpha and beta) [56572] (68 folds)
6. f: Membrane and cell surface proteins and peptides [56835] (57 folds)
7. g: Small proteins [56992] (92 folds)
8. h: Coiled coil proteins [57942] (7 folds)
9. i: Low resolution protein structures [58117] (25 folds)
10. j: Peptides [58231] (129 folds)
11. k: Designed proteins [58788] (44 folds)

Click for information about changes to SCOP(e) design and size.



Top of the SCOP hierarchy
SCOP *Classes*

Proteins in the same SCOP *Class* normally share similar overall secondary structure compositions

SCOPe: Structural Classification of Proteins — extended. Release 2.05 (updated 2016-01-14, stable release February 2015)

Browse | Stats & History | ASTRAL Subsets | Downloads | Related Resources | References | Help | About

Search **SCOPe** (example): Search

Lineage for Class c: Alpha and beta proteins (a/b)

1. Root: SCOPe 2.05
2. Class c: Alpha and beta proteins (a/b) [51349] (148 folds)

Folds:

1. c.1: TIM beta/alpha-barrel [51350] (33 superfamilies)
   contains parallel beta-sheet barrel, closed; n=8, S=8; strand order 12345678
   the first seven superfamilies have similar phosphate-binding sites
2. c.2: NAD(P)-binding Rossmann-fold domains [51734] (1 superfamily)
   core: 3 layers, a/b/a; parallel beta-sheet of 6 strands, order 321456
   The nucleotide-binding modes of this and the next two folds/superfamilies are similar
3. c.3: FAD/NAD(P)-binding domain [51904] (1 superfamily)
   core: 3 layers, b/b/a; central parallel beta-sheet of 5 strands, order 32145; top antiparallel beta-sheet of 3
4. c.4: Nucleotide-binding domain [51970] (1 superfamily)
   3 layers: a/b/a; parallel beta-sheet of 5 strands, order 32145; Rossmann-like
5. c.5: MurCD N-terminal domain [51983] (1 superfamily)
   3 layers: a/b/a; parallel beta-sheet of 5 strands, order 32145; incomplete Rossmann-like fold; binds UDP
6. c.6: 7-stranded beta/alpha barrel [51988] (3 superfamilies)
   variant of beta/alpha barrel; parallel beta-sheet barrel, closed, n=7, S=8; strand order 1234567; some me
7. c.7: PFL-like glycyl radical enzymes [51997] (1 superfamily)
   contains: barrel, closed; n=10, S=10; accommodates a hairpin loop inside the barrel
8. c.8: The "swivelling" beta/beta/alpha domain [52008] (10 superfamilies)
   3 layers: b/b/a; the central sheet is parallel, and the other one is antiparallel; there are some variations in
   this domain is thought to be mobile in most multi-domain proteins known to contain it
9. c.9: Barstar-like [52037] (2 superfamilies)
   2 layers, a/b; parallel beta-sheet of 3 strands, order 123
10. c.10: Leucine-rich repeat, LRR (right-handed beta-alpha superhelix) [52046] (3 superfamilies)
    2 curved layers, a/b; parallel beta-sheet; order 1234...N; there are sequence similarities between different
11. c.12: Ribosomal proteins L15p and L18e [52079] (1 superfamily)
    core: three turns of irregular (beta-beta-alpha)n superhelix
12. c.13: SpoIIaa-like [52086] (2 superfamilies)
    core: 4 turns of a (beta-alpha)n superhelix
13. c.14: ClpP/crotonase [52095] (1 superfamily)
    core: 4 turns of (beta-beta-alpha)n superhelix
14. c.15: BRCT domain [52112] (1 superfamily)
    3 layers, a/b/a; core: parallel beta-sheet of 4 strands, order 2134
15. c.16: Lumazine synthase [52120] (1 superfamily)
    3 layers, a/b/a; core: parallel beta-sheet of 4 strands, order 2134
16. c.17: Caspase-like [52128] (1 superfamily)
    3 layers, a/b/a; core: mixed beta-sheet of 6 strands, order 213456, strand 6 is antiparallel to the rest
17. c.18: Uracil-DNA glycosylase-like [52140] (1 superfamily)
    3 layers, a/b/a; core: parallel beta-sheet of 4 strands, order 2134
18. c.19: FabD/lysophospholipase-like [52150] (1 superfamily)
    core: 3 layers, a/b/a; mixed beta-sheet of 6 strands, order 432156; strand 4 is antiparallel to the rest
19. c.20: Initiation factor IF2/eIF5b, domain 3 [52155] (1 superfamily)
    3 layers, a/b/a; core: parallel beta-sheet of 4 strands, order 2134

SCOP *Classes* are divided into *Folds*.

