

Homology 3D modeling

Miguel Andrade

Faculty of Biology,

Institute of Organismic and Molecular Evolution

Johannes Gutenberg University

Mainz, Germany

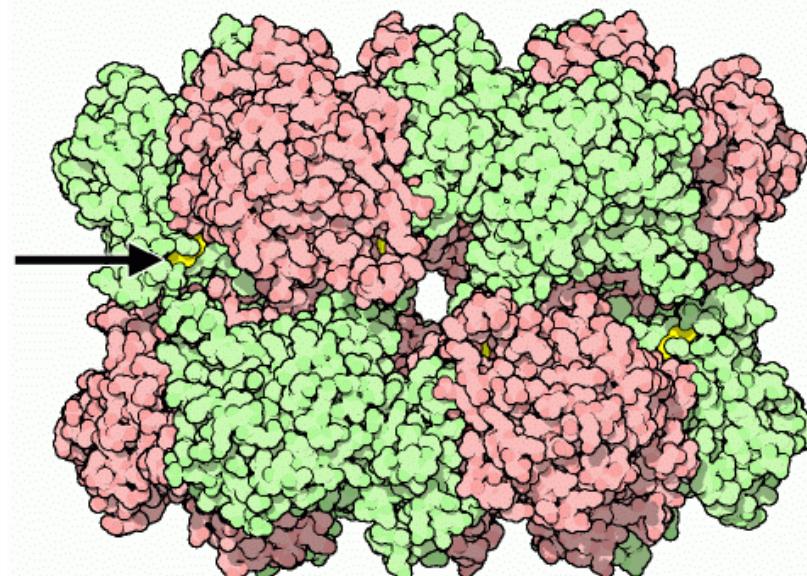
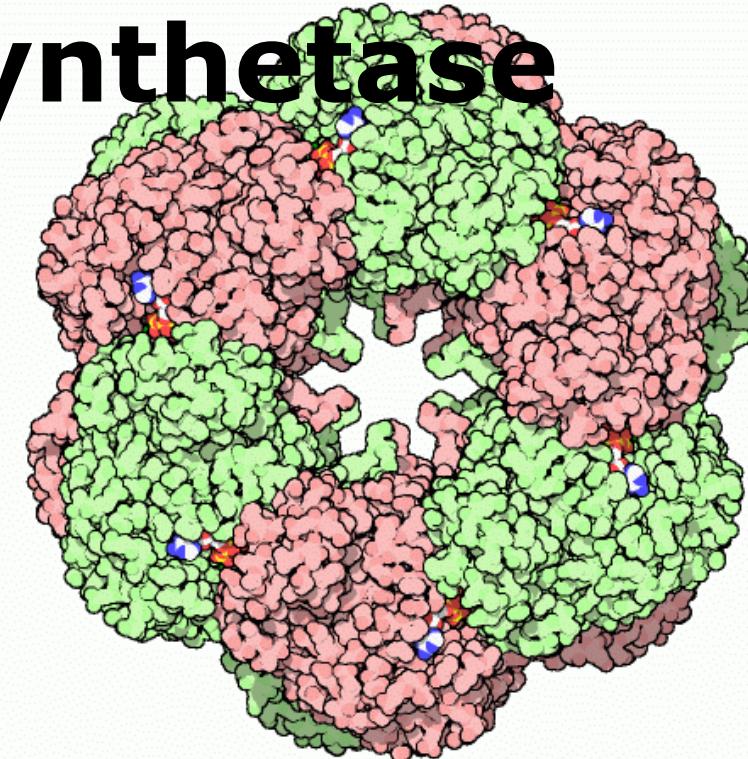
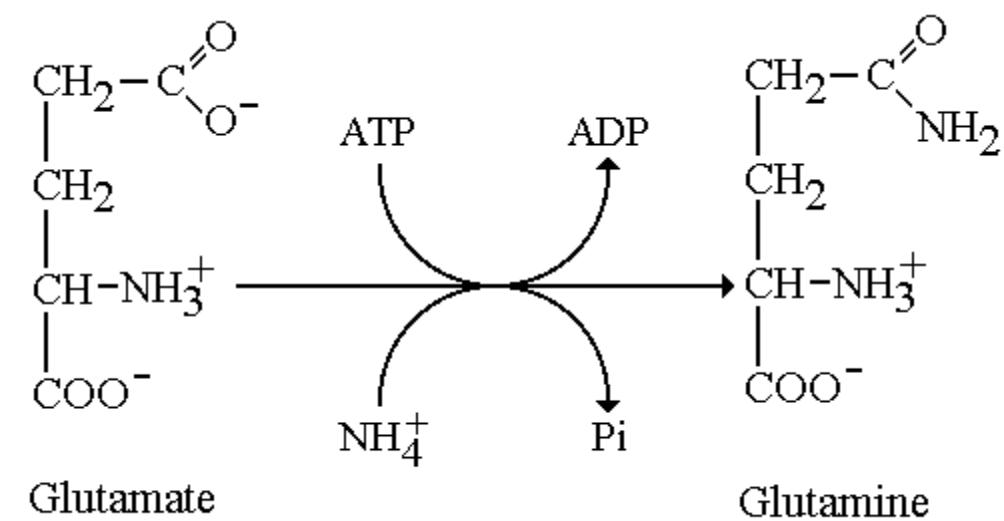
andrade@uni-mainz.de

