

# Introduction to Structural Life Science Data Cloud

Akira R. KINJO

Institute for Protein Research, Osaka University

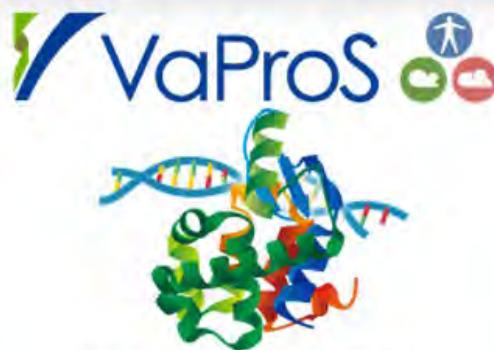
2015-09-13 (Sun)



# 創薬等支援技術基盤プラットフォーム >>> 情報拠点



情報拠点は、これまでの構造生物学関連ナショナルプロジェクトから継承したデータベース・解析ツールの拡充に加えて、構造生命科学データクラウドを新たに開発・提供することによって、創薬等支援技術基盤プラットフォーム (PDIS) の一翼を担って行きます。

[HOME](#)[プロジェクトの概要](#)[情報拠点の人々](#)[役立つDBと解析ツール](#)

構造生命科学データクラウド  
Variation effect on PROtein Structure and function  
Ver 2.1.0

## 役に立つデータベースと解析ツール



情報拠点 Web サイトを含む Web サイトで  
提供されている有用な情報資源

Ver 2.0

## ニュースウォッチ



構造生命科学関連ニュースと  
情報拠点からの新着情報  
最終更新日 2015.08.31

## 情報拠点を知る



私たちが考えていること

### 代表機関

五條堀 孝 教授

国立遺伝学研究所  
遺伝情報分析研究室

### 分担機関

金城 玲 准教授

大阪大学蛋白質研究所  
蛋白質情報科学研究室

### 分担機関

木下 賢吾 教授

東北大学大学院  
情報科学研究科  
生命情報システム科学分野

### 分担機関

近藤 るみ 准教授

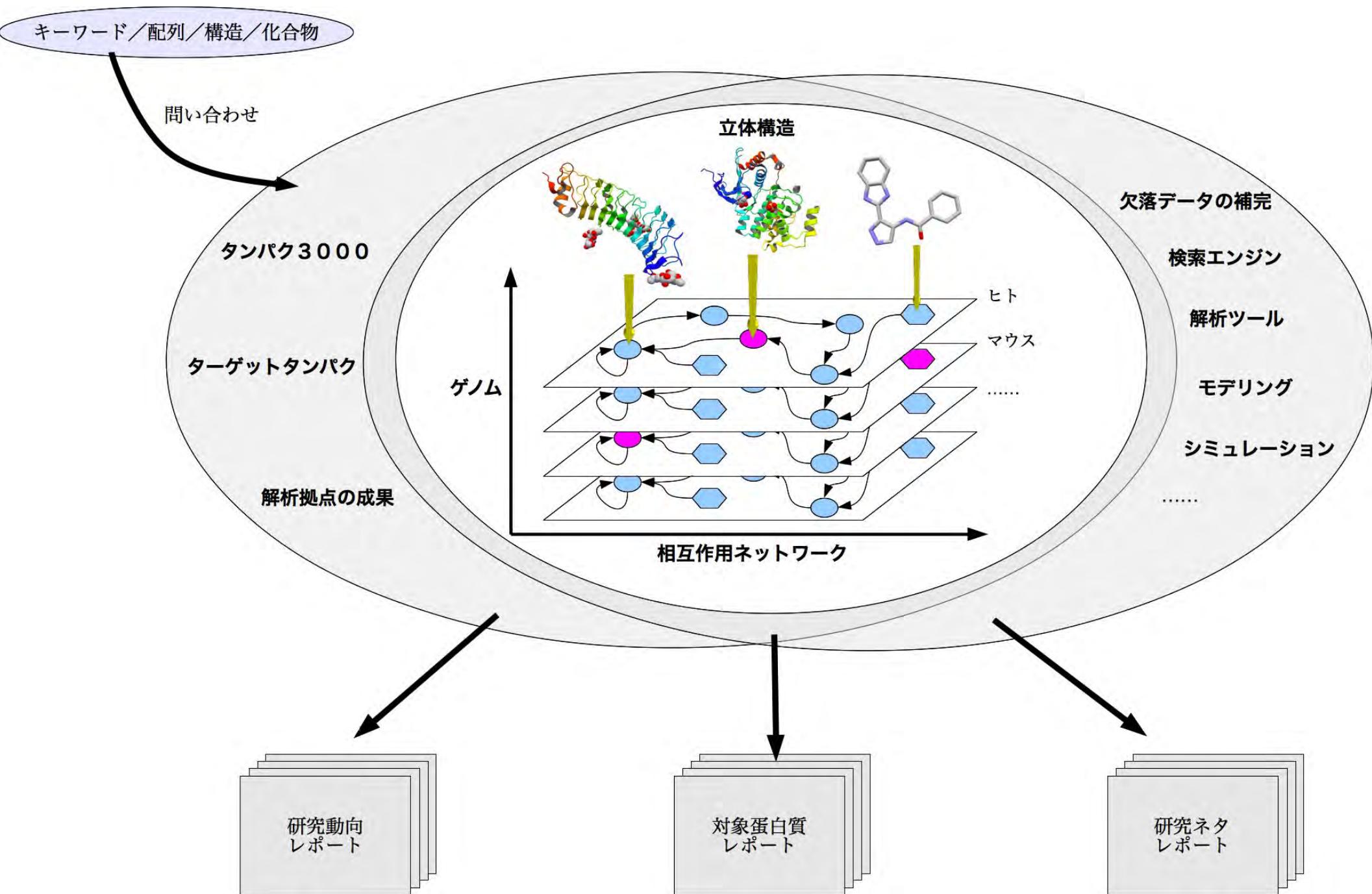
お茶の水女子大学  
人間文化創成科学研究科

### 分担機関

永田 宏次 准教授

東京大学大学院  
農学生命科学研究所  
応用生命化学専攻

本成果は、国立研究開発法人日本医療研究開発機構 創薬等ライフサイエンス研究支援基盤事業  
(創薬等支援技術基盤プラットフォーム事業)の支援により得られました。



VaProS | Home

p4d-info.nig.ac.jp/vapro/

Akira

**EXECUTION HISTORY**

Variation effect on PROtein Structure and function

**VaProS**

HOME ABOUT TUTORIAL

Platform for Drug Discovery, Informatics, and Structural Life Science

**Keyword Sequence**

Keywords  Search

Query: gene/protein, compound or disease name, UniProtKB ID/AC, EntrezGene ID, EMBL, Ensembl, Ensembl\_TRS, Ensembl\_PRO, OMIM ID, ...  
(e.g. beta galactosidase, galactosemia, BGAL\_HUMAN, ENSP00000306920, M27507, AC112211)

## VaProS: VAriation effect on PROtein Structure and function

VaProS, VAriation effect on PROtein Structure and function, is a new data cloud for Structural Life Science and is the core technology to lead the collaboration between the discipline in Structural Biology and the whole Life Sciences. Led by the initiative of National Institute of Genetics, VaProS has been developed around the Integrated Structural Biology Database at Institute for Protein Research in Osaka University, together with the selected outcomes from Protein 3000 Project, Targeted Proteins Research Program, Genome Network Project and Cell Innovation Project. [more...](#)

### hGtoP



A database of human proteins identified in the genome. This database focuses on a structural annotation of each protein.