Proteins in the same SCOP *Fold* share similar arrangements (topologies) of secondary structures.

SCOPe: Structural Classification of Proteins — extended. Release 2.05 (updated 2016-01-14, stable release February 2015)

Browse | Stats & History | ASTRAL Subsets | Downloads | Related Resources | References | Help | About

Search **SCOPe** (example): Search

Lineage for Fold c.1: TIM beta/alpha-barrel

1. Root: SCOPe 2.05
2. Class c: Alpha and beta proteins (a/b) [51349] (148 folds)
3. Fold c.1: TIM beta/alpha-barrel [51350] (33 superfamilies)
   contains parallel beta-sheet barrel, closed; n=8, S=8; strand order 12345678
   the first seven superfamilies have similar phosphate-binding sites

Superfamilies:

1. c.1.1: Triosephosphate isomerase (TIM) [51351] (2 families)
2. c.1.2: Ribulose-phoshate binding barrel [51366] (7 families)
3. c.1.3: Thiamin phosphate synthase [51391] (2 families)
   automatically mapped to Pfam PF02581
4. c.1.4: FMN-linked oxidoreductases [51395] (2 families)
5. c.1.5: Inosine monophosphate dehydrogenase (IMPDH) [51412] (2 families)
   The phosphate moiety of substrate binds in the 'common' phosphate-binding site
6. c.1.6: PLP-binding barrel [51419] (2 families)
   circular permutation of the canonical fold: begins with an alpha helix and ends with a beta-strand
7. c.1.7: NAD(P)-linked oxidoreductase [51430] (2 families)
8. c.1.8: (Trans)glycosidases [51445] (15 families)
9. c.1.9: Metallo-dependent hydrolases [51556] (19 families)
   the beta-sheet barrel is similarly distorted and capped by a C-terminal helix
   has transition metal ions bound inside the barrel
10. c.1.10: Aldolase [51569] (9 families)
    Common fold covers whole protein structure
11. c.1.11: Enolase C-terminal domain-like [51604] (3 families)
    binds metal ion (magnesium or manganese) in conserved site inside barrel
    N-terminal alpha+beta domain is common to this superfamily
12. c.1.12: Phosphoenolpyruvate/pyruvate domain [51621] (8 families)
13. c.1.13: Malate synthase G [51645] (1 family)
14. c.1.14: RuBisCo, C-terminal domain [51649] (2 families)
    automatically mapped to Pfam PF00016
15. c.1.15: Xylose isomerase-like [51658] (8 families)
    different families share similar but non-identical metal-binding sites
16. c.1.16: Bacterial luciferase-like [51679] (5 families)
    consists of clearly related families of somewhat different folds
17. c.1.17: Nicotinate/Quinolinate PRTase C-terminal domain-like [51690] (3 families)
    incomplete beta/alpha barrel with parallel beta-sheet of 7 strands
18. c.1.18: PLC-like phosphodiesterases [51695] (4 families)
19. c.1.19: Cobalamin (vitamin B12)-dependent enzymes [51703] (4 families)

SCOP *Folds* are divided into *Superfamilies*:

Proteins in the same SCOP *Superfamily* share similar folds and are likely to have a common evolutionary ancestor and common function

SCOPe: Structural Classification of Proteins — extended. Release 2.05 (updated 2016-01-14, stable release February 2015)

Browse | Stats & History | ASTRAL Subsets | Downloads | Related Resources | References | Help | About

Search **SCOPe** (example): Search

Lineage for Superfamily c.1.2: Ribulose-phoshate binding barrel

1. Root: SCOPe 2.05
2. Class c: Alpha and beta proteins (a/b) [51349] (148 folds)
3. Fold c.1: TIM beta/alpha-barrel [51350] (33 superfamilies)
   contains parallel beta-sheet barrel, closed; n=8, S=8; strand order 12345678
   the first seven superfamilies have similar phosphate-binding sites
4. Superfamily c.1.2: Ribulose-phoshate binding barrel [51366] (7 families) *S*

Families:

1. c.1.2.1: Histidine biosynthesis enzymes [51367] (3 protein domains)
   structural evidence for the gene duplication within the barrel fold
   automatically mapped to Pfam PF00977
2. c.1.2.2: D-ribulose-5-phosphate 3-epimerase [51372] (1 protein)
   automatically mapped to Pfam PF00834
3. c.1.2.3: Decarboxylase [51375] (4 protein domains)
4. c.1.2.4: Tryptophan biosynthesis enzymes [51381] (4 protein domains)
5. c.1.2.5: NanE-like [117362] (1 protein)
   Pfam PF04131
6. c.1.2.6: PdxS-like [141755] (2 protein domains)
   Pfam PF01680; SOR/SNZ
7. c.1.2.0: automated matches [191350] (1 protein)
   not a true family

More info for Superfamily c.1.2: Ribulose-phoshate binding barrel

Timeline for Superfamily c.1.2: Ribulose-phoshate binding barrel:

- Superfamily c.1.2: Ribulose-phoshate binding barrel first appeared (with stable ids) in SCOP 1.55
- Superfamily c.1.2: Ribulose-phoshate binding barrel appears in SCOPe 2.04

SCOPe
scope@compbio.berkeley.edu

SCOP *Superfamilies* are divided into *Families*:

Proteins in the same SCOP *Family* share similar sequences >30% Identity and are highly likely to have similar functions.

SCOPe: Structural Classification of Proteins — extended. Release 2.05 (updated 2016-01-14, stable release February 2015)

Browse | Stats & History | ASTRAL Subsets | Downloads | Related Resources | References | Help | About

Search **SCOPe** (example): Search

Lineage for Family c.1.2.1: Histidine biosynthesis enzymes

1. Root: SCOPe 2.05
2. Class c: Alpha and beta proteins (a/b) [51349] (148 folds)
3. Fold c.1: TIM beta/alpha-barrel [51350] (33 superfamilies)
   contains parallel beta-sheet barrel, closed; n=8, S=8; strand order 12345678
   the first seven superfamilies have similar phosphate-binding sites
4. Superfamily c.1.2: Ribulose-phoshate binding barrel [51366] (7 families) $\mathcal{S}$
5. Family c.1.2.1: Histidine biosynthesis enzymes [51367] (3 protein domains)
   structural evidence for the gene duplication within the barrel fold
   automatically mapped to Pfam PF00977

Protein Domains:

1. Cyclase subunit (or domain) of imidazoleglycerolphosphate synthase HisF [51370] (4 species)
   1. Species Baker's yeast (Saccharomyces cerevisiae), His7 [TaxId:4932] [69379] (4 PDB entries)
   2. Species Pyrobaculum aerophilum [TaxId:13773] [69380] (1 PDB entry)
   3. Species Thermotoga maritima [TaxId:2336] [51371] (4 PDB entries)
   4. Species Thermus thermophilus [TaxId:274] [82237] (1 PDB entry)
2. Phosphoribosylformimino-5-aminoimidazole carboxamide ribotite isomerase HisA [51368] (2 species)
   1. Species Streptomyces coelicolor [TaxId:1902] [141744] (1 PDB entry)
      Uniprot P16250 2-240
   2. Species Thermotoga maritima [TaxId:2336] [51369] (2 PDB entries)
3. automated matches [190186] (9 species)
   not a true protein
   1. Species Arthrobacter aurescens [TaxId:43663] [260126] (1 PDB entry)
   2. Species Corynebacterium efficiens [TaxId:152794] [226694] (1 PDB entry)
   3. Species Mycobacterium tuberculosis [TaxId:83332] [189657] (4 PDB entries)
   4. Species Streptomyces coelicolor [TaxId:1902] [189237] (2 PDB entries)
   5. Species Streptomyces sp. [TaxId:465541] [259607] (2 PDB entries)
   6. Species Streptomyces sviceus [TaxId:463191] [258540] (2 PDB entries)
   7. Species Thermotoga maritima [TaxId:2336] [186925] (4 PDB entries)
   8. Species Actinomyces urogenitalis [TaxId:525246] [269148] (1 PDB entry)
   9. Species Streptomyces coelicolor [TaxId:100226] [277529] (1 PDB entry)

SCOP *Families*
Are made up of protein *Domains:*

SCOP *Domains* are parts of individual PDB structures

More info for Family c.1.2.1: Histidine biosynthesis enzymes

Timeline for Family c.1.2.1: Histidine biosynthesis enzymes:

- Family c.1.2.1: Histidine biosynthesis enzymes first appeared (with stable ids) in SCOP 1.55
- Family c.1.2.1: Histidine biosynthesis enzymes appears in SCOPe 2.04

SCOPe: Structural Classification of Proteins — extended. Release 2.05 (updated 2016-01-14, stable release February 2015)

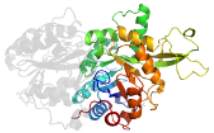
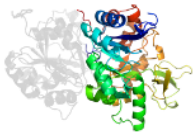
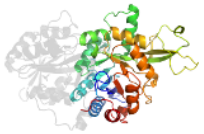
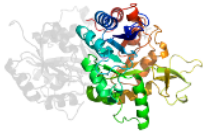
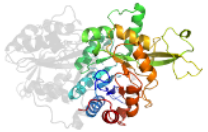
Browse | Stats & History | ASTRAL Subsets | Downloads | Related Resources | References | Help | About

Search **SCOPe** (example): Search

Lineage for Species: Baker's yeast (Saccharomyces cerevisiae), His7 [TaxId: 4932]

1. Root: SCOPe 2.05
2. Class c: Alpha and beta proteins (a/b) [51349] (148 folds)
3. Fold c.1: TIM beta/alpha-barrel [51350] (33 superfamilies)
   contains parallel beta-sheet barrel, closed; n=8, S=8; strand order 12345678
   the first seven superfamilies have similar phosphate-binding sites
4. Superfamily c.1.2: Ribulose-phoshate binding barrel [51366] (7 families) $\mathcal{S}$
5. Family c.1.2.1: Histidine biosynthesis enzymes [51367] (3 protein domains)
   structural evidence for the gene duplication within the barrel fold
   automatically mapped to Pfam PF00977
6. Protein Cyclase subunit (or domain) of imidazoleglycerolphosphate synthase HisF [51370] (4 species)
7. Species Baker's yeast (Saccharomyces cerevisiae), His7 [TaxId:4932] [69379] (4 PDB entries)

PDB entries in Species: Baker's yeast (Saccharomyces cerevisiae), His7 [TaxId: 4932]:

1. Domain(s) for 1jvn:
   1. 
      Domain d1jvna1: 1jvn A:230-552 [67355]
      Other proteins in same PDB: d1jvna2, d1jvnb2
      complexed with ni, pop, so4
   2. 
      Domain d1jvnb1: 1jvn B:230-552 [67357]
      Other proteins in same PDB: d1jvna2, d1jvnb2
      complexed with ni, pop, so4
2. Domain(s) for 1ox4:
   1. 
      Domain d1ox4a1: 1ox4 A:230-550 [87497]
      Other proteins in same PDB: d1ox4a2, d1ox4b2
      complexed with ni, pop, so4
   2. 
      Domain d1ox4b1: 1ox4 B:230-550 [87499]
      Other proteins in same PDB: d1ox4a2, d1ox4b2
      complexed with ni, pop, so4
3. Domain(s) for 1ox5:
   1. 
      Domain d1ox5a1: 1ox5 A:230-550 [87501]
      Other proteins in same PDB: d1ox5a2, d1ox5b2
      complexed with 1pr, ni
   2. 
      Domain d1ox5b1: 1ox5 B:230-550 [87503]
      Other proteins in same PDB: d1ox5a2, d1ox5b2
      complexed with 1pr, ni

SCOP *Domains*
Are derived from *Entries:*

SCOP *Entries* are individual PDB (Protein Data Bank) files

For PDB see http://pdbe.org

SCOP Summary (2009, but makes the point)