Mount Everest



Age: 60M years

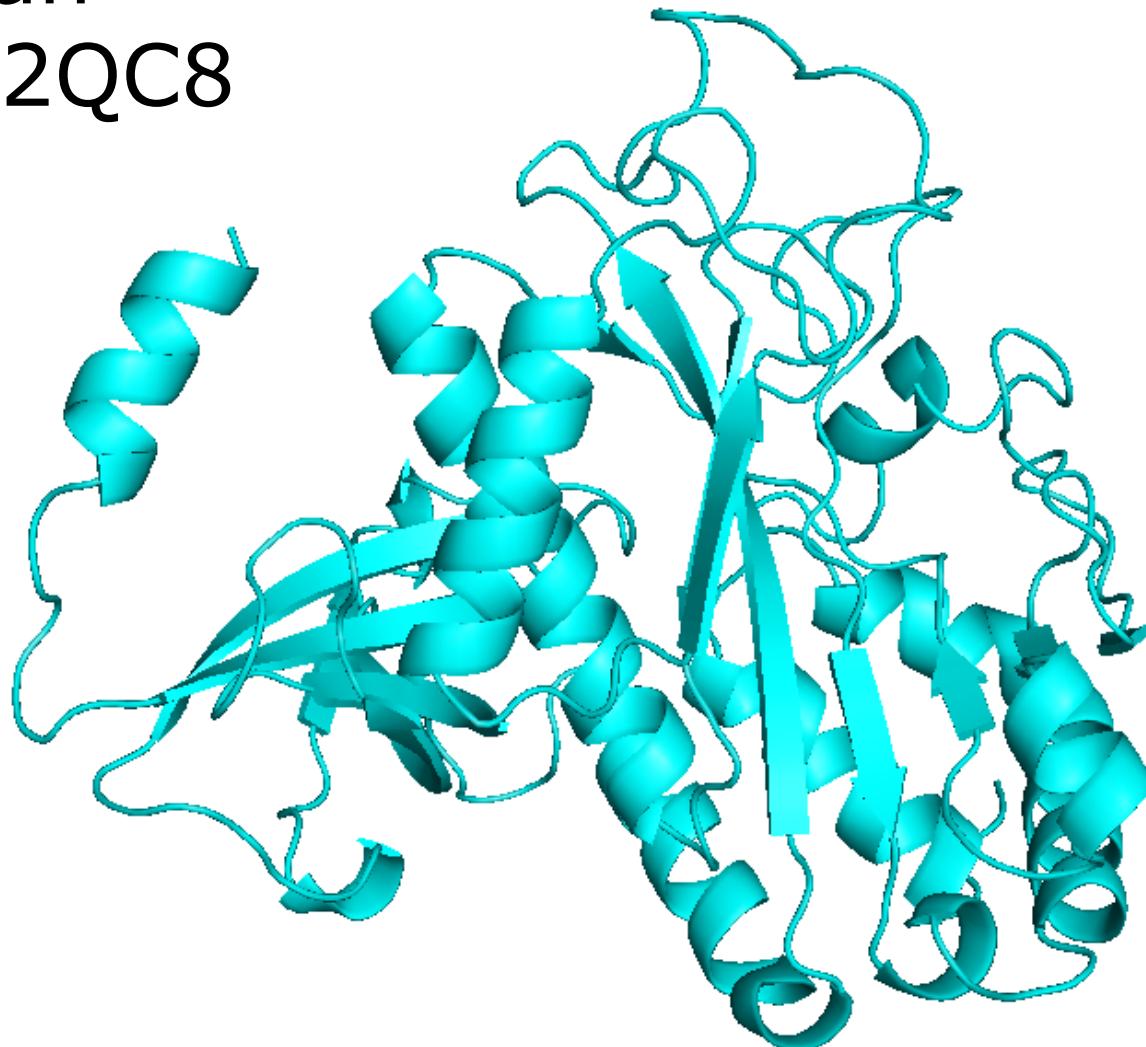
Glutamine synthetase



Age: +3500M years

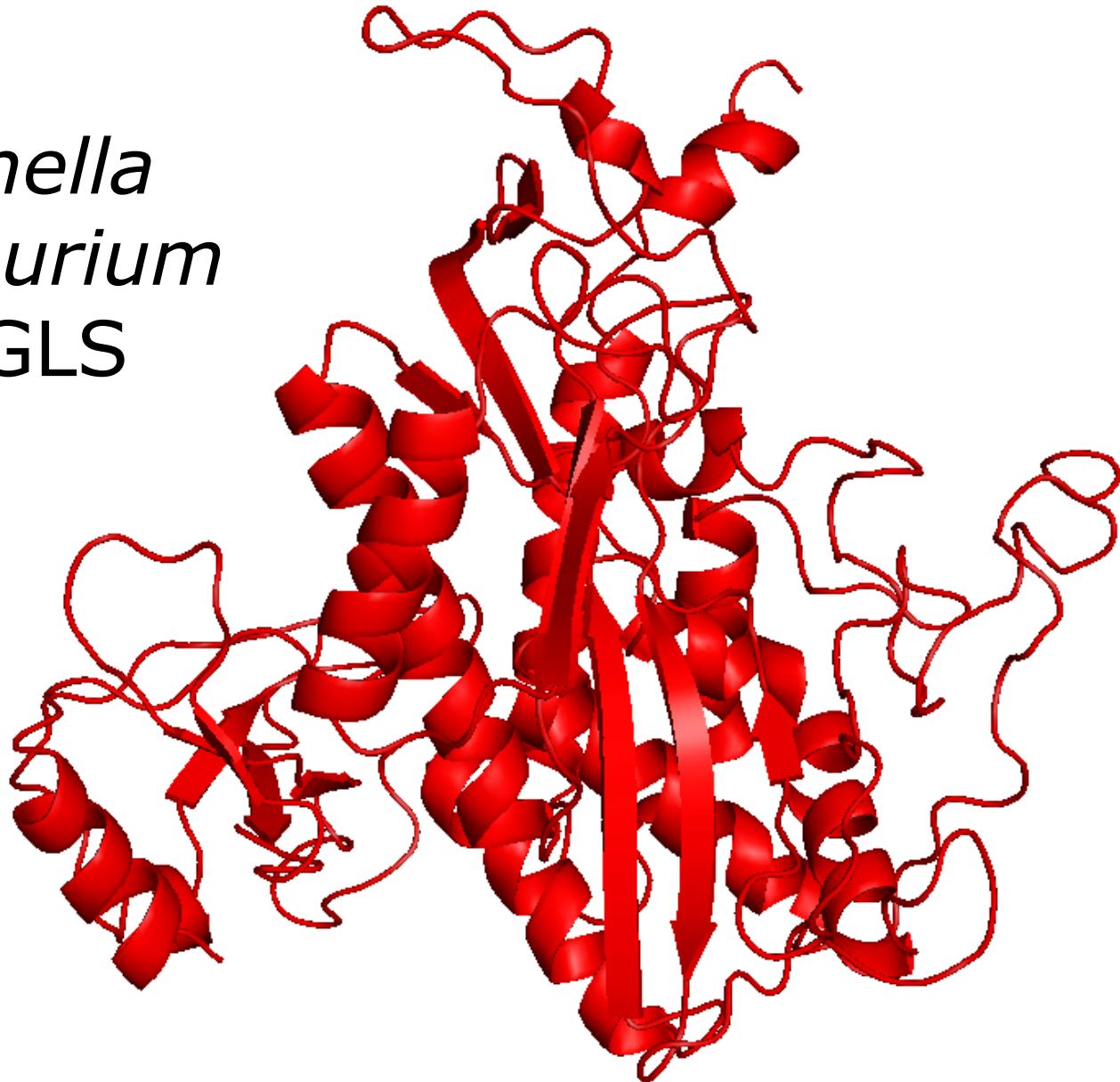
Glutamine synthetase

Human
PDB:2QC8

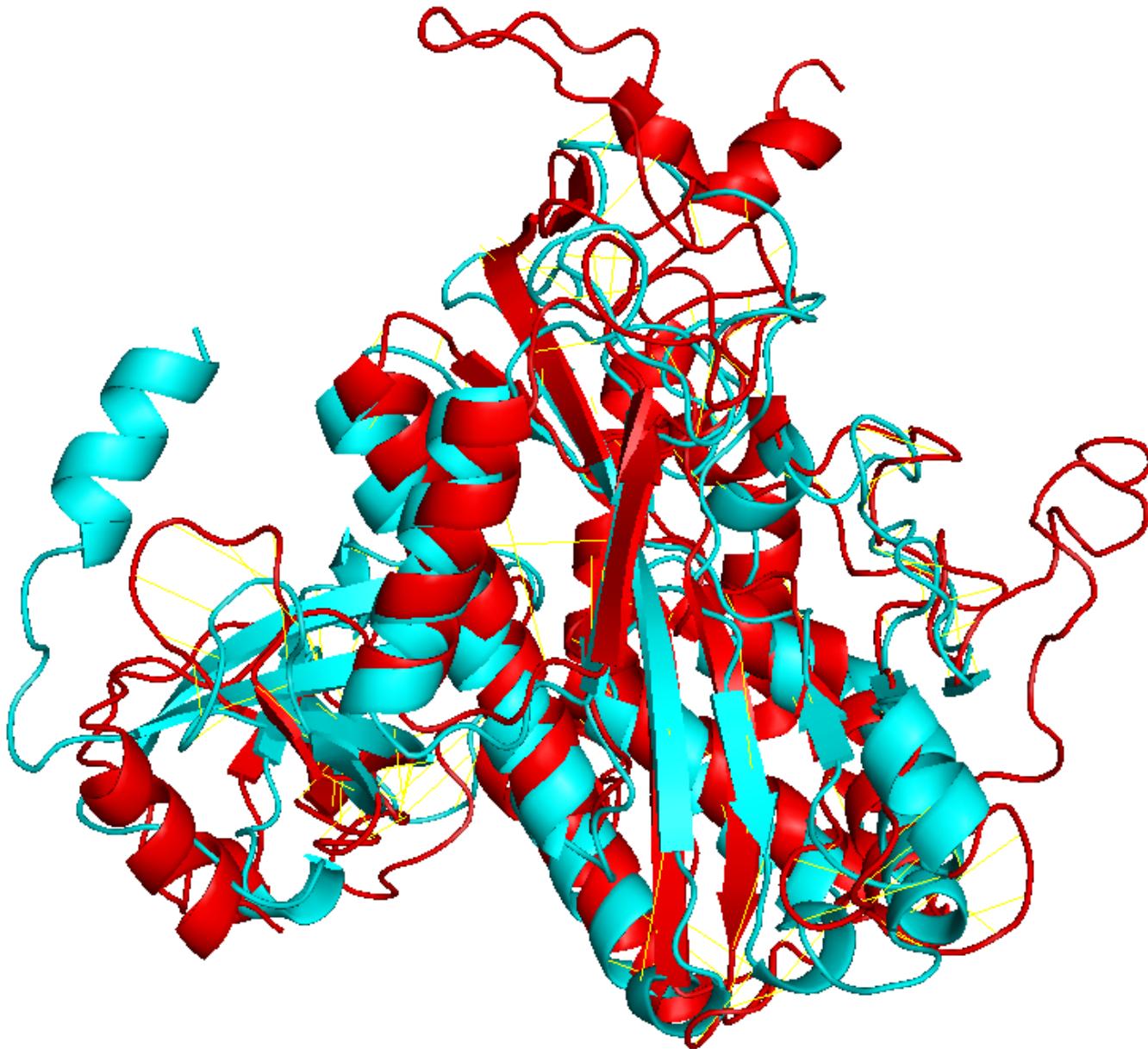


Glutamine synthetase

Salmonella
typhimurium
PDB:2GLS



Glutamine synthetase



Time line

Earth: 4.6 By

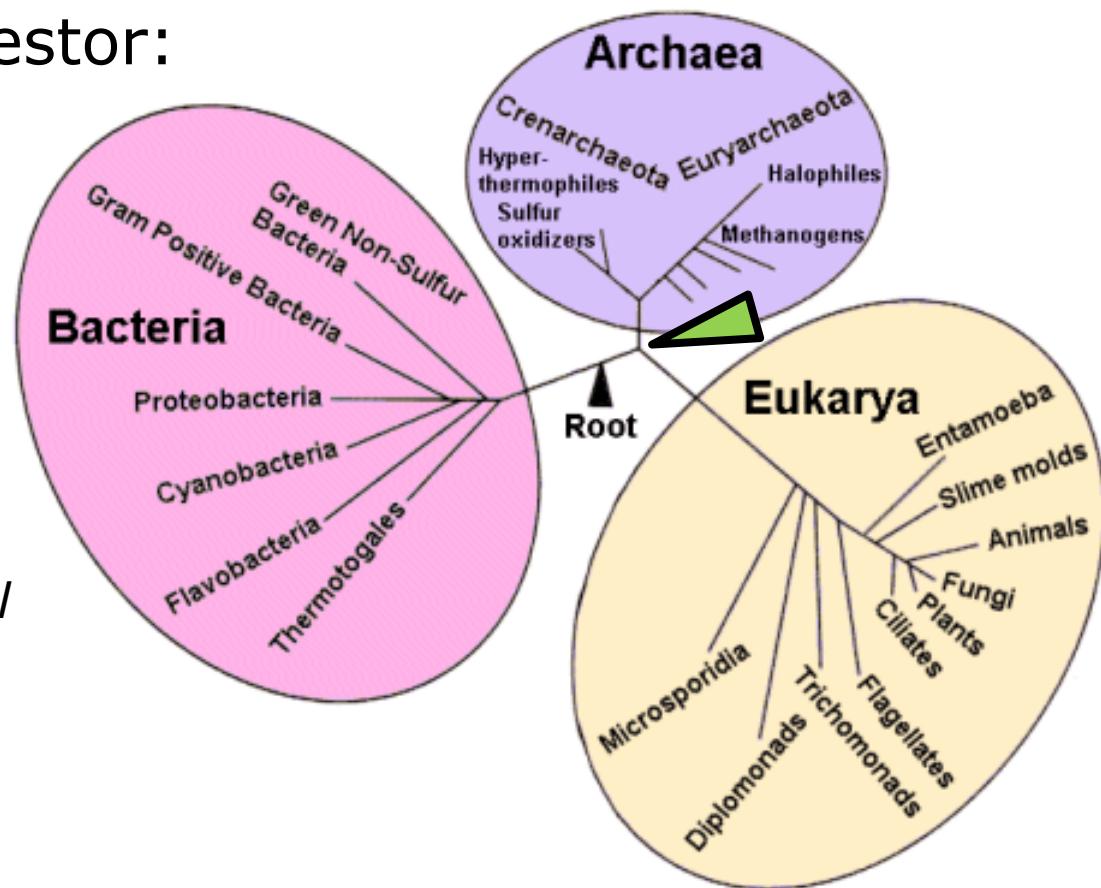
Origin of life: 3.9 By – 3.5 By

Last Common Ancestor:
3.5 – 3.8 By

Glansdorff & Labedan
(2008) *Biology Direct*

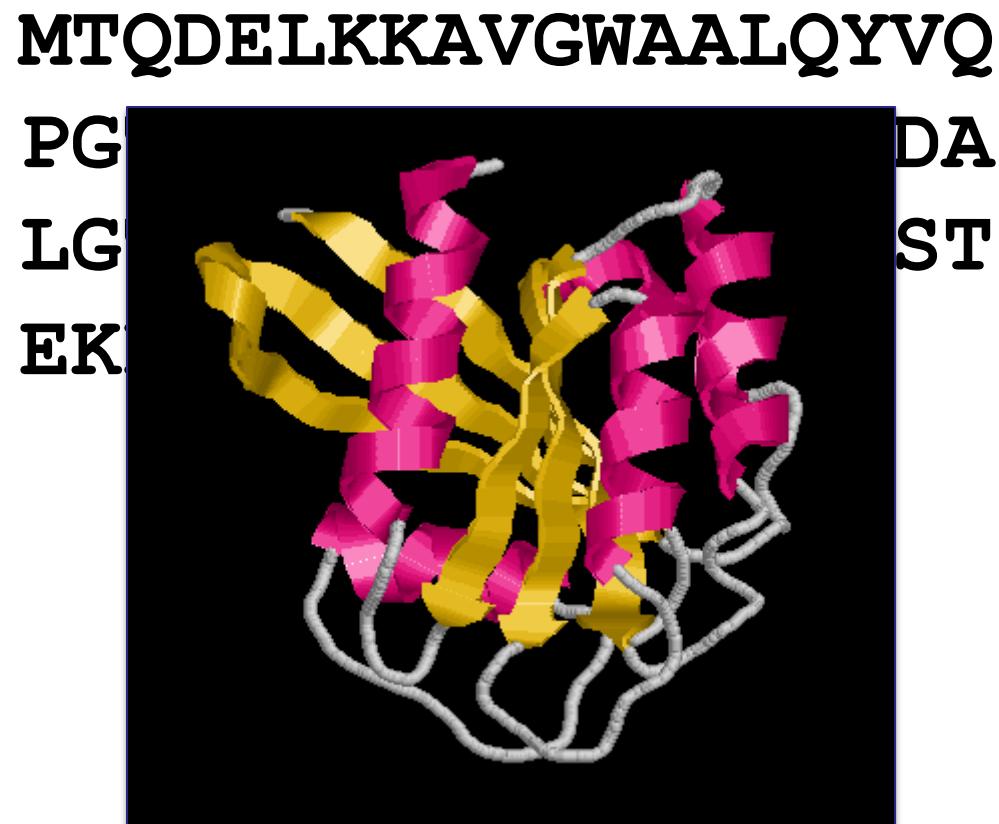
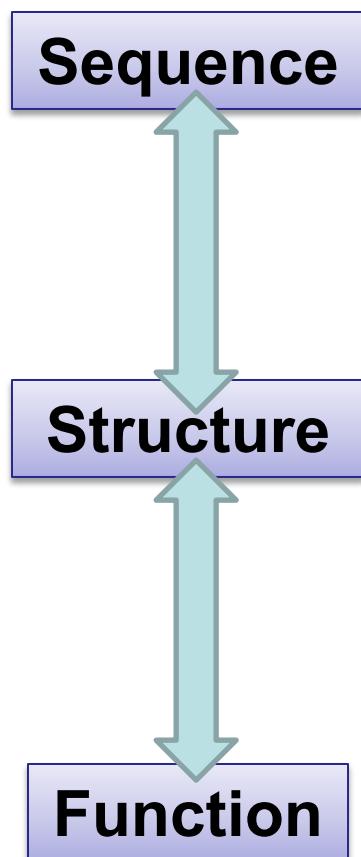
4.29 By

Sheridan et al. (2003)
Geomicrobiology Journal



Sequence and function

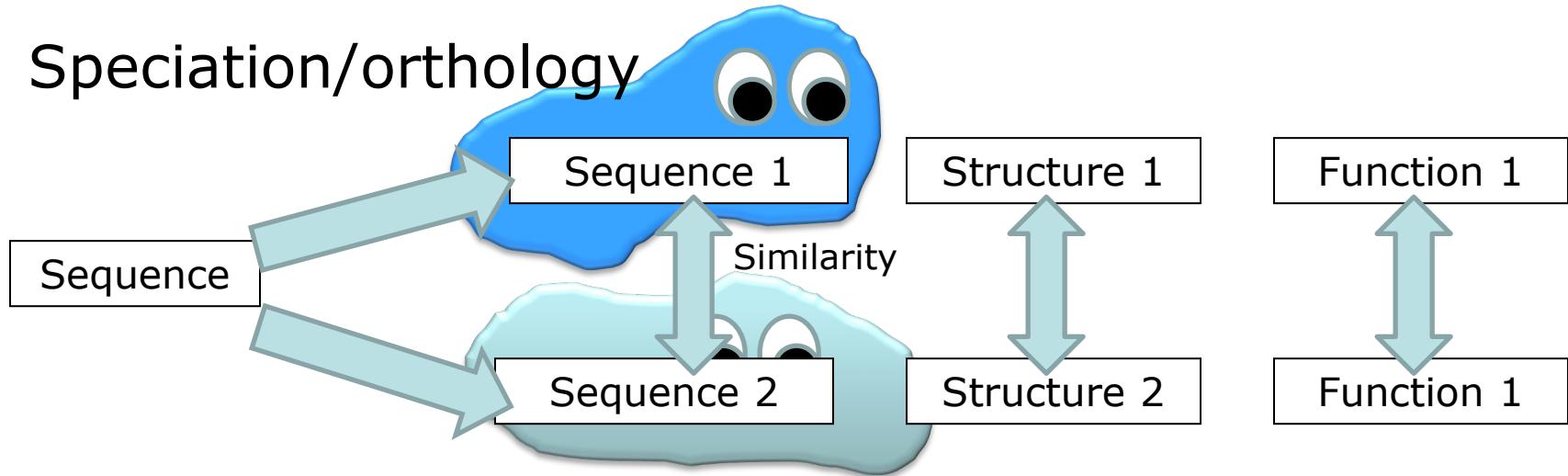
Evolutionary constraints



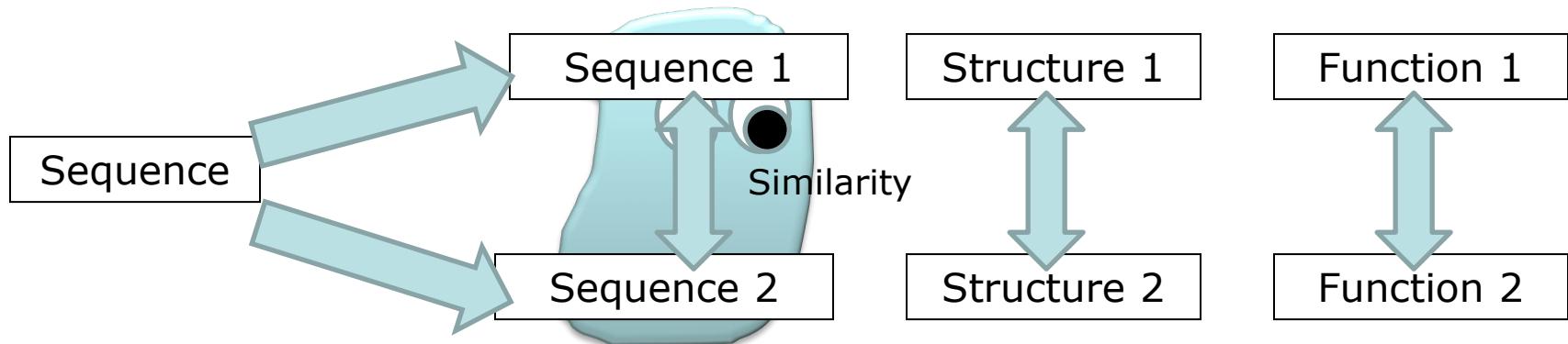
Sequence and function

Evolutionary constraints

Speciation/orthology



Gene duplication/paralogy



Sequence pairwise alignment

```
>gs_human gi|74271837|ref|NP_001028216.1| glutamine synthetase [Homo sapiens]