### 3D Interaction



A database of human proteins with molecular interaction information. Interactions between a human protein and a protein, nucleic acid, or ligand can be obtained which is derived from a homology search against PDB.

### Natural Ligand Database



A database of protein-ligand interactions in three-dimension. The ligand in the database is a "natural ligand" generated by the knowledge of enzymatic reactions appeared in KEGG combined with a three-dimensional data of protein-analog interactions in PDB.

### COXPRESdb



A database of gene coexpression. DNA microarray or RNA-seq analyses produce information on relative expression levels for thousands of genes simultaneously. In addition, large collections of expression data contain information about concerted changes in transcript levels in these datasets beyond the original purpose of each dataset.

### S-VAR



A tool to assess functional impact of missense and multiple mutations on a protein. The assessment enables one to prioritize the mutations for further studies.

### Autophagy



A database of up-to-date proteins related to autophagy (self-digestion process in eukaryotic cells). The information here includes from protein sequences to protein-protein interactions and is human curated.

# HOMCOS :複合体立体構造の検索・モデリングサーバ

- ・PDBに収納された複合体の立体構造データを活用するためのサーバ
- ・配列相同意検索はBLAST、化学構造類似性検索はKCOMBUを使用

## タンパク質に対する結合分子検索

<http://homcos.pdbj.org>

アミノ酸配列を入力として、PDBに対してBLAST検索、ホモログが結合する分子のリストを作成



### 予測結合分子のコンタクトバー表示

予測結合分子のコンタクトバー表示

結合が予測される他のタンパク質

結合が予測される化合物

1 M E  
2 E H b  
3 F H b

### サイトごとの情報をまとめたテーブル

1行1サイトでの表示	
SITE_281_F	0.0 zwhs_C
SITE_280_A	1. zwhs_C
SITE_281_D	1. zwhs_C
SITE_282_N	0.0 zwhs_C metal
SITE_283_O	0.0 zwhs_C
SITE_284_P	0.0 zwhs_C ARGENTINYLGLY
SITE_285_E	0.0 zwhs_C compound
SITE_286_T	0.0 zwhs_C ATLOVIMIDENONE
SITE_287_M	11.6 Usfh_A compound
SITE_288_R	25.8 Usfh_A metal compound

288番目のArgに対するまとめ

288 Arg on 288-th site (TEMPLATE: PDB\_1dz assembly\_M1)

288番目のArgが結合部位となる化合物

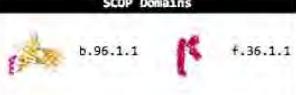
hGtoP :hsap0 :ACHA4\_HUMAN  
 p4d-info.nig.ac.jp/hGTOP\_db/mas?id=P43681-2

**hTOP**

**Homo sapiens(Human)**  
**Recname:Neuronal acetylcholine receptor subunit alpha-4**  
**AccessionID:[P43681-2](#)**  
**Gene:ACHA4\_HUMAN**