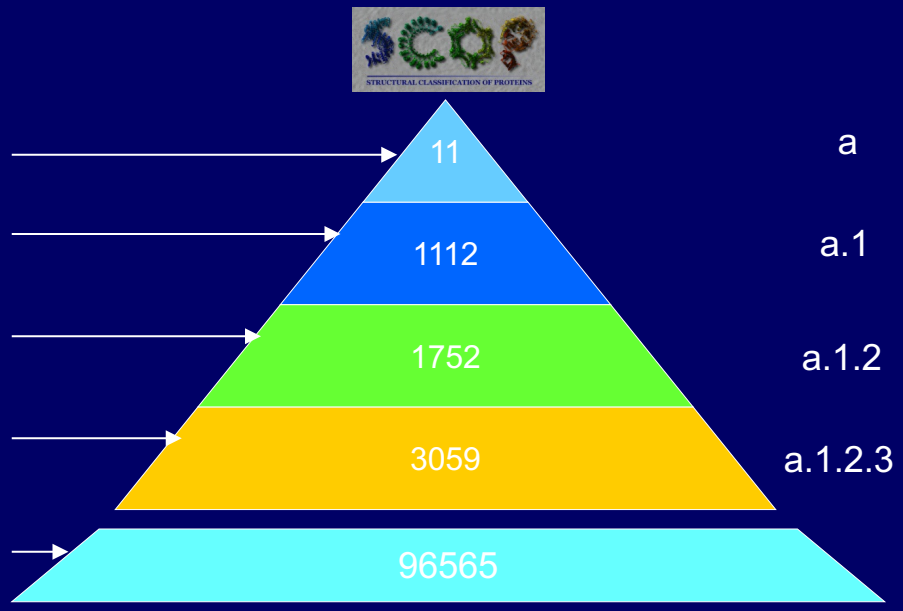
Class: Similar secondary structure composition

Fold: Same major secondary structures in same arrangement with same topology

Superfamily: low sequence identity but similar structure and function imply common evolutionary origin

Family: >30 % sequence identity or similar structure and function

Total number of domains:



CATH

- Similar hierarchy to scop with the addition of *Architecture* between Class and topology (fold).
- Much nicer looking interface than SCOP. Very interactive and modern.

CATH / Gene3D

26 million protein domains classified into 2,738 superfamilies

Browse » Search » Download » Take the Tour »

What is CATH?

CATH is a classification of protein structures downloaded from the Protein Data Bank. We group protein domains into superfamilies when there is sufficient evidence they have diverged from a common ancestor.

- Search CATH by text, ID or keyword
- Search CATH by protein sequence (FASTA)
- Search CATH by PDB structure
- Browse CATH Hierarchy
- CATH Release Statistics
- CATH Tutorials

Example pages

- PDB "2bop"
- Domain "1cukA01"
- Relatives of "1cukA01"
- Superfamily "HUPs"
- Functional Family
- FunFam Alignment
- Search for "enolase"
- Superfamily Comparison

Latest Release Statistics

CATH v4.0 based on PDB dated March 26, 2013	
235,858	CATH Domains
2,738	CATH Superfamilies
69,058	Annotated PDBs

Gene3D v12 released March 18, 2012	
6,131	Cellular Genomes
21,662,155	Protein Sequences
25,615,754	CATH Domain Predictions

Citing CATH

If you find this resource useful, please consider citing the reference that describes this work:

CATH: comprehensive structural and functional annotations for genome sequences.

Sillitoe I, Lewis, TE, Cuff AL, Das S, Ashford P, Dawson NL, Furnham N, Laskowski RA, Lee D, Lees J, Lehtinen S, Studer R, Thornton JM, Orengo CA

Nucleic Acids Res. 2015 Jan doi: 10.1093/nar/gku947

Browse CATH-Gene3D Hierarchy

BROWSE LINKS

Browse Hierarchy

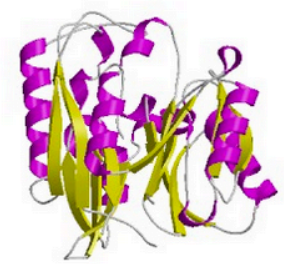
Highly Diverse Superfamilies

Superfamily Comparison

Select a CATH node...

A 5-stranded Propeller

CATH ID	
Topologies	1
Superfamilies	2
Domains	135
Example Domain	2ci1A00 [PDB]



 Tree  Sunburst

Top of CATH Hierarchy (4 Classes)