MTTSASSHLNKGIKQVYMSLPQGEKVQAMYIWIDGTGEGLRCKTRTLDSEPKCVEELPEWNFDGSSTLQS
EGSNSDMYLVPAAMFRDPFRKDPNKLVLCEVFKNRRPAETNLRHTCKRIMDMVSNQHPWFGMEQEYTL
GTDGHPGWPSNGFPGPQGPYYCGVGADRAYGRDIVEAHYRACLYAGVKIAGTNAEVMPAQWEFQIGPCE
GISMGDHlwvarfilhrvcedfgviatfdpkpipingwnngagchtnfstkamreenglkyieeaieklskr
HQYHIRAYDPKGGLDNARRLTGFHETSNINDFSAGVANRSASIRIPRTVGQEKKGYFEDRRPSANCDPFS
VTEALIRTCLLNETGDEPFQYKN
```

```
>gs_salmonella gi|16767272|ref|NP_462887.1| glutamine synthetase [Salmonella
enterica subsp. enterica serovar Typhimurium str. LT2]
MSAEHVLTMLNEHEVKFVDLRFTDTKGKEQHVTIPAHQVNAEFFEEGKMFDGSSIGGWKGINESDMVLMP
DASTAVIDPFFADSTLIIRCDILEPGTLQGYDRDPRSIAKRAEDYL RATGIADTVLFGPEPEFFLFDDIR
FGASISGSHVAIDDIEGAWNSSTKYEGGNKGHRPGVKGGYFPVPPVDSAQDIRSEMCLVMEQMGLVVEAH
HHEVATAGQNEVATRFNTMTKADEIQIYKVVHNVAHFGKTATFMPKPMFGDNGSGMHCHMSIAKNGT
NLFGDKYAGLSEQALYYIGGVVIKHAKAINALANPTTNSYKRLVPGYEAPVMLAYSARNRSASIRIPVVA
SPKARRIEVRFDPAAANPYLCFAALLMAGLDGIKNKIHPGEAMDKNLYDLPEEAKEIPQVAGSLEEALN
ALLDLDREFLKAGGVFTDEAIDAYIALRREEDDRVMTPHPVEFELYYSV
```