Organism	HUM	PRI	ROD	MAM	VRT	INV	PLN	BCT	ARC
Hits	52	99	57	73	318	206	4	0	0

**SCOP Domains**



**Database** | **Range** | **Top-Hit** | **Cross-Reference** | **Evalue** | **SeqID**

Database	Range	Top-Hit	Cross-Reference	Evalue	SeqID
555 amino acids					
SECSTR					
PSIPRED					
DISOPRED					
BLT:PDB	56->276	2bg9A	PDB1	2e-61	47.1%
BLT:PDB	490->547	2bg9A	PDB1	5e-15	50.0%
RPS:PDB	43->171	4aoeA	PDB1	2e-36	22.0%
RPS:PDB	169->203	2k58B	PDB1	2e-12	80.0%
RPS:PDB	238->262	3mrqA	PDB1	3e-10	72.0%
RPS:PDB	498->526	2bg9A	PDB1	3e-15	55.2%
RPS:SCOP	48->170	119bA	b.96.1.1	3e-33	23.8%
RPS:SCOP	172->278	1oedA	f.36.1.1	3e-40	63.8%
RPS:SCOP	519->547	1oedA	f.36.1.1	2e-06	44.8%
HMM:SCOP	172->551	1oedC	Neurotransmitter-gated ion-channel transmembrane pore	1.3e-84	0.0%
HMM:SCOP	57->171	119bA	Nicotinic receptor ligand binding domain-like	1.8e-39	0.0%
RPS:PFM	58->172	PF02931	PF02931	3e-22	45.4%
RPS:PFM	179->336	PF02932	PF02932	1e-15	50.6%
RPS:PFM	442->547	PF02932	PF02932	1e-06	41.5%
HMM:PFM	179->547	PF02932	Neur chan memb	2.6e-106	63.1%
HMM:PFM	58->171	PF02931	Neur chan LBD	3e-36	42.1%
BLT:SWISS	56->555	ACHA4_PANTR	-	0.0	94.6%
PROS	183->186	PS00001	ASN GLYCOSYLATION		
PROS	288->291	PS00004	CAMP PHOSPHO SITE		
PROS	44->46	PS00005	PKC PHOSPHO SITE		
PROS	185->187	PS00005	PKC PHOSPHO SITE		
PROS	149->151	PS00005	PKC PHOSPHO SITE		
PROS	263->265	PS00005	PKC PHOSPHO SITE		
PROS	350->352	PS00005	PKC PHOSPHO SITE		
PROS	461->463	PS00005	PKC PHOSPHO SITE		
PROS	474->476	PS00005	PKC PHOSPHO SITE		
PROS	479->481	PS00005	PKC PHOSPHO SITE		
PROS	515->517	PS00005	PKC PHOSPHO SITE		
PROS	24->27	PS00006	CK2 PHOSPHO SITE		
PROS	125->128	PS00006	CK2 PHOSPHO SITE		
PROS	353->356	PS00006	CK2 PHOSPHO SITE		
PROS	492->485	PS00006	CK2 PHOSPHO SITE		
PROS	515->518	PS00006	CK2 PHOSPHO SITE		
PROS	495->502	PS00007	TYR PHOSPHO SITE		
PROS	15->28	PS00008	MYRISTYL		
PROS	32->37	PS00008	MYRISTYL		
PROS	145->150	PS00008	MYRISTYL		
PROS	388->393	PS00008	MYRISTYL		
PROS	409->414	PS00008	MYRISTYL		
PROS	433->438	PS00008	MYRISTYL		
PROS	98->104	PS00236	NEUROTR ION CHANNEL		
TM					
SEG					
555 amino acids					

# Summary of homologs, structures, motifs, etc.



SEARCH

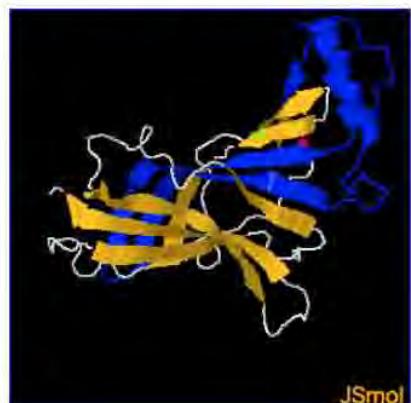
## Result of RPS-Blast-PDB "4aoeA"

*Homo sapiens(Human)*

Rename:Neuronal acetylcholine receptor subunit alpha-4

AccessionID:[P43681-2](#)

Gene:ACHA4\_HUMAN



PDB INFORMATION	
ENTRY	4aoeA
LENGTH	205
MOLNAME	ACETYLCHOLINE-BINDING PROTEIN TYPE 2
SCOP_TAXONOMY	x.x.x

LIGAND INFORMATION					
No	Compound_ID	ATOM	Chain	ATOM Number	Residue Number
1	PRO	ATOM	A	1659	205
2	PRO	ATOM	B	1659	205
3	PRO	ATOM	C	1659	205
4	PRO	ATOM	D	1659	205
5	PRO	ATOM	E	1659	205

