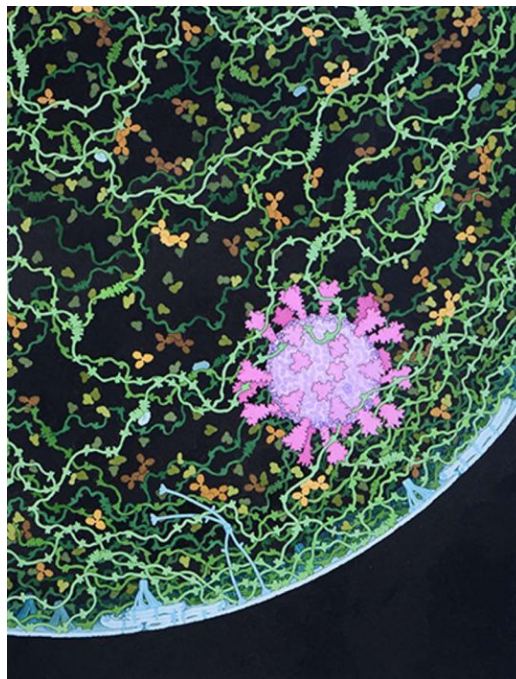


Dane strukturalne

Protein Data Bank (1971, 7 struktur)

- Pierwsze źródło danych biologicznych w historii
- Globalne archiwum struktur:
 - białek
 - DNA/RNA
- Otwarty dostęp do >210 000 struktur
- Kordynatorzy:
 - RCSB PDB (US)
 - PRBe (EMBL-EBI)
 - PDBj (Japonia)
 - PDBc (Chiny)
 -
 - EMDB (3DEM)
 - BMRB (NMR)

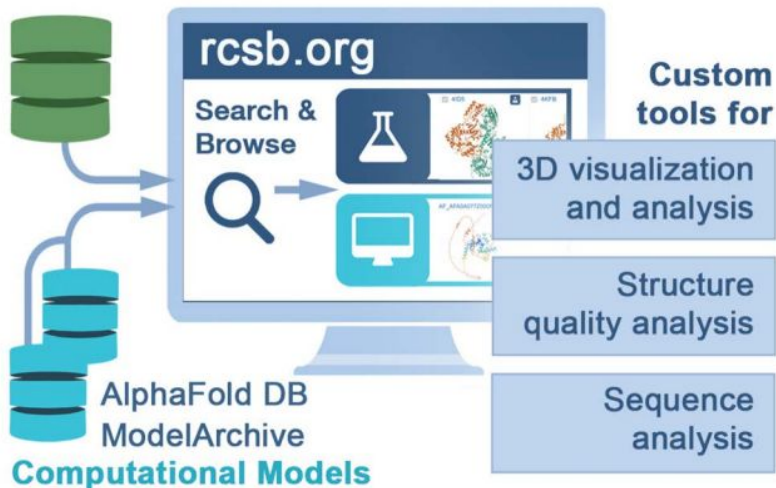


Dostępne dane

rcsb.org:

- >210 000 struktur PDB
- >1 000 000 modeli komputerowych (Computed Structure Models, CSM)
 - AlphaFold DB
 - ModelArchive
- Narzędzia do:
 - wizualizacji i analizy
 - oceny jakości struktur
 - analizy sekwencji