Sequence pairwise alignment

BLAST (Altschul et al, 1990)

>lcl|39919 unnamed protein product

Length=469

Score = 70.5 bits (171), Expect = 1e-17, Method: Compositional matrix adjust.
Identities = 102/363 (28%), Positives = 138/363 (38%), Gaps = 96/363 (26%)

Query	62	FDGSSTLQSEGSN-SDMYLVPA--MFRDPFRKDPNKLVLCEVFK-----YNRRP----	108
		FDGSS +G N SDM L+P A DPF D ++ C++ + Y+R P	
Sbjct	50	FDGSSIGGWKGINESDMVLMVDASTAVIDPFFADSTLIIRCDILEPGTLQGYDRDPRSIA	109
Query	109	--AETNLRHTCKRIMDMVSNQHPWFGMEQEYTLMGTDGHPPFGWPSNGF-----	154
		AE LR T I D V FG E E+ L D FG +G	
Sbjct	110	KRAEDYLRATG--IADTV---LFGPEPEFFLF--DDIRFGASISGSHVAIDDIEGAWN	160
Query	155	-----PGPQGPYYCGVGADRAYGRDI-----VEAHYRACLYAG	187
		PG +G Y+ D A +DI VEAH+ AG	
Sbjct	161	SSTKYEGGNKGHRPGVKGGYFPVPPVDSA--QDIRSEMCLVMEQMGLVVEAHHEVATAG	218
Query	188	VKIAGTNAEVMPAQWEFQIGPCEGISMGDHILWVARFILHRVCEDFGVIATFDPKPIPG-N	246
		T M + D + + ++++H V FG ATF PKP+ G N	
Sbjct	219	QNEVATRFNTMTKK-----ADEIQIYKVHHNVAHRGKTATFMPKPMFGDN	265
Query	247	WNGAGCHTNFSTKAMREENGLKYIEEAIKEKLSKRHQYHIRAYDPKGGLDNA-----	297
		+G CH + + +G KY LS++ Y+I NA	
Sbjct	266	GSGMHCHMSLAKNGTNLFSGDKY----AGLSEQALYYIGGVIKHAKAINALANPTTNSY	320
Query	298	RRLTGFHETSNIIDFSAGVANRSASIRIPRTVGQEKKGYFEDRRPSANCDFPSVTEALIR	357
		+RL +E + +SA NRSASIRIP V K E R P +P+ AL+	
Sbjct	321	KRLVPGYEAPVMLAYSA--NRNSASIRIP-VVASPKARRIEVRFPDPAANPYLCFAALLM	377
Query	358	TCL 360	
		L	
Sbjct	378	AGL 380	

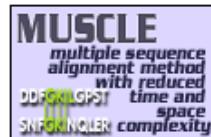
Multiple sequence alignment

```
>gs_human gi|74271837|ref|NP_001028216.1| glutamine synthetase [Homo sapiens]
MTSASSHLNKGKQVYMSLPQGEKVQAMYIWIDGTGEGLCKTRTLDEPKCVELPEWNFDGSSTLQS
EGSNSDMLVPAAMFRDPFRKDPNKLVLCEVFKNRRPAETNLRHTCKRIMDMVSNQHPWFGMEQEYTL
GTDGHPFGWPSNGFPGPQGPYYCGVGADRAYGRDIVEAHYRACLYAGVKIAGTNAEVMPAQWEFQIGPCE
GISMGDHLWVARFILHRVCEDFGVIATFDPKPIPENWGAGCHTNFSTKAMREENGKYIEEAIEKLSKR
HQYHIRAYDPKGGLDNARRLTGFHETSNINDFSAGVANRSASIRIPRTVGQEKKGYFEDRRPSANCDPFS
VTEALIRTCLLNGETGDEPFQYKN

>gs_vulca gi|307594850|ref|YP_003901167.1| glutamine synthetase [Vulcanisaeta
distributa DSM 14429]
MPTRNLEIEPADLWRILKASGIKYVKFIIVDINGAPRSEIVPIDMAKDLFIDGMPFDASSIPSYSTVNKS
DFVAYVDPRAVYVEYWQDGKVADVFTMVDIAKPSPLDPVVNLDAEQARSKGYEFLMGVEVEFFVIK
EDGGKPVADPGIYFDGWNVTVQSQFMKELITAIADAGINYTKTHEVAPSQYEVNIGATDPLRLADQIV
YFKIMAKDIARKYGLVATFMPKPFWGVNGSGAHTHISVWKDGKNLFQSSTGKITEECGYAISAILSARA
LSSFVAPLVNSYKRLVPHYEAPTRIVWGYANRSAMIRIPQYKMRINRIEYRHPDPSMNPYLAFTAIIKTM
IRGLEEKKEPPPTEEVAYELANALETPATLEDLKELSKSFLATELPSELVNAYIKIKQNEWEDYLTNV
GPWEKTWNIIITQWEYNKYLVT

>gs_salmonella gi|16767272|ref|NP_462887.1| glutamine synthetase [Salmonella
enterica subsp. enterica serovar Typhimurium str. LT2]
MSAEHVLTMLNEHEVKFVDLRFTDTKGKEQHVTIPAHQVNAEFFEEGKMFDSIGGWKGGINESDMVLMP
DASTAVIDPFFADSTLIIRCDELEPGTLQGYDRDPRSIAKRAEDYL RATGIADTVLFGPEPEFFLFDDIR
FGASISGSHVAIDDIEGAWNSSTKYEGGNKGHRPGVKGGYFPVPPVDSAQDIRSEMCLVMEQMGLVVEAH
HHEVATAGQNEVATRFNTMTKKADEIQIYKVHHNVHNRFGKTATFMPKPMFGDNGSGMHCHMSLAKNGT
NLFGSGDKYAGLSEQALYYIGGVIKHAKAINALANPTTNSYKRLVPGYEAPVMLAYSARNRSASIRIPVVA
SPKARRIEVRFPDPAANPYLCFAALLMAGLDGIKNKIHPGEAMDKNLYDLPPEEAKEIPQVAGSLEEALN
ALLDLDREFLKAGGVFTDEAIDAYIALRREDDRVRMTPHPVEFELYYSV

>gs_yeast gi|330443748|ref|NP_015360.2| Gln1p [Saccharomyces cerevisiae S288c]
MAEASIEKTQILQKYLELDQRGRIIAEYVWIDGTGNLRSKGRTLKKRITSIDQLPEWNFDGSSTNQAPGH
DSDIYLKPVAYYPDPFRRGDNIVVIAACYNNNDGTPNKFNHRHEAAKLFAAHKDEEIWFGLEQEYTLFDMY
DDVYGPWKGGYPAPQGPYYCGVGAGKVYARDMIEAHYRACLYAGLEISGINAEVMPSQWEFQVGPCTGID
MGDQLWMARYFLHRVAEEFGIKISFHPKPLKGDWNGAGCHTNVSTKEMRQPQGMKYIEQAIEKLSKRHAE
HIKLYGSDNDMRLTGRHETASMTAFSSGVANRGSSIRIPRSVAKEGYGYFEDRRPASNIDPYLVTGIMCE
TVCGAIDNADMTEFERESS
```



EBI > Tools > Multiple Sequence Alignment > MUSCLE

MUSCLE - Multiple Sequence Alignment

MUSCLE stands for MULTiple Sequence Comparison by Log-Expectation. MUSCLE is claimed to achieve both better average accuracy and better speed than ClustalW2 or T-Coffee, depending on the chosen options.

Internet Explorer users: If button presses (including copy/paste operations) don't appear to work please try enabling Compatibility View.

Use this tool

STEP 1 - Enter your input sequences

Enter or paste a set of sequences in any supported format:

Or upload a file: No file chosen

STEP 2 - Set your Parameters

OUTPUT FORMAT:

The default settings will fulfill the needs of most users and, for that reason, are not visible.

(Click here, if you want to view or change the default settings.)