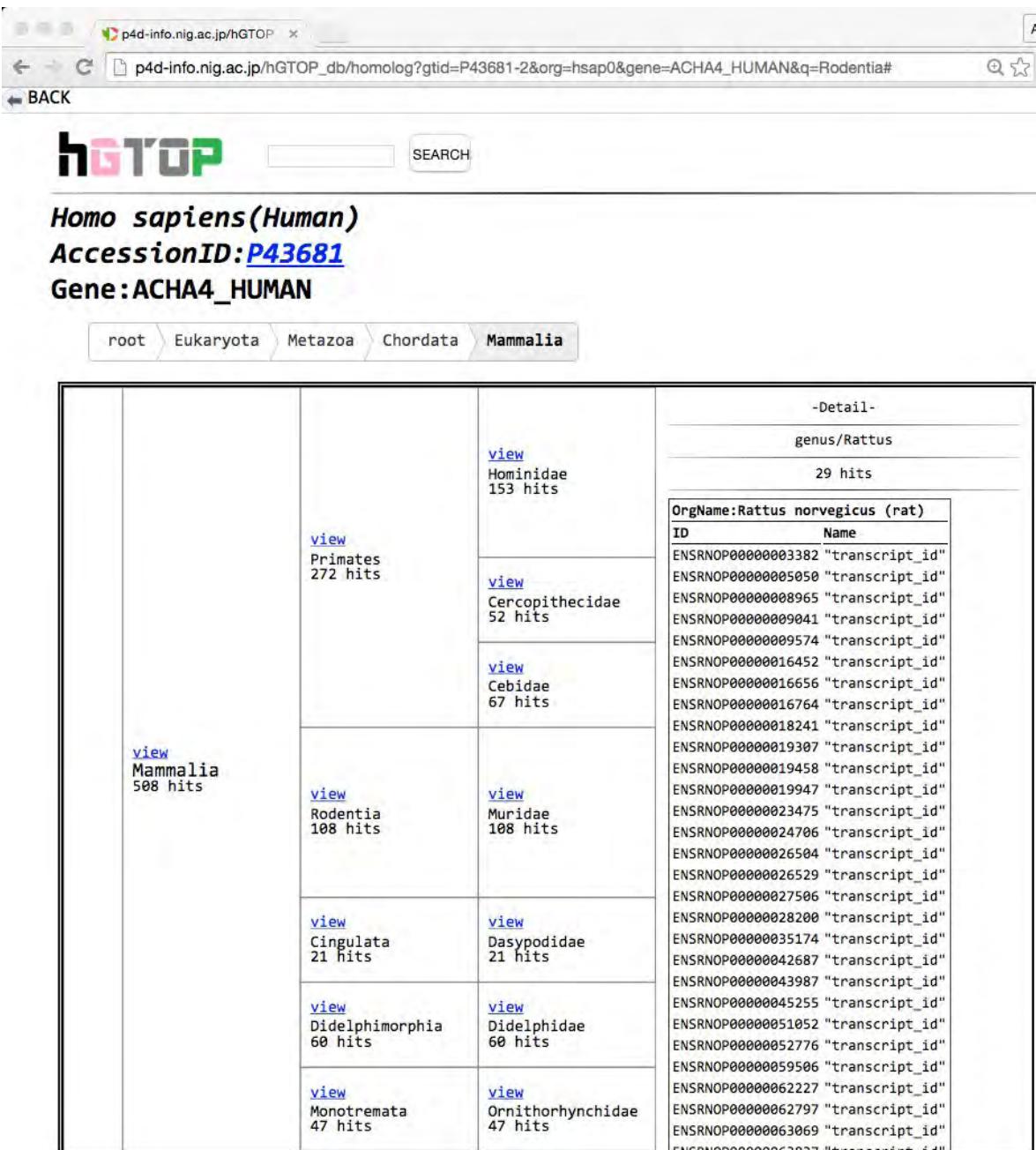
ALIGNMENT				
SeqID	51.9%	Eval	2e-36	Ncomp 128 Naa_query 555 Naa_target 205

AASEC	43:LSGRIWTAGCPGEGADGDFAVTHLTKAHLFHDGRVQWTTPAIYKSSCSIDVTFFFDQQNCTMKFGSWTYI
TARGET	79:LPSKYIWLPHIEVYNSIGKPGIHSQQLVRVYKDGTVTFVPQYTIRFCALEN-VTTEQGAACTLKFGPWTYI
SECSTR	:EEGGGcccccccEETccccccEEccccEEEEETcEEEEccEEEEEEccccTT cccccccEEEEEEcccc

[View On Hits](#)