Experimental Models Protein Data Bank

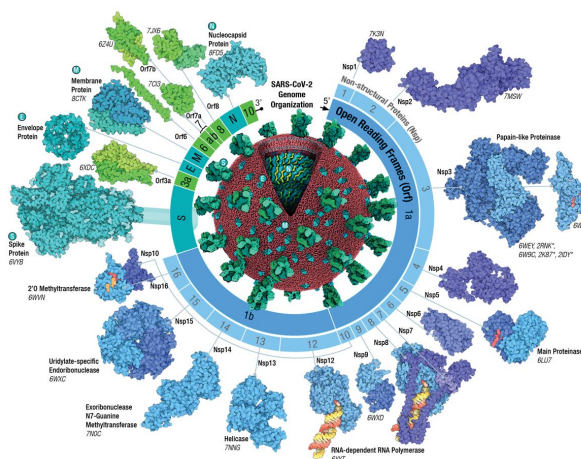


rscb.org



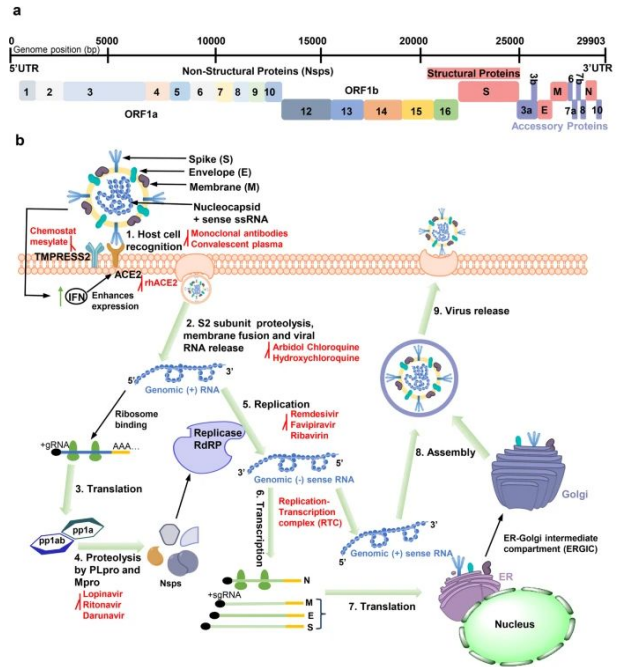
Znaczenie danych strukturalnych

- SARS-CoV 2002
>170 SARS-CoV structuresPD
- MERS-CoV 2012
>100 MERS-CoV structuresP
- COVID-19 2019
>3,000 SARS-CoV-2 structures
- Określenie pokrewieństwa
- Projektowanie szczepionek
- Identyfikacja leków przeciwwirusowych



Koronawirus SARS-CoV-2 genom

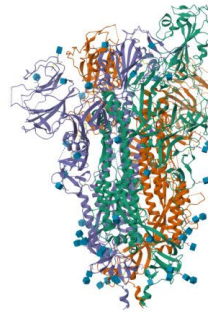
- Genom: RNA
- Białka strukturalne kodowane oddzielne
- Białka funkcjonalne kodowane jako poliptyd
- MP: main protease
- PL: papain like proteinase



Yan, W., Zheng, Y., Zeng, X. et al. Structural biology of SARS-CoV-2: open the door for novel therapies. *Sig Transduct Target Ther* 7, 26 (2022). <https://doi.org/10.1038/s41392-022-00884-5>

Strategie wykorzystujące dane strukturalne

- Szczepionka identyfikacja immunogenego elementu
- Zastosowanie leków przeciwwirusowych



- 6VSB
Prefusion
2019-nCoV spike
glycoprotein with
a single
receptor-binding
domain up

- 6LU7
The crystal structure of
COVID-19 main protease
in complex with an
inhibitor N3