STEP 3 - Submit your job

Be notified by email *(Tick this box if you want to be notified by email when the results are available)*

<http://www.ebi.ac.uk/Tools/msa/muscle/>

>gs_human gi|74271837|ref|NP_001028216.1| glutamine synthetase [Homo sapiens]
MTTSASSHLNKGKQVYMSLPQGEKVQAMYIWIDGTGEGLRCKTRTLDSEPKCVELPEW
N-FDGSSLQSEGSNSD---MYLVPAAMFRDPFRKDPNKVLCEVFKYNRPA-ETNLRH
TCKRIMDMVSNQH---PWFGMEQEYTLMGT-----DGHPFGW-----
-PSNGFPGPQGP--YYCGVGADRAYGRDIVEAHYRACLYAGVKIAGTNAEVMPA-QWEFQ
IGPCEGISMGDHLWVARFILHRVCEDFGVIATFDPKPIPGNWNGAGCHTNFKAMREEN
GLKYIEEAIKEKLSKRHQYHIRAYDPKGG-----LDNARRLTGFHETSNIINDSAGV
ANRSASIRIPRTVGQEKKGYFEDRRPSANCDFPSVTEALIRT-CLLNETGDEP-----

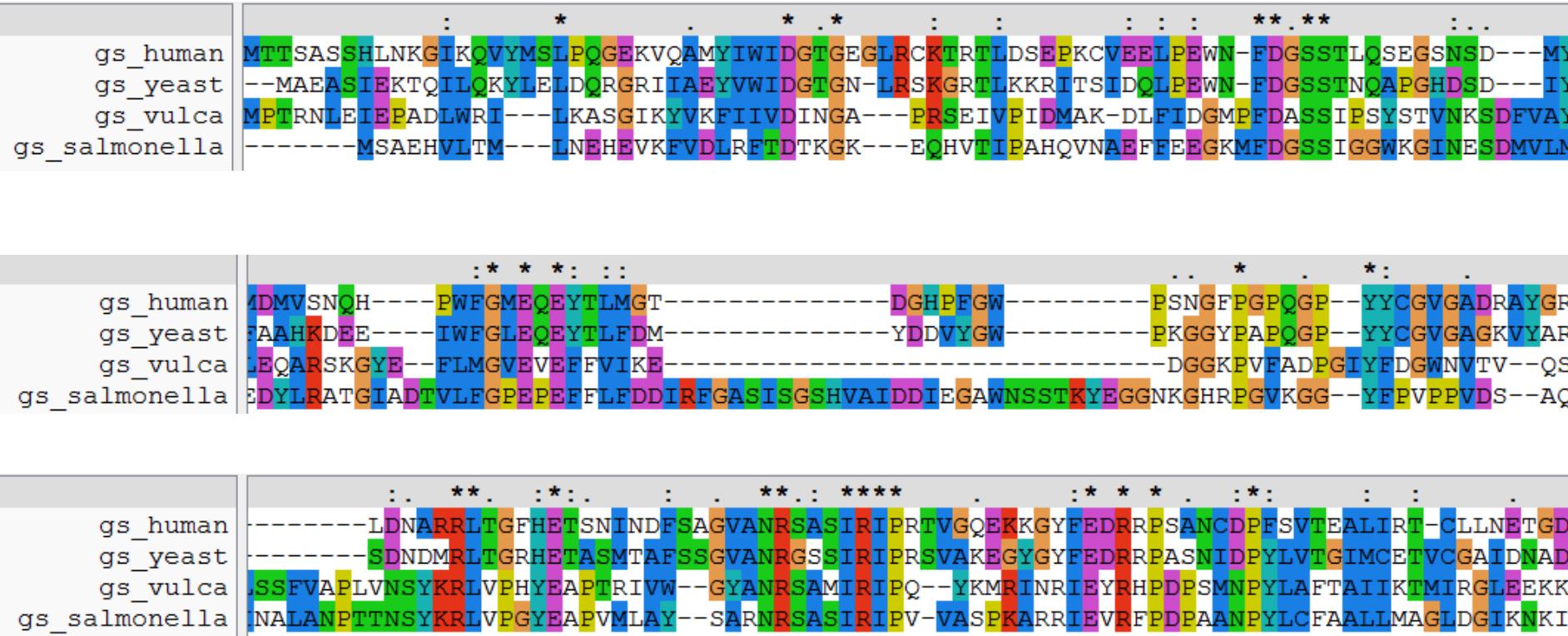
-----FQYKN-----

>gs_yeast gi|330443748|ref|NP_015360.2| Gln1p [Saccharomyces cerevisiae S288c]
--MAEASIEKTQILQKYLELDQRGRIIAEYVWIDGTGN-LRSKGRTLKKRITSIDQLPEW
N-FDGSSTNQAPGHDS---IYLKPVAYYPDPFRRGDNIVVLAACYNNNDGTPN-KFNHRH
EAALKFAAHKDEE---IWFGLEQEYTLFDM-----YDDVYGW-----
-PKGGYPAPQGP--YYCGVGAGKVYARDMIEAHYRACLYAGLEISGINAEVMPS-QWEFQ
VGPCTGIDMGDQLWMARYFLHRVAEEFGIKISFHPKPLKGDWNGAGCHTNVSTKEMRQPG
GMKYIEQQAIEKLSSKRHAEHIKLYG-----SDNDMRLTGRHETASMTAFSSGV
ANRGSSIRIPRSVAKEGYGYFEDRRPASNIDPYLVTGIMCETVCGAIDNADM-----

-----KEFERESS-----

>gs_vulca gi|307594850|ref|YP_003901167.1| glutamine synthetase [Vulcanisaeta distributa DSM 14429]
MPTRNLEIEPADLWRI---LKASGIKYVKFIIVDINGA---PRSEIVPIDMAK-DLFIDG
MPFDASSIPSYSTVNKSDFVAYVDPRAVYVEYWQDGKVADVFTMVSDIADKPS-PLDPRR
VLNDALEQARSKGYE--FLMGVEVEFFVIKE-----
--DGGKPVFADPGIYFDGWNNTV--QSQFMKELITAIADAGINYTKTHEVAPS-QYEVN
IGATDPLRLADQIVYFKIMAKDIARKYGLVATFMPKFWGV-NGSGAHTHIS---VWKDG
KNLF-QSSTGKITEECGYAISAILSARALSSVAPLVNSYKRLVPHYEAPTRIVW--GY
ANRSAMIRIPQ--YKMRINRIEYRHDPDSMNPyLAFTAIIKTMIRGLEEKKEPPPTEEV
AYELA--NALETP---ATLEDTLK--ELSKSFLATE--LPSELVNAYIKIKQNEWEDYLT
NVGPWEKTWNIIITQWEYNKYLVTAA

>gs_salmonella gi|16767272|ref|NP_462887.1| glutamine synthetase [Salmonella enterica
-----MSAEHVLTM---LNEHEVKFVDLRFTDTKGK---EQHVTIPAHQVNAEFFEEG
KMFDGSSIGGWKGINESDMVLMFDASTAVIDPFFADSTLIIRCDILEPGTLQGYDRDPRS
IAKRAEDYLRATGIADTVLFGPEPEFFLFDDIRFGASISGSHVAIDDIEGAWNSSTKYEG
GNKGHRPGVKGG--YFPVPPVDS--AQDIRSEMCLVMEQMGLVVEAHHEVATAGQNEVA
TRFNTMTKKADEIQIYKYVHNVAHRFGKTATFMPKPMFGD-NGSGMHCHMS---LAKNG
TNLFSGDKYAGLSEQALYYIGGVIKHAKAINALANPTTNSYKRLVPGYEAPVMLAY--SA
RNRSASIRIPV-VASPKARRIEVRFPDPAANPYLCFAALLMAGLDGIKNKIHPGEAMDKN
LYDLPPEEAKEIPQVAGSLEEALNALDLDREFLKAGGVFTDEAIDAYIALRREEDDRVRM
TPHP-----VEFELYYSV-



ClustalW, JalView

Determination of protein structure

X-ray crystallography (111K in PDB)

- need crystals

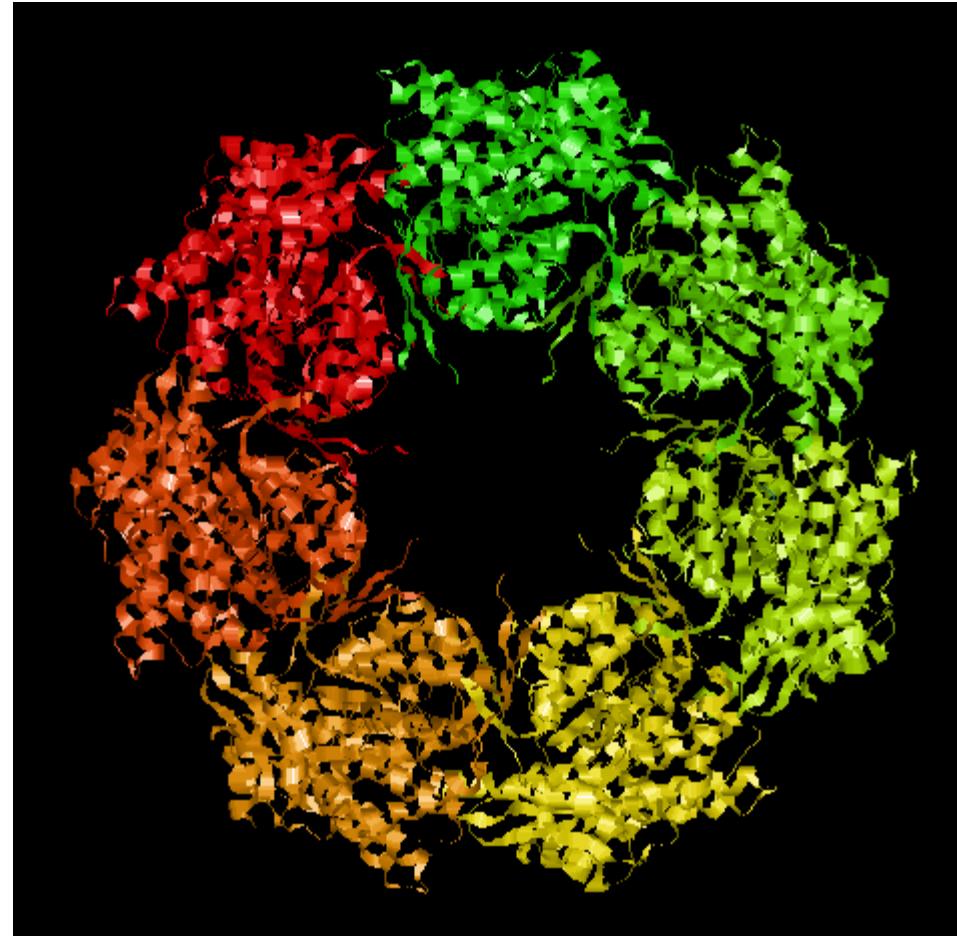
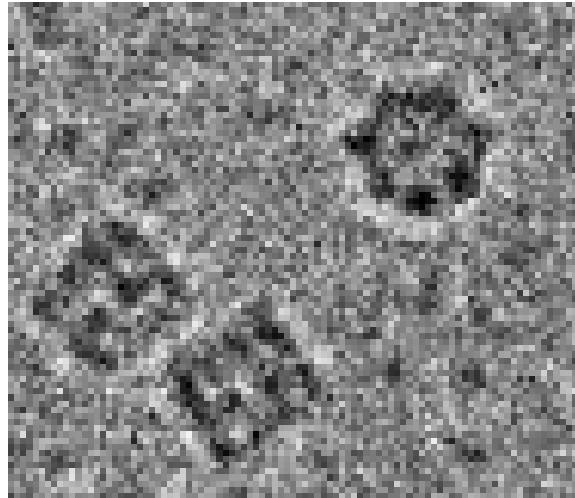
Nuclear Magnetic Resonance (NMR)
(10.4K)