- C 1 Mainly Alpha — *5 Architectures, 397 Folds, 907 Superfamilies, 48121 Domains*
- C 2 Mainly Beta — *20 Architectures, 241 Folds, 547 Superfamilies, 58944 Domains*
- C 3 Alpha Beta — *14 Architectures, 626 Folds, 1158 Superfamilies, 125772 Domains*
  - A 3.10 Roll — *58 Folds, 101 Superfamilies, 9748 Domains*
  - A 3.15 Super Roll — *3 Folds, 3 Superfamilies, 5 Domains*
  - A 3.20 Alpha-Beta Barrel — *18 Folds, 46 Superfamilies, 10515 Domains*
  - A 3.30 2-Layer Sandwich — *223 Folds, 495 Superfamilies, 34581 Domains*
  - A 3.40 3-Layer(aba) Sandwich — *126 Folds, 287 Superfamilies, 49841 Domains*
  - A 3.50 3-Layer(bba) Sandwich — *11 Folds, 17 Superfamilies, 2313 Domains*
  - A 3.55 3-Layer(bab) Sandwich — *6 Folds, 6 Superfamilies, 24 Domains*
  - A 3.60 4-Layer Sandwich — *16 Folds, 18 Superfamilies, 3478 Domains*
  - A 3.65 Alpha-beta prism — *1 Folds, 2 Superfamilies, 405 Domains*
  - A 3.70 Box — *1 Folds, 1 Superfamilies, 173 Domains*
  - A 3.75 5-stranded Propeller — *1 Folds, 2 Superfamilies, 135 Domains*
  - A 3.80 Alpha-Beta Horseshoe — *3 Folds, 4 Superfamilies, 257 Domains*
  - A 3.90 Alpha-Beta Complex — *158 Folds, 175 Superfamilies, 14167 Domains*
  - A 3.100 Ribosomal Protein L15; Chain: K; domain 2 — *1 Folds, 1 Superfamilies, 130 Domains*
- C 4 Few Secondary Structures — *1 Architectures, 111 Folds, 126 Superfamilies, 3021 Domains*

Browse CATH-Gene3D Hierarchy

BROWSE LINKS

Browse Hierarchy

Highly Diverse Superfamilies

Superfamily Comparison

Select a CATH node...

T TIM Barrel

CATH ID	
Superfamilies	29
Domains	10050
Example Domain	2vxnA00 [PDB]

 Tree  Sunburst

Top of CATH Hierarchy (4 Classes)

- C 1 Mainly Alpha — *5 Architectures, 397 Folds, 907 Superfamilies, 48121 Domains*
- C 2 Mainly Beta — *20 Architectures, 241 Folds, 547 Superfamilies, 58944 Domains*
- C 3 Alpha Beta — *14 Architectures, 626 Folds, 1158 Superfamilies, 125772 Domains*
  - A 3.10 Roll — *58 Folds, 101 Superfamilies, 9748 Domains*
  - A 3.15 Super Roll — *3 Folds, 3 Superfamilies, 5 Domains*
  - A 3.20 Alpha-Beta Barrel — *18 Folds, 46 Superfamilies, 10515 Domains*
    - T 3.20.10 D-amino Acid Aminotransferase; Chain A, domain 2 — *1 Superfamilies, 131 Domains*
    - T 3.20.14 L-fucose Isomerase; Chain A, domain 3 — *1 Superfamilies, 15 Domains*
    - T 3.20.16 Serine Protease, Human Cytomegalovirus Protease; Chain A — *1 Superfamilies, 47 Domains*
    - T 3.20.19 Aconitase; domain 4 — *1 Superfamilies, 38 Domains*
    - T 3.20.20 TIM Barrel — *29 Superfamilies, 10050 Domains*
      - H 3.20.20.10 Alanine racemase — *181 Domains*
      - H 3.20.20.20 Dihydropteroate (DHP) synthetase — *134 Domains*
      - H 3.20.20.30 FMN dependent fluorescent proteins — *54 Domains*
      - H 3.20.20.40 Glycosyl hydrolases family 6, cellulases — *59 Domains*
      - H 3.20.20.60 Phosphoenolpyruvate-binding domains — *499 Domains*
      - H 3.20.20.70 Aldolase class I — *3485 Domains*
      - H 3.20.20.80 Glycosidases — *2048 Domains*
      - H 3.20.20.100 NADP-dependent oxidoreductase — *411 Domains*
      - H 3.20.20.105 tRNA-guanine (tRNA-G) transglycosylase — *89 Domains*
      - H 3.20.20.110 Rubisco — *328 Domains*
      - H 3.20.20.120 Enolase superfamily — *1097 Domains*
      - H 3.20.20.140 Metal-dependent hydrolases — *704 Domains*
      - H 3.20.20.150 Divalent-metal-dependent TIM barrel enzymes — *394 Domains*
      - H 3.20.20.190 Phosphatidylinositol (PI) phosphodiesterase — *113 Domains*
      - H 3.20.20.210 *Not yet named* — *81 Domains*
      - H 3.20.20.220 *Not yet named* — *46 Domains*
      - H 3.20.20.240 *Not yet named* — *41 Domains*
      - H 3.20.20.300 *Not yet named* — *67 Domains*
      - H 3.20.20.330 Homocysteine S-methyltransferase — *24 Domains*
      - H 3.20.20.350 *Not yet named* — *20 Domains*
      - H 3.20.20.360 Malate synthase, domain 3 — *9 Domains*
      - H 3.20.20.370 Glycoside hydrolase/deacetylase — *51 Domains*
      - H 3.20.20.380 CutC-like — *20 Domains*
      - H 3.20.20.390 FMN-linked oxidoreductases — *38 Domains*