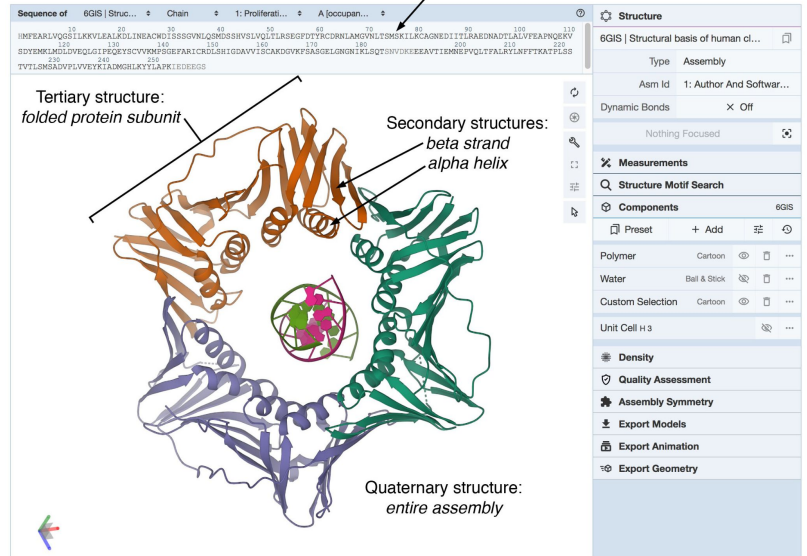


Organizacja molekularna

6GIS

Structural basis of human clamp sliding on DNA

Primary structure:
amino acid sequence

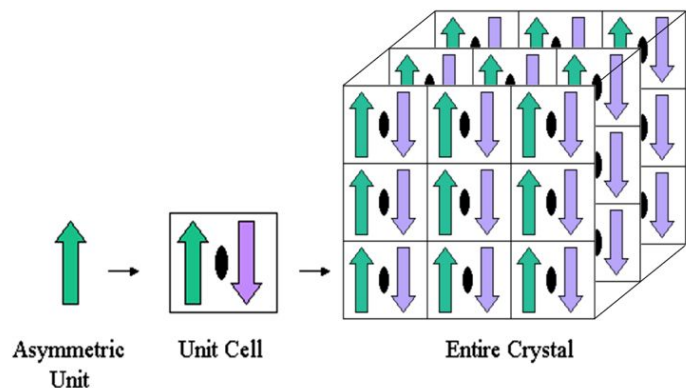


Organizacja molekularna

Jednostka asymetryczna jest najmniejszą częścią struktury krystalicznej, do której można zastosować operacje symetrii w celu wygenerowania kompletnej komórki elementarnej (jednostki powtarzalnej kryształu).

Operacje symetrii najczęściej spotykane w kryształach makrocząsteczek biologicznych to rotacje, translacje i osie śrubowe (kombinacje rotacji i translacji).

Zastosowanie operacji symetrii krystalograficznej do jednostki asymetrycznej daje jedną komórkę elementarną, która po przetłumaczeniu w trzech wymiarach tworzy cały kryształ.



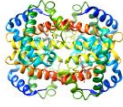
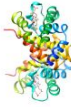
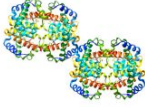
Znaczenie biologiczne

Krystaliczna jednostka asymetryczna może zawierać:

- jeden zespół biologiczny
- część zespołu biologicznego
- wiele zespołów biologicznych

Zawartość jednostki asymetrycznej zależy od pozycji krystalizowanej cząsteczki i jej konformacji w komórce elementarnej. W zależności od warunków krystalizacji i lokalnego upakowania mogą wystąpić dwa różne scenariusze:

- Kopie makrocząsteczki lub kompleksu w krystalicznej komórce elementarnej mają identyczne konformacje i zajmują pozycje związane z symetrią. W rezultacie zespół biologiczny może składać się z jednej kopii makrocząsteczki/kompleksu lub może składać się z dwóch lub więcej cząsteczek/kompleksów powiązanych symetrią, tworzących większy zespół.
- Kopie makrocząsteczki lub kompleksu przyjmują nieco inne konformacje i zajmują unikalne pozycje w krystalicznej jednostce asymetrycznej. W rezultacie każda z różnych pozycji makrocząsteczki/kompleksu może odpowiadać strukturalnie podobnym, ale nie identycznym zespołom biologicznym.

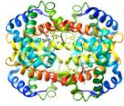
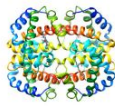
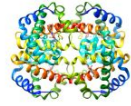
Asymmetric unit with one biological assembly	Asymmetric unit with a portion of a biological assembly	Asymmetric unit with multiple biological assemblies
		
Entry 2hhb contains one hemoglobin molecule (4 chains) in the asymmetric unit.	Entry 1out contains half a hemoglobin molecule (2 chains) in the asymmetric unit. A crystallographic two-fold axis generates the other 2 chains of the hemoglobin molecule.	Entry 1hv4 contains two hemoglobin molecules (8 chains) in the asymmetric unit.

Interpretacja komórki elementarnej

Zespół biologiczny (czasami określany również jako jednostka biologiczna) to zespół makromolekularny, który okazał się być lub uważa się, że jest funkcjonalną formą cząsteczki. Na przykład, funkcjonalna forma hemoglobiny ma cztery łańcuchy.