- proteins in solution
- lower size limit (600 aa)

Electron microscopy (1.2K)

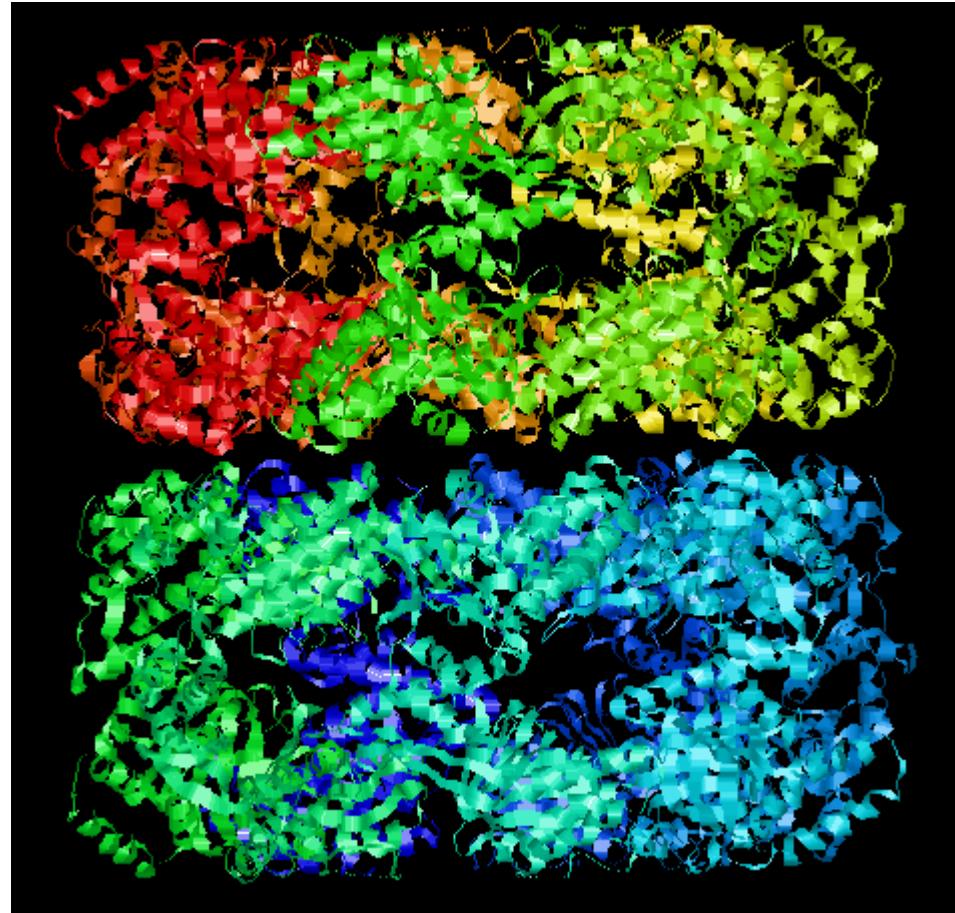
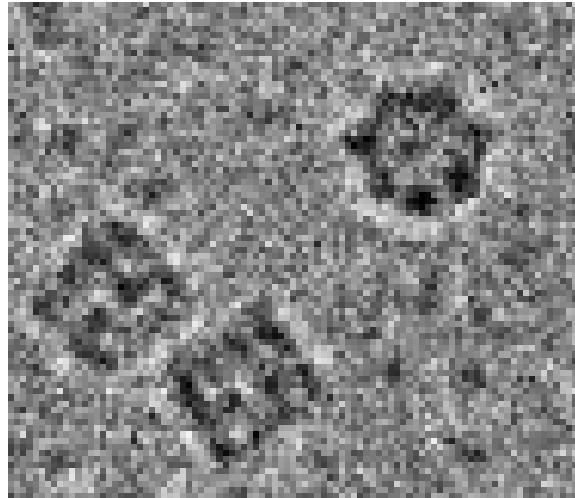
- Low resolution (>5A)