**Recname:Neuronal acetylcholine receptor alpha 4**  
**AccessionID:[P43681-2](#)**  
**Gene:ACHA4\_HUMAN**

Organism	HUM	PRI	ROD	MAM	VRT	INV	PLN	BCT	ARC
Hits	52	99	57	73	318	206	4	0	0

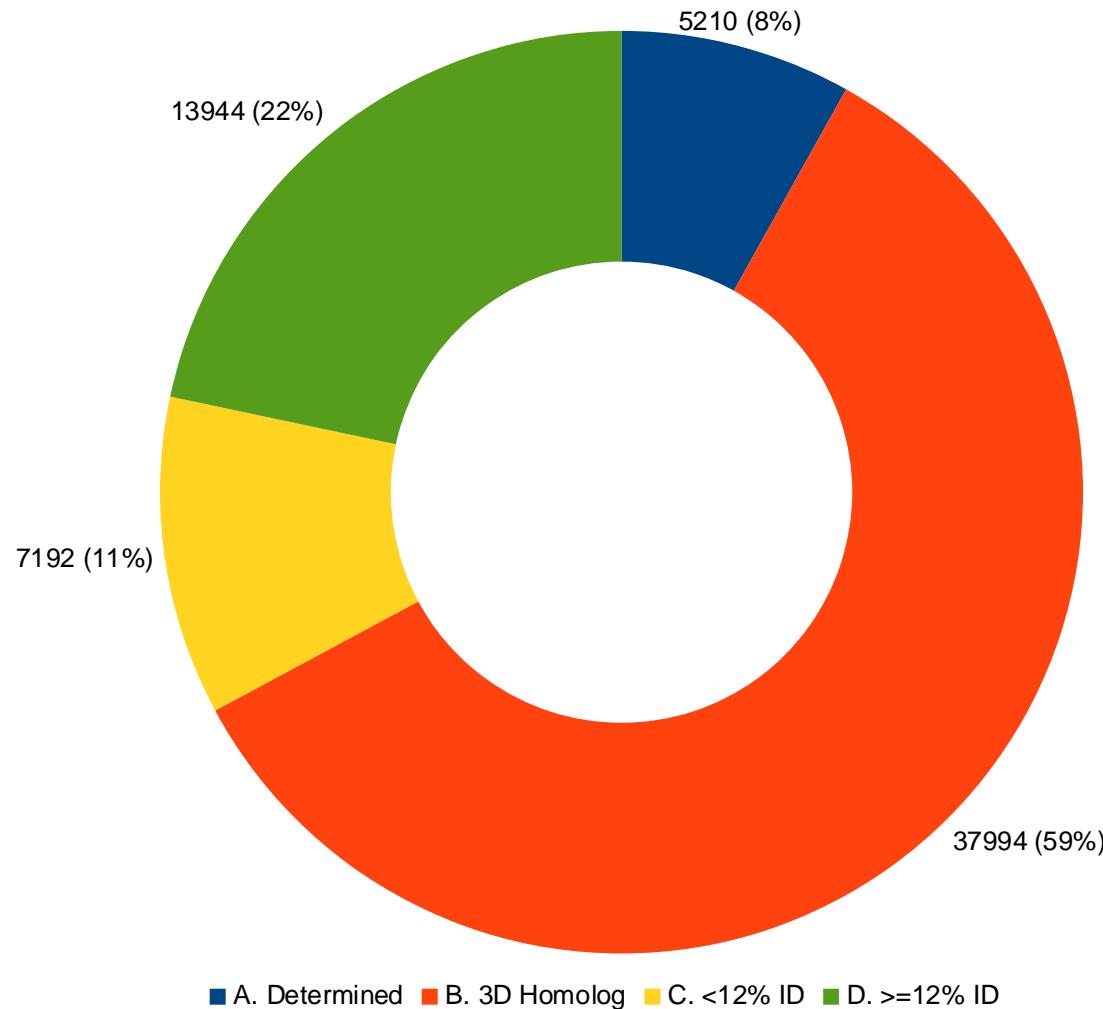

 A screenshot of a web browser displaying the hGTOP database. The URL in the address bar is `p4d-info.nig.ac.jp/hGTOP_db/homolog?gtid=P43681-2&org=hsap0&gene=ACHA4_HUMAN&q=Rodentia#`. The page title is "hGTOP". The main content shows search results for the gene *ACHA4\_HUMAN* in *Homo sapiens (Human)*. The accession ID is [P43681](#). The gene name is **ACHA4\_HUMAN**. The navigation path is root > Eukaryota > Metazoa > Chordata > Mammalia. On the right, there is a detailed view for the genus *Rattus*, which has 29 hits. Below this, a table lists transcript IDs for various rat species. The table has columns for ID and Name, with entries starting with ENSRNOP0000003382 "transcript\_id".

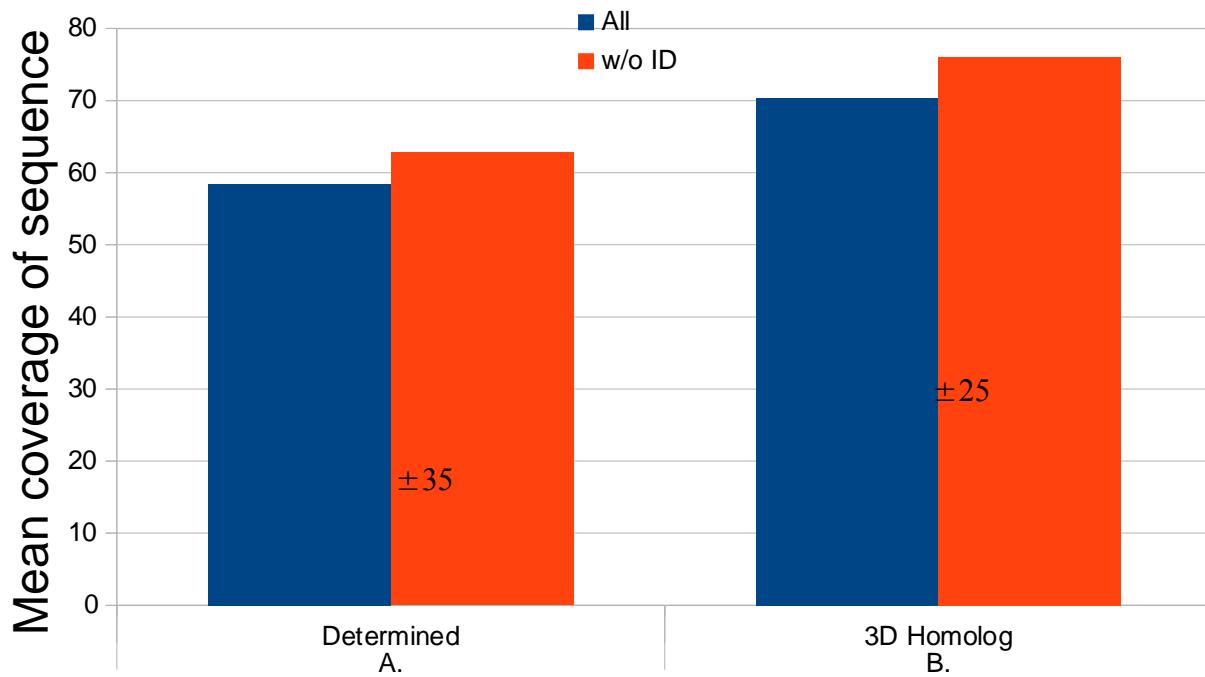
-Detail-	
genus/Rattus	
29 hits	
OrgName:	Rattus norvegicus (rat)
ID	Name
ENSRNOP0000003382	"transcript_id"
ENSRNOP0000005050	"transcript_id"
ENSRNOP0000008965	"transcript_id"
ENSRNOP0000009041	"transcript_id"
ENSRNOP0000009574	"transcript_id"
ENSRNOP00000016452	"transcript_id"
ENSRNOP00000016656	"transcript_id"
ENSRNOP00000016764	"transcript_id"
ENSRNOP00000018241	"transcript_id"
ENSRNOP00000019307	"transcript_id"
ENSRNOP00000019458	"transcript_id"
ENSRNOP00000019947	"transcript_id"
ENSRNOP00000023475	"transcript_id"
ENSRNOP00000024706	"transcript_id"
ENSRNOP00000026504	"transcript_id"
ENSRNOP00000026529	"transcript_id"
ENSRNOP00000027506	"transcript_id"
ENSRNOP00000028200	"transcript_id"
ENSRNOP00000035174	"transcript_id"
ENSRNOP00000042687	"transcript_id"
ENSRNOP00000043987	"transcript_id"
ENSRNOP00000045255	"transcript_id"
ENSRNOP00000051052	"transcript_id"
ENSRNOP00000052776	"transcript_id"
ENSRNOP00000059506	"transcript_id"
ENSRNOP00000062227	"transcript_id"
ENSRNOP00000062797	"transcript_id"
ENSRNOP00000063069	"transcript_id"
ENSRNOP00000063077	"transcript_id"