W zależności od konkretnej struktury krystalicznej, operacje symetrii składające się z rotacji, translacji lub ich kombinacji mogą wymagać wykonania w celu uzyskania kompletnego zespołu biologicznego. Alternatywnie, może być konieczne wybranie podzbioru zdeponowanych współrzędnych w celu przedstawienia zespołu biologicznego. W ten sposób zespół biologiczny może być zbudowany z:

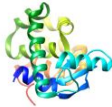
- jednej kopii jednostki asymetrycznej
- wielu kopii jednostki asymetrycznej
- części jednostki asymetrycznej

Biological assembly composed of one copy of the asymmetric unit	Biological assembly composed of multiple copies of the asymmetric unit	Multiple biological assemblies in the asymmetric unit
		
In entry 2hhb , the biological assembly is equivalent to the asymmetric unit.	In entry 1out the biological assembly includes two asymmetric units.	In entry 1hv4 the biological assembly is one-half of the asymmetric unit.
No operations are necessary.	Application of a crystallographic symmetry operation (a 180 rotation around a crystallographic two-fold axis) produces the complete biological assembly.	The entry contains two structurally similar, but not entirely identical copies of the biological assembly within the crystal asymmetric unit.

(Nad)interpretacja komórki elementarnej

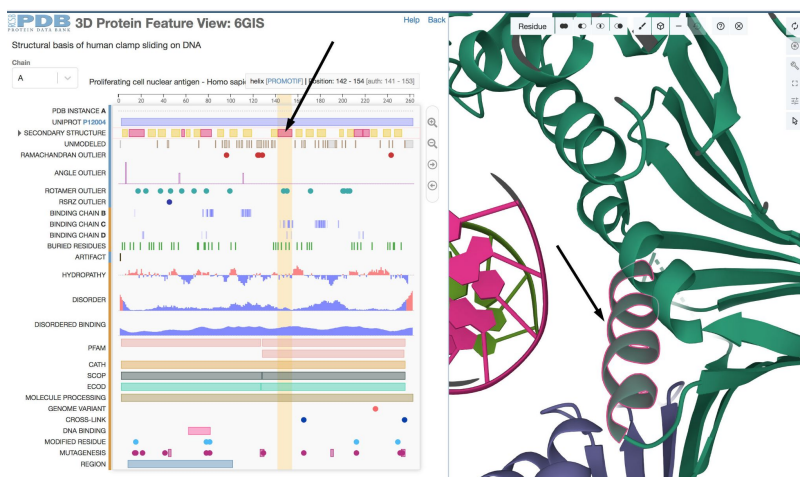
Cząsteczka może czasami wydawać się multimeryczna w kryształach na podstawie upakowania kryształów. Jednak może nie być żadnych dowodów lub biologicznego znaczenia na poparcie stanu multimerycznego w roztworze.

Gdy wpis jest przetwarzany, wszystkie prawdopodobne złożenia są obliczane na podstawie zakopanej powierzchni i energii interakcji. Te przewidywane złożenia mogą, ale nie muszą pokrywać się z tym, co autor uważa za biologicznie istotne dla cząsteczki. Zespoły biologiczne zgłoszone we wpisie zawierają uwagę wyjaśniającą, czy jest to "dostarczone przez autora", "określone przez oprogramowanie", czy oba.

Asymmetric unit (monomer)	Author & Software Determined Biological Assembly (monomer)	Software Determined Biological Assembly (dimer)
		
The asymmetric unit is a monomer. These are the deposited coordinates.	The "author provided" and "software determined" biological assemblies are both monomers.	The software, PISA, predicts that this molecule may also form a dimer. Hence the second biological assembly is only "software determined".

Identyfikacja elementów kluczowych

- Struktury białek zorganizowane są w motywy
- CATH
Class(C), Architecture(A),
Topology(T), and Homologous
Superfamily(H)
- SCOP
Structural Classification of Proteins



CATH

- Wyszukiwanie
 - Text or ID
 - FASTA
 - Structure
- Wynik
 - "domeny"
 - "elementy funkcjonalne"
 - "ścieżka ewolucyjna"

Home Search Browse Download About Support

Search CATH by keywords or ID

CATH / Gene3D v4.3

151 million protein domains classified into 5,841 superfamilies

Search by keywords, PDB code, GO term, etc.