Determination of protein structure



resolution 2.4 Å

Determination of protein structure

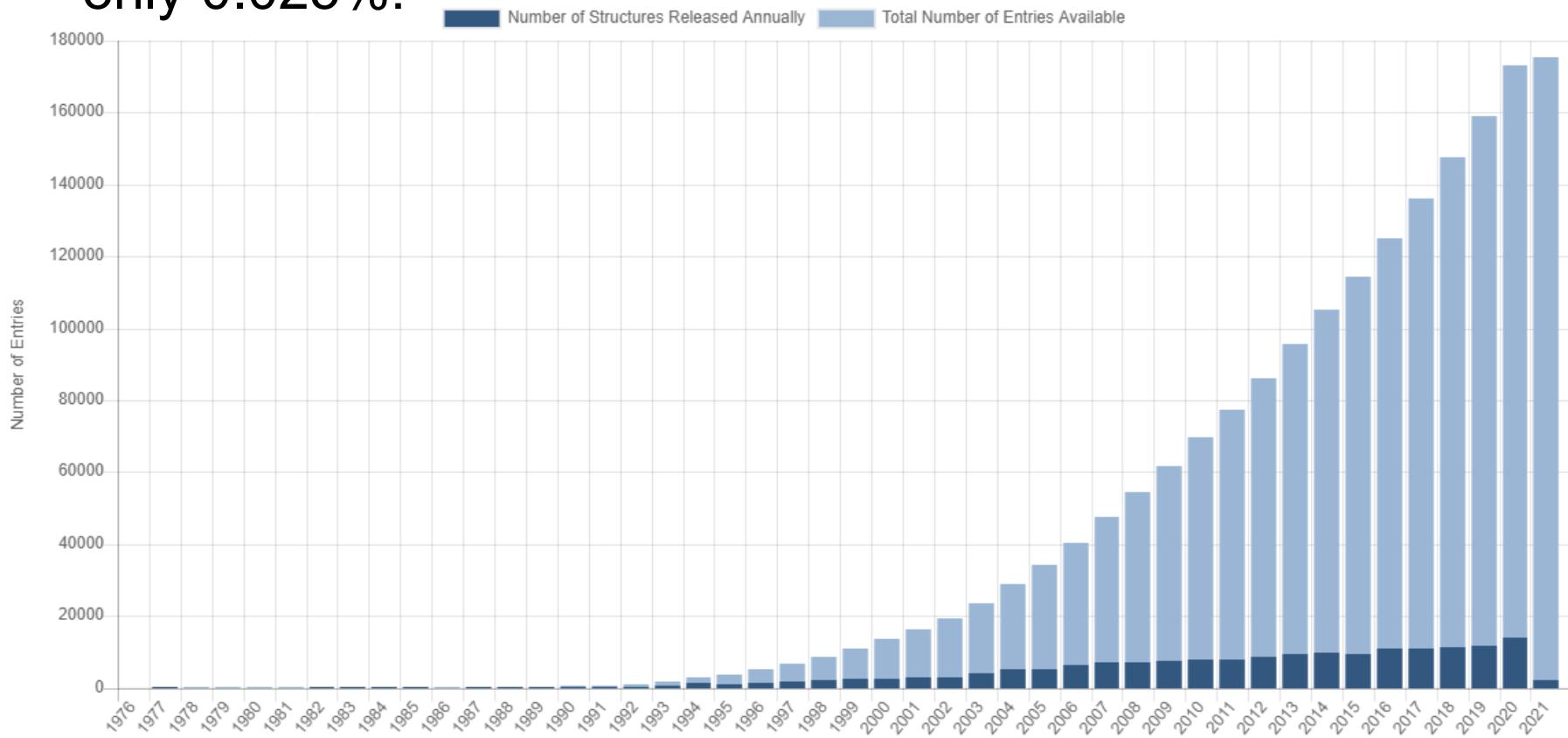


resolution 2.4 Å

Structural genomics

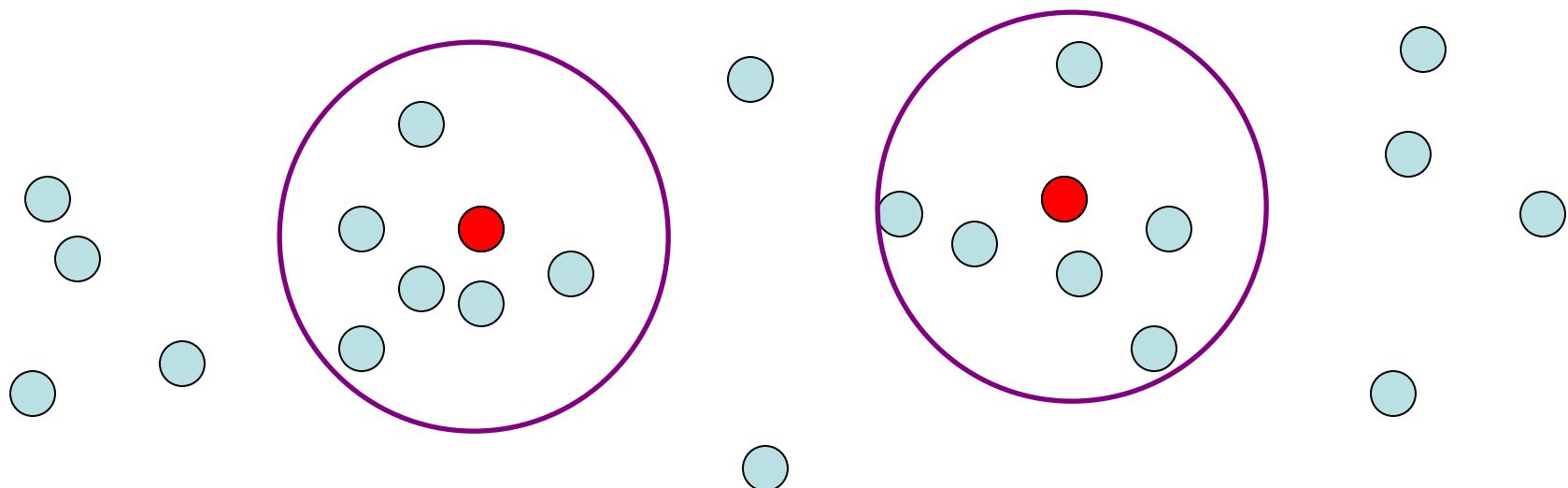
Currently: 175K protein 3D structures
from around 52.7K sequences in UniProt (how do I know?)
208M sequences in UniProt

only 0.025%!



Structural genomics

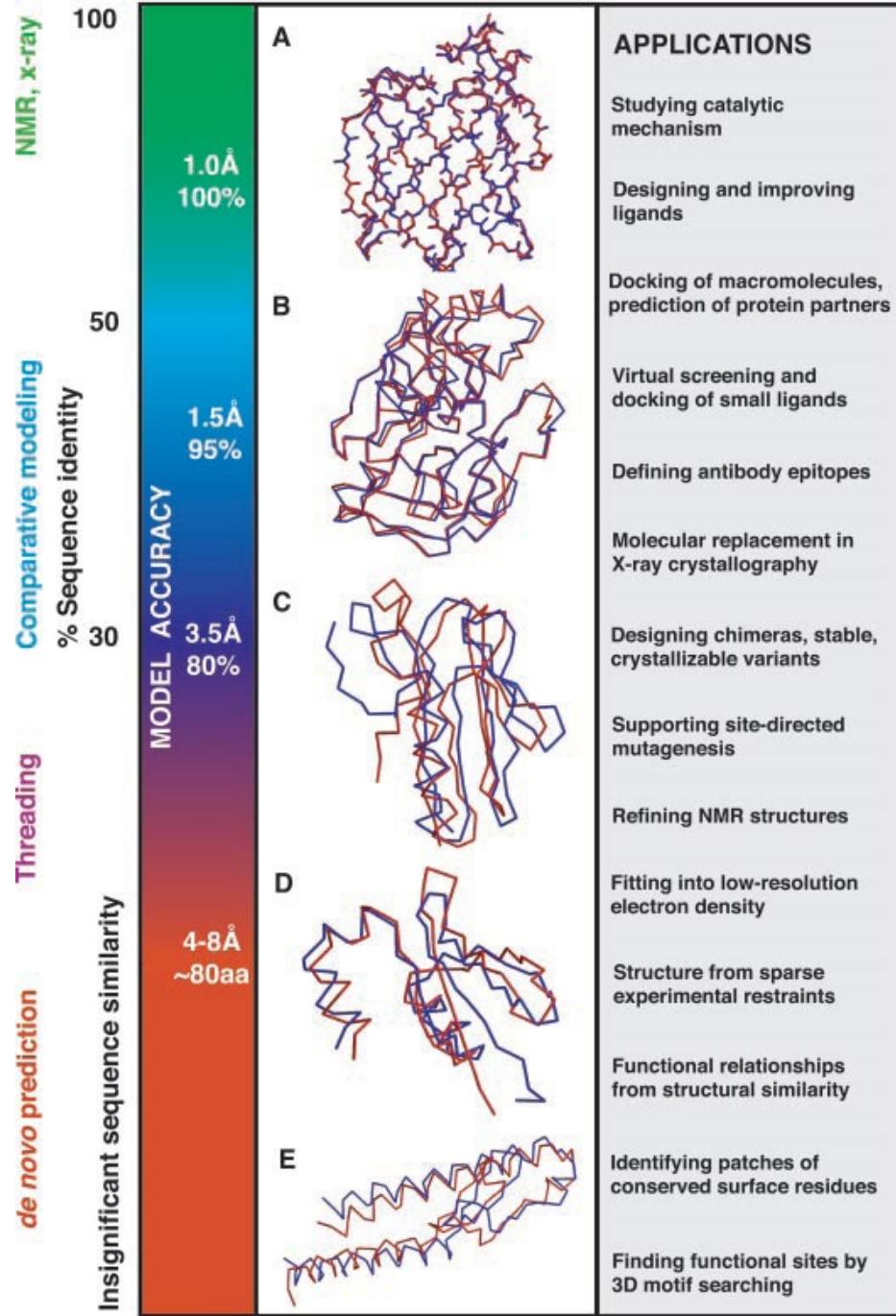
Currently: 175K protein 3D structures
from around 52.7K sequences in UniProt (how do I know?)
208M sequences in UniProt
only 0.025%!



50% sequences covered (25% in 1995)

Relation between sequence identity and accuracy/applications

From:
Baker and Sali (2001)
Science



Homology modelling

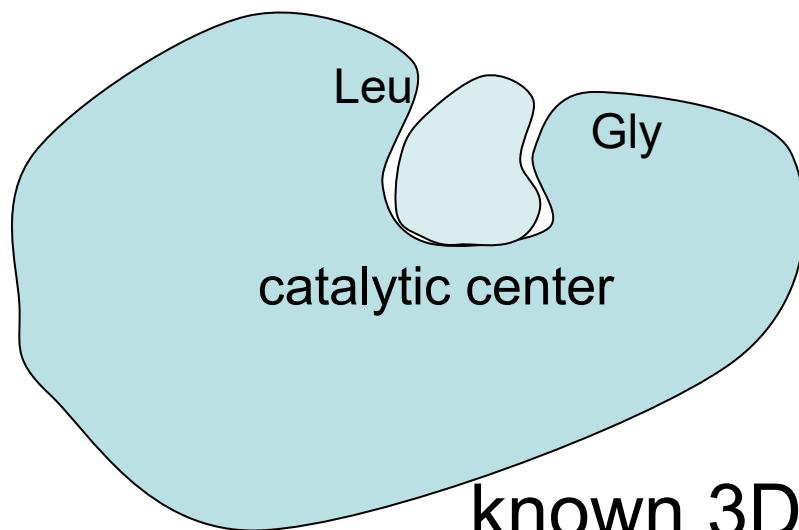
Applications: target design

Query sequence

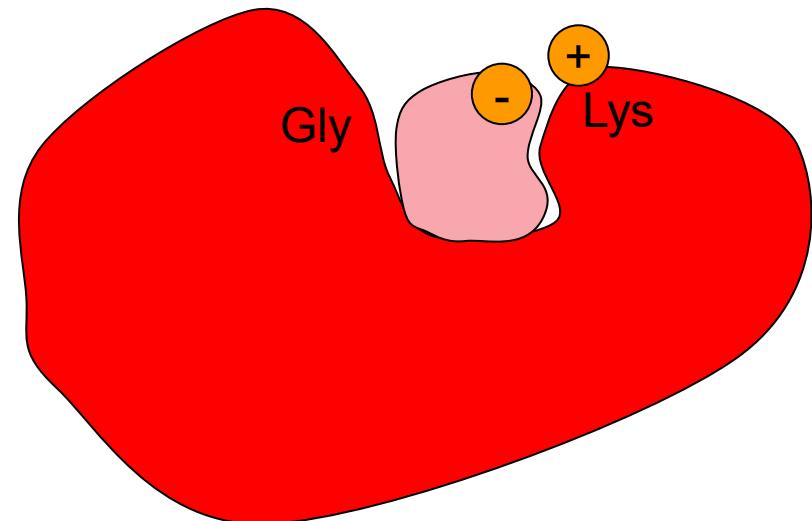
G K

similar to

L G



known 3D

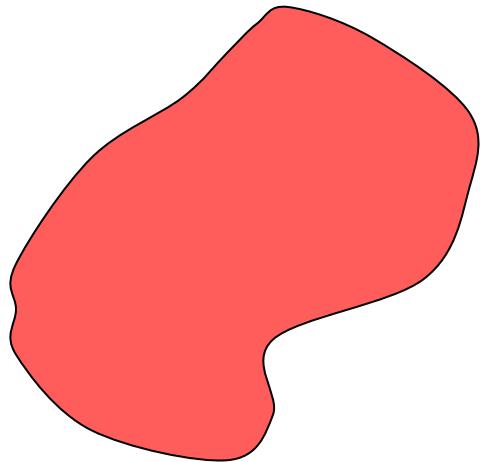


model 3D by
homology

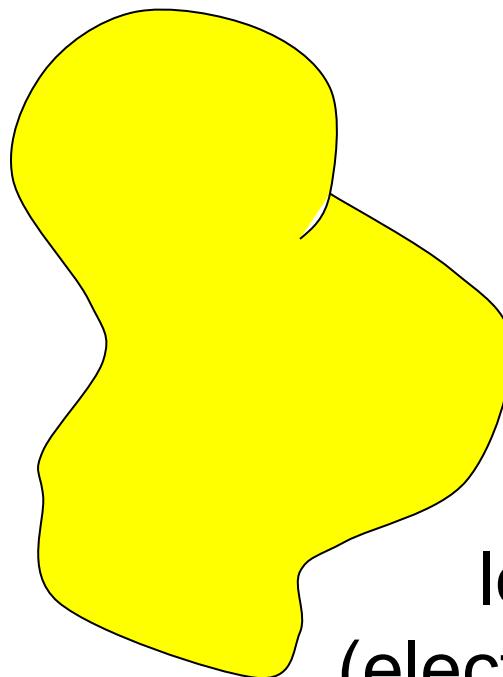
Homology modelling

Applications: fit to low res 3D

Query sequence 1



Query sequence 2



low resolution 3D
(electron microscopy)