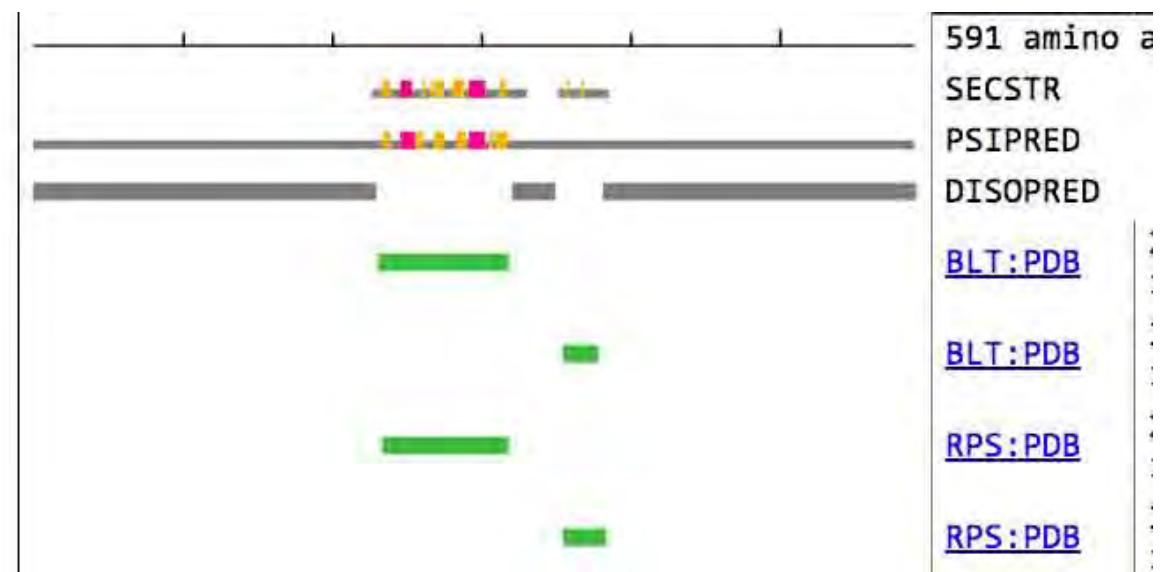
# Human Structurome Today (64340 proteins)

~70% have some 3D information!