CATH annotations for the 21 model organisms predicted by AlphaFold (v2) are available to download (doi:10.1038/s42003-023-04488-9). Core classification files for the latest version of CATH-Plus (v4.3) are available to download. Daily updates of our very latest classifications are also available.

3D Structure

Find out what 3D structure your protein adopts

Find out more

Protein Evolution

Learn about a particular protein family and how it evolved

Find out more

Protein Function

Investigate the function of your protein

Find out more

Conserved Sites

Look at protein sites that are highly conserved and implicated in function

Find out more

Download Data

Download data files and query CATH via web services

Go

Learn more

Find out how CATH is created and maintained, how to link to CATH and more

Go

What is CATH-Gene3D?

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
- Search CATH by PDB structure
- Browse CATH Hierarchy
- CATH Release Statistics
- CATH Tutorials

Gene3D uses the information in CATH to predict the locations of structural domains on millions of protein sequences available in public databases. This allows us to include additional annotations to the CATH-Gene3D database such as functional information and active site residues.

- Go to Gene3D
- Compare Genomes
- Download Gene3D Data
- Learn how Gene3D is created

If you have any questions, comments or suggestions please get in touch via [Twitter](#), [ask a question in our online forum](#) or visit our [support page](#).

Latest Release Statistics

	CATH-Plus 4.3.0	CATH (daily snapshot)
PDB Release	01-07-2019	5 months ago
Domains	590235	538613
Superfamilies	5401	6031
Annotated PDBs	15005	106379

	Gene3D v21
Protein Sequences	32,885,324
CATH Domain Predictions	151,513,797

6VSI

>6VSB
glycoprotein
syndrom

CATH Domain

Home / Superfamily 3.30.70

DOMAIN LINKS

Summary

Structure

Sequence

Neighbourhood

CATH Superfamily 3.30.70.1840

Spike protein, C-terminal core receptor binding subdomain

Home / Superfamily 3.30.70.1840

SUPERFAMILY LINKS

Summary

Superfamily Superposition

Classification / Domains

Functional Families

Structural Neighbourhood

Functional Families

Overview of the Structural Clusters (SC) and Functional Families within this CATH Superfamily. Clusters with a representative structure are represented by a filled circle.

☐ Putative spike glycoprotein
☐ Putative spike glycoprotein
☐ Spike glycoprotein

Superfamily: Spike protein, C-terminal core receptor binding subdomain

The newly-emerging Middle East respiratory syndrome coronavirus (MERS-CoV) can cause severe and fatal acute respiratory disease in humans. Despite global efforts, the potential for an associated pandemic in the future cannot be excluded. Since September 2012, 1,368 laboratory-confirmed cases of human infection by the Middle East respiratory syndrome coronavirus (MERS-CoV) have officially been reported to World Health Organization as of 17 July 2015, including at least 490 related deaths with a fatality rate of ~35.8%.

MERS-CoV belongs to the beta-genus and uses human dipeptidyl peptidase 4 (DPP4) as its host receptor. Receptor recognition is a major determinant of coronavirus host range and tropism. An envelope-anchored trimeric spike protein is responsible for coronavirus entry into host cells via binding to the host receptor and subsequently fusing viral and host membranes. The spike protein consists of a receptor-binding S1 subunit and a membrane fusion S2 subunit. The S1 subunit consists of an N-terminal and a C-terminal domain, both of which can function as receptor-binding domains (RBD). However, it is the C-domain that interacts with DPP4.

This entry represents the core-subdomain of the RBD which is a five-stranded antiparallel beta-sheet with several short alpha helices.