Unique feature in CATH:

FunFams

Gene3D extends this to proteins of unknown structure where possible

CATH Home Search Browse Download About Support Search CATH by keywords or ID

CATH Superfamily 3.20.20.60

Phosphoenolpyruvate-binding domains

View in Gene3D

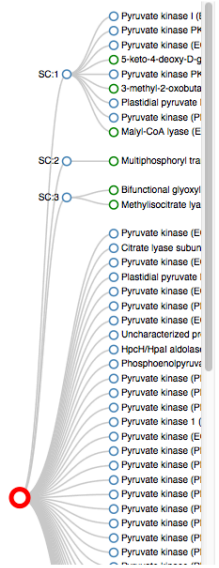
Home / Superfamily 3.20.20.60

SUPERFAMILY LINKS

- Summary
- Superfamily Superposition
- Classification / Domains
- Alignments
- Structural Neighbourhood
- Functional Annotations
- Taxonomy Browser
- Multi-Domain Organisation

Functional Families

Overview of the Structural Clusters (SC) and Functional Families (FF) within this CATH Superfamily



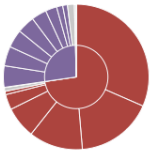
GO Diversity

Unique GO annotations

115 Unique GO terms

EC Diversity

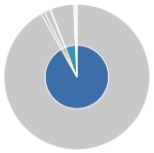
Unique EC annotations



45 Unique EC terms

Species Diversity

Unique species annotations



7056 Unique species

Structural Diversity

Structural domains within this superfamily

Domain Organisation

View multi-domain architectures via ArchSchema (Laskowski/EBI)

ArchSchema (requires Java)

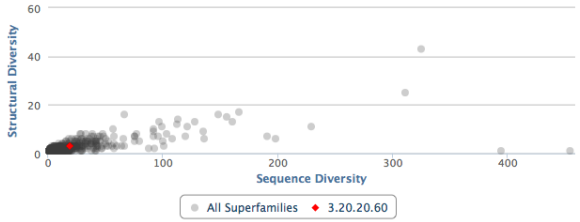
Enzyme Function

Evolution of Enzyme Function via FunTree (Furnham/EBI)

FunTree (opens new window)

Sequence/Structure Diversity

Overview of the sequence / structure diversity of this superfamily compared to other superfamilies in CATH. Click on the chart to view the data in more detail.



Superfamily Summary

A general summary of information for this superfamily.

Structures	
Domains:	499
Domain clusters (>95% seq id):	59
Domain clusters (>35% seq id):	19
Unique PDBs:	139
Alignments	
Structural Clusters:	3
FunFam Clusters:	53
Function	
Unique EC:	45
Unique GO:	115
Taxonomy	
Unique Species:	7056

CATH News — Support, Jobs

Get Started — Documentation, Tutorials

Download — WebServices, Software

About — Orengo Group, Web accessibility

Comparison of SCOP and CATH

- SCOP
- Largely *manually* created hierarchy with some automation to help
- A domain is only defined if it has been seen as a stand-alone protein and is though to be functionally independent.
- Early 1990s-style interface

- CATH
- Largely *automatically* created hierarchy – follows defined rules – some manual editing to help
- A domain is thought of as an independent folding unit and does not have to be seen as a stand-alone protein.
- Slicker interface with nice graphics and excellent links to other resources!

SCOP and CATH are both useful so consult both

For Sequence-based domain assignments

- Pfam
  - Collection of domain assignments made entirely from sequence data
- InterPro
  - Collection of different databases of domain assignments