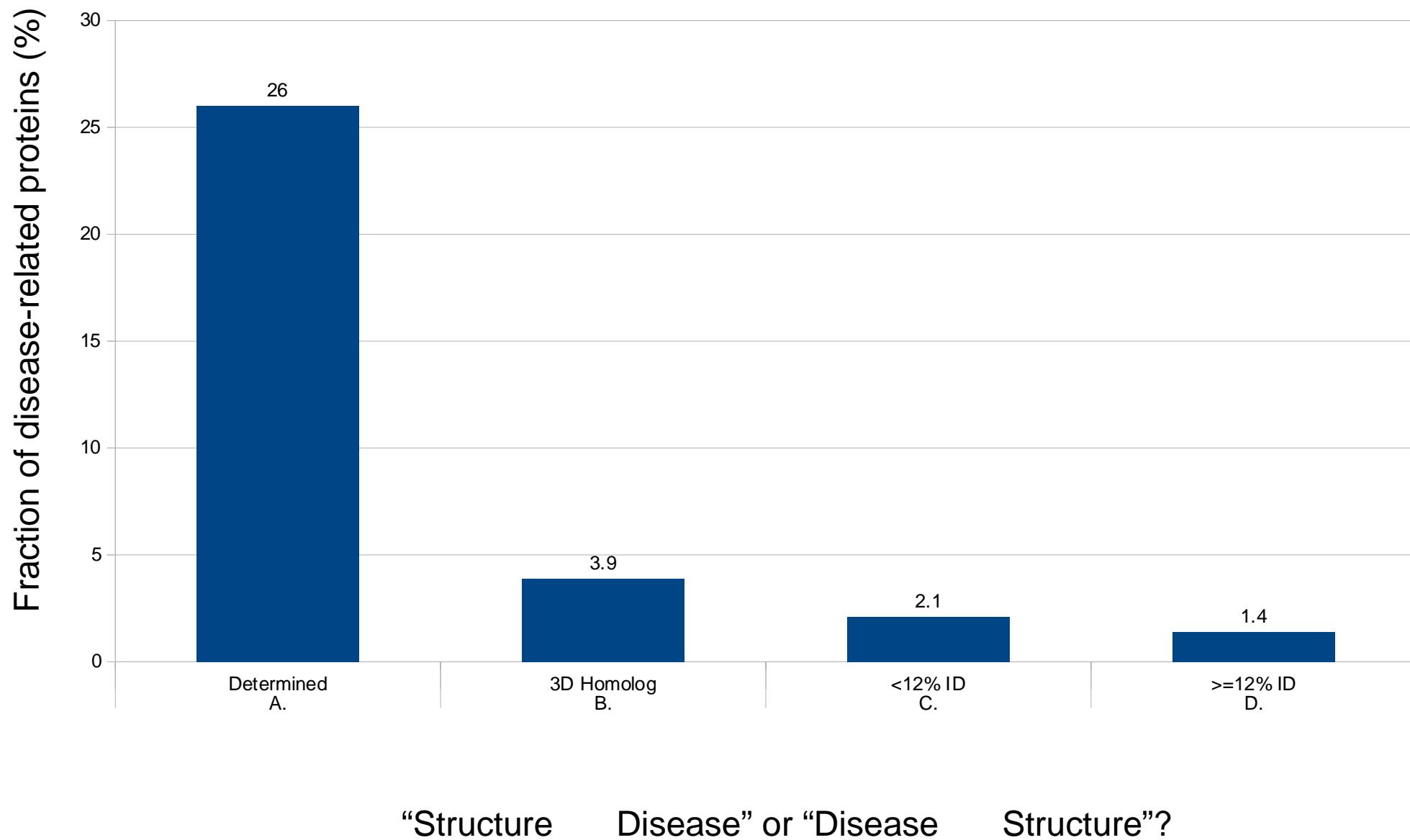




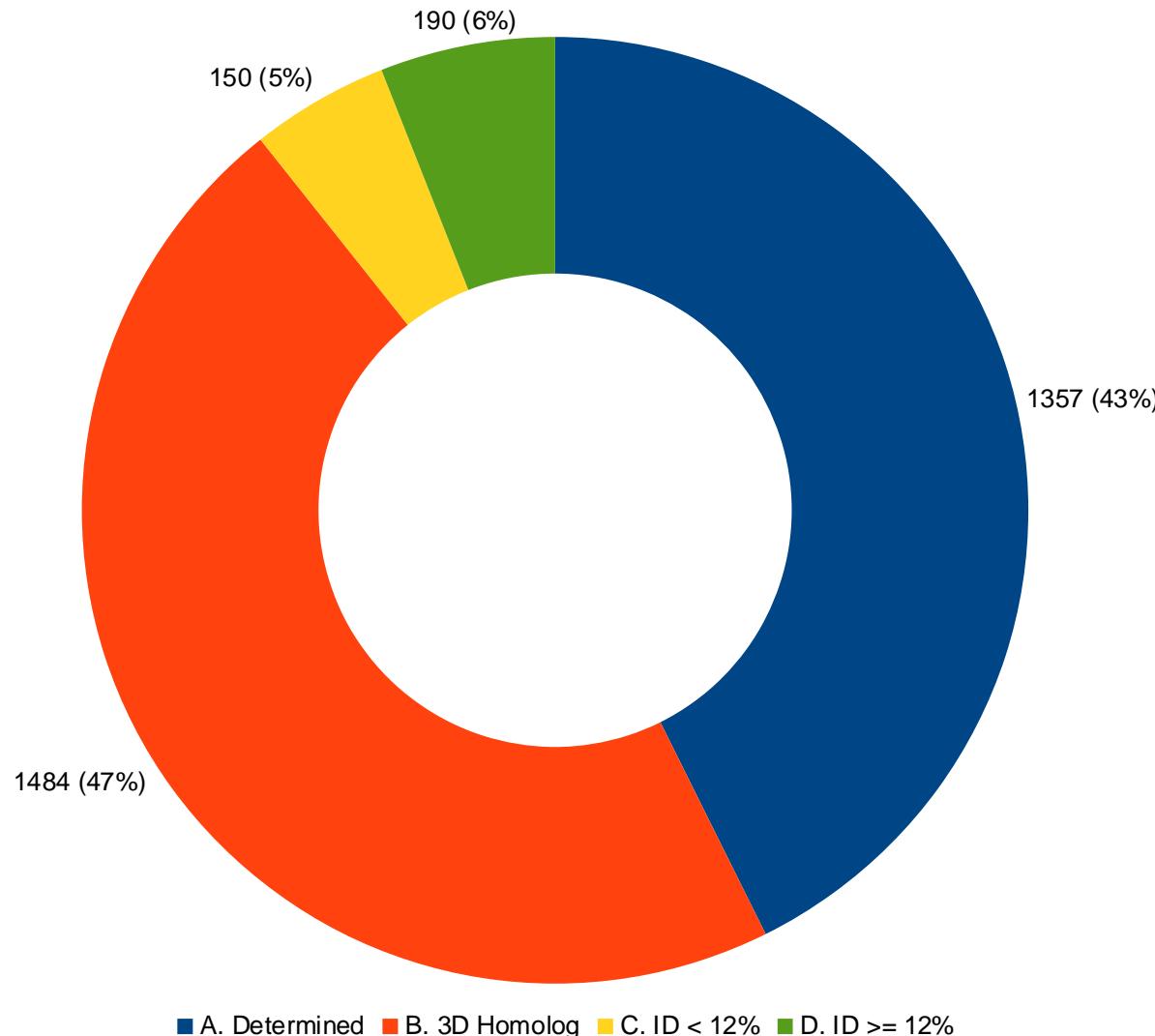
Mean coverage of sequence:  
60~70%



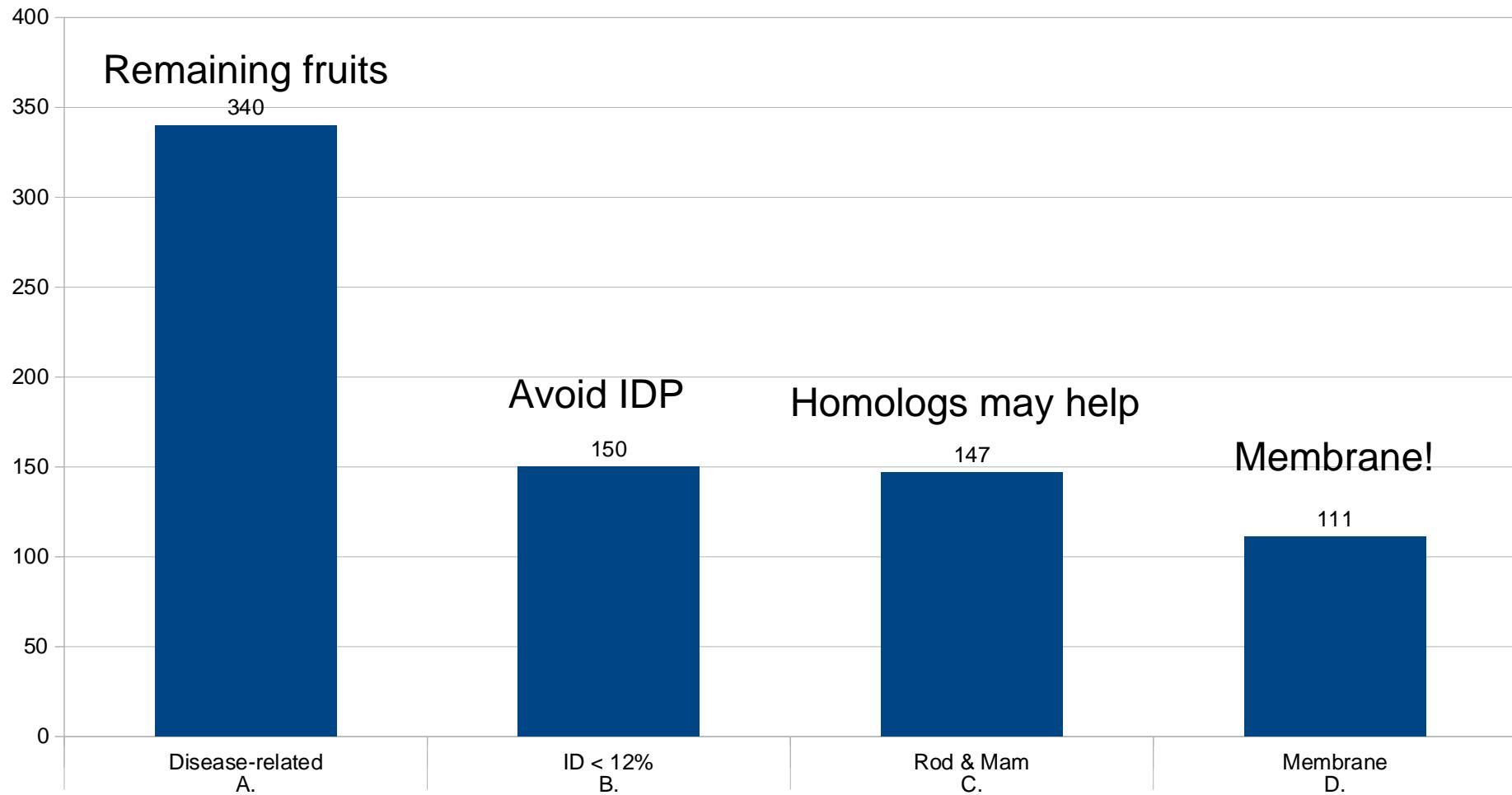
# People love diseases (?)



Out of 3181 disease-related (OMIM) proteins,  
90% have some structural information.



# Screening possible targets for structure determination



IDPs are not negligible.

## Some examples of disease-related membrane proteins with no structure

Accession	Name	# diseases	Diseases (e.g.)
P23942	Peripherin-2	6	CHOROIDAL DYSTROPHY, CENTRAL AREOLAR 2, etc
Q8WVP7	Limb region 1 protein homolog	4	SYNDACTYLY, TYPE IV, etc.
P49279	Natural resistance- associated macrophage protein 1	2	MYCOBACTERIUM TUBERCULOSIS, etc
Q8TDI8	Clarin-1	2	USHER SYNDROME, TYPE IIIA, etc.
Q9HCJ1	Progressive ankylosis protein homolog	2	CHONDROCALCINOSIS 2, etc.

- Grasping trends.
- Screening potential targets.
- Screening potential drugs.
- and much more.