PMD:23803533 PMD:20391699

GO Diversity

Unique GO annotations

Unique GO terms

Unique GO terms

Unique GO terms

Unique GO terms

Unique GO terms

Unique GO terms

Unique GO terms

Unique GO terms

Unique GO terms

Unique GO terms

Unique GO terms

Unique GO terms

Unique GO terms

Unique GO terms

Unique GO terms

Unique GO terms

Unique GO terms

Unique GO terms

Unique GO terms

Unique GO terms

Unique GO terms

Unique GO terms

Unique GO terms

Unique GO terms

Unique GO terms

EC Diversity

Unique EC annotations

Unique EC terms

Unique EC terms

Unique EC terms

Unique EC terms

Unique EC terms

Unique EC terms

Unique EC terms

Unique EC terms

Unique EC terms

Unique EC terms

Unique EC terms

Unique EC terms

Unique EC terms

Unique EC terms

Unique EC terms

Unique EC terms

Unique EC terms

Unique EC terms

Unique EC terms

Unique EC terms

Unique EC terms

Unique EC terms

Unique EC terms

Unique EC terms

Unique EC terms

Species Diversity

Unique species annotations

Unique species

Unique species

Unique species

Unique species

Unique species

Unique species

Unique species

Unique species

Unique species

Unique species

Unique species

Unique species

Unique species

Unique species

Unique species

Unique species

Unique species

Unique species

Unique species

Unique species

Unique species

Unique species

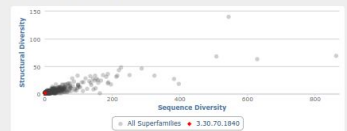
Unique species

Unique species

Unique species

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: 58

Domain clusters (~95% seq id): 6

Domain clusters (~35% seq id): 2

Unique PDBs: 28

Alignments

Structural Clusters (SA): 1

Structural Clusters (SA): 1

FunFam Clusters: 3

Function

Unique EC: 10

Unique GO: 10

Taxonomy

Unique Species: 121

Structure

JD8

[PDBSum](#)
[Proteopedia](#)

RAY DIFFRACTION

schierichia

Structure of Severe Acute Respiratory Syndrome Coronavirus Receptor-Binding Domain complexed with neutralizing Antibody

abakaran, P., Gan, J.,

ng, Y., Zhu, Z., Choudhry,

, Xiao, X., Ji, X., Dimitrov,

S.

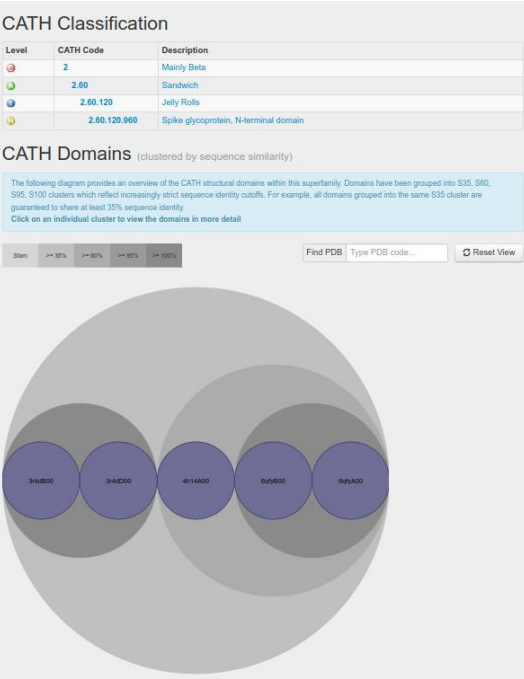
3iol.Chem.

API Help

Superfamily



Classification
Domains



Functional families

Browse Functional Families

Displaying 2 FunFams

Filter FunFams by Keywords

ID	Function Family (FunFam) Name	Total Sequences	Enzyme?	Structure?	Structural Representative	PDB Sites?	Alignment Diversity (0-100)
1	Spike glycoprotein	32			-	-	5.8
2	Spike glycoprotein	19		3D	3r4dD00	-	65.4

Search SCOP

[Keyword and ID search](#) [Sequence search](#)

Search results for 6vsb

[Domains \[12 \]](#) [Represented \[24 \]](#)

- [8092466](#) [6VSB A:335-526](#)
- [8092467](#) [6VSB A:335-526](#)
- [8092468](#) [6VSB A:27-293](#)
- [8092469](#) [6VSB A:27-293](#)
- [8092470](#) [6VSB A:717-1068](#)
- [8092471](#) [6VSB A:717-1068](#)
- [8092472](#) [6VSB A:1074-1146](#)
- [8092473](#) [6VSB A:1074-1146](#)
- [8092660](#) [6VSB A:529-592](#)
- [8092661](#) [6VSB A:529-592](#)
- [8092662](#) [6VSB A:593-699](#)
- [8092663](#) [6VSB A:593-699](#)

Domena