Homology modelling GenTHREADER

David Jones http://bioinf.cs.ucl.ac.uk/psipred_new/
Input sequence or MSA



Choose Prediction Methods

PSIPRED v3.3 (Predict Secondary Structure)
 pGenTHREADER (Profile Based Fold Recognition)

 BioSerf v2.0 (Automated Homology Modelling)
 FFPred v2.0 (Eukaryotic Function Prediction)
 MEMPACK (SVM Prediction of TM Topology and Helix Packing)
 DomSerf v2.0 (Automated Domain Modelling by Homology)
[Help...](#)

DISOPRED3 & DISOPRED2 (Disorder Prediction)
 MEMSAT3 & MEMSAT-SVM (Membrane Helix Prediction)

 DomPred (Protein Domain Prediction)
 GenTHREADER (Rapid Fold Recognition)
 pDomTHREADER (Fold Domain Recognition)

Input Sequence (Single sequence or Multiple Sequence alignments; as raw sequence or fasta format)

Typically 30 minutes, up to two hours
GenTHREADER Jones (1999) *J Mol Biol*

Homology modelling Phyre



Mike Sternberg <http://www.sbg.bio.ic.ac.uk/phyre2/>

Kelley et al (2000) *J Mol Biol*

Kelley et al (2015)

Nature Protocols

The screenshot shows the Phyre2 web interface. At the top center is the large "Phyre²" logo. Below it is the text "Protein Homology/analogy Recognition Engine V 2.0". To the right is a "Subscribe to Phyre at Google Groups" section with fields for "Email:" and a "Subscribe" button, along with a link "Visit Phyre at Google Groups". Below the logo are several icons: a bar chart, a magnifying glass, a question mark, an envelope, and a book. In the middle left is a dark grey sidebar with fields for "E-mail Address" and "Optional Job description". On the right is a large white input area for "Amino Acid Sequence". At the bottom is a control panel with "Modelling Mode" set to "Normal" (radio button selected), "Intensive" (radio button unselected), and buttons for "Phyre Search" and "Reset".

Processing time can be hours

Homology modelling

Static solutions

Datasets of precomputed models / computations

Not flexible

Variable coverage

But you don't have to wait



Homology modelling MODbase

Andrej Sali

<http://modbase.compbio.ucsf.edu/>

 **ModBase: Database of Comparative Protein Structure Models** 

• [Sali Lab Home](#) • [ModWeb](#) • [ModLoop](#) • [ModBase](#) • [ModEval](#) • [PCSS](#) • [FoXS](#) • [IMP](#) • [ModPipe](#) •

[ModBase Home](#) • [ModBase Datasets for User:Anonymous](#) • [User Login](#) • [Help](#) • [News](#) • [Contact](#) • [Current Datasets](#) •

• [General Information](#)
• [Statistics and Genome Datasets](#)
• [News](#)
• [Project Pages](#)
• [Authors and Acknowledgements](#)
• [Publications](#)
• [Related Resources](#)

Please address inquiries to:
modbase@salilab.org

MODBASE contains theoretically calculated models, not experimentally determined structures. The models may contain significant errors.

ModBase Search

ModBase is a database of comparative protein structure models, calculated by our modeling pipeline ModPipe.

Search type  Display type 

To include the academic (comprehensive) dataset, go to '[Current Datasets](#)'!

All available datasets are selected 

Search by properties
Property 
Organism  or

[Advanced search](#)

Homology modelling MODbase

Sequence Overview

[Go to Model Overview](#)

Search Summary

Search Input: database_id: sorcs3_human

Organism(s):

Homo sapiens

1 match found.

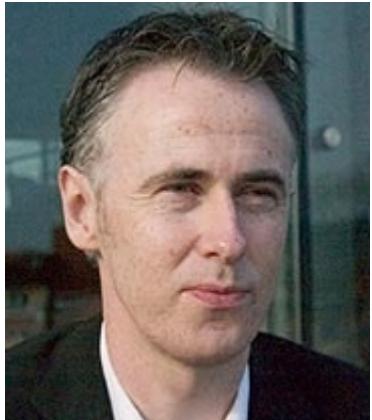
Perform Action on Selected Model(s) Check model(s), then select option

Covered For	TARGET						MODEL DATA				TEMPLATE		
	Model Icon	Model/Fold Reliability	Sequence Database Link	Database Annotation	Organism	Protein Size	Modeled Segment	Size	Seq Id(%)	PDB code	PDB Segment	PDB Comment	
			<input type="checkbox"/>	Q5VXF9	vps10 domain receptor protein sorcs 3 (sorcs3)	Homo sapiens	1222	198-643	446	16.00	1sqjA	8-581	crystal structure analysis of oligoxyloglucan reducing-end-specific cellobiohydrolase (oxg-rcbh)
			<input type="checkbox"/>	Q5VXF9	vps10 domain receptor protein sorcs 3 (sorcs3)	Homo sapiens	1222	798-915	118	35.00	1wqoA	5-122	solution structure of the pkd domain from human vps10 domain-containing receptor sorcs2
			<input type="checkbox"/>	Q5VXF9	vps10 domain receptor protein sorcs 3 (sorcs3)	Homo sapiens	1222	198-712	515	12.00	1sqjA	8-730	crystal structure analysis of oligoxyloglucan reducing-end-specific cellobiohydrolase (oxg-rcbh)

Aquaria

Sean O'Donoghue

<http://aquaria.ws/>



AQUARIA Updated: 1 Jan 2015

About | Report an Issue | Help

SPECIFY A PROTEIN

Protein name or ID
Synonyms: TET1_HUMAN, Methylcytosine dioxygenase TET1, CXXC-type... [+]
Gene: TET1
Organism
Homo sapiens, Human

ABOUT TET1_HUMAN

FUNCTION: Dioxygenase that catalyzes the conversion of the modified genomic... [+]
CATALYTIC ACTIVITY:

- DNA 5-methylcytosine + 2-oxoglutarate + O(2) = DNA 5-

[+]
COFACTOR:

- Binds 1 Fe(2+) ion per subunit.
- Binds 3 zinc ions per subunit. The ... [+]

SUBUNIT: Interacts with HCFC1 and more details (5)

3D STRUCTURE TET1_HUMAN sequence aligned onto TET2 structure from PDB 4nm6-A (64% sequence identity) ?

Color Representation View

Crystal structure of TET2-DNA complex: insight into TET-mediated 5mC oxidation.

Hu et al., Cell (2013)

Abstract: TET proteins oxidize 5-methylcytosine (5mC) on DNA and play important roles in various biological processes. Mutations of TET2 are frequently observed in myeloid malignancy. Here, we present the crystal structure of human TET2 bound to methylated DNA at 2.02 Å resolution. The structure shows that two zinc fingers bring the Cys-rich and DSBH domains together to form a compact catalytic domain. The Cys-rich domain stabilizes the DNA above the DSBH core. TET2 specifically recognizes CpG dinucleotide and shows substrate preference for 5mC in a CpG context. 5mC is inserted into the... [+]

Determined by: X-ray diffraction at 2.03 Å resolution

Chain A: TET2 (Methylcytosine dioxygenase TET2)

SELECTION A: L(1340)

MATCHING STRUCTURES in PDB 15 FEATURES for TET1_HUMAN 3D

200 400 600 800 2136

65% 64% 34% 26%

1 2 3

The figure displays the Aquaria interface for the protein TET1_HUMAN. On the left, there's a sidebar for specifying a protein with fields for name, ID, gene, organism, and a detailed 'About' section. The main area shows a 3D ribbon model of the TET2 protein bound to DNA, with specific residues highlighted in yellow. A selection box is overlaid on the protein structure. Below the 3D view, a 'Matching Structures' panel lists three other PDB entries (1, 2, 3) with their sequence identities (65%, 64%, 34%) and feature counts (2136). The right side contains a detailed abstract from a paper by Hu et al. in Cell (2013) about the crystal structure of TET2 and its role in 5mC oxidation.

O'Donoghue et al (2015) *Nature Methods*

Homology modelling Protein Model Portal



Torsten Schwede

PSI | The Protein Model Portal

Home Interactive Modeling Quality Estimation Protein Modeling 101 More ▾

Welcome to the

Protein Model Portal (PMP)

PMP gives access to various models computed by comparative modeling methods provided by different partner sites, and provides access to various interactive services for model building, and quality assessment.

Please enter your query.

Examples: [UniProt AC] [UniProt ID] [RefSeq] [PDBID] [Sequence] [Free Text]

Haas et al. (2013) *Database*

Domains

Protein domains are structural units (average 160 aa) that share:

Function
Folding
Evolution

Proteins normally are multidomain (average 300 aa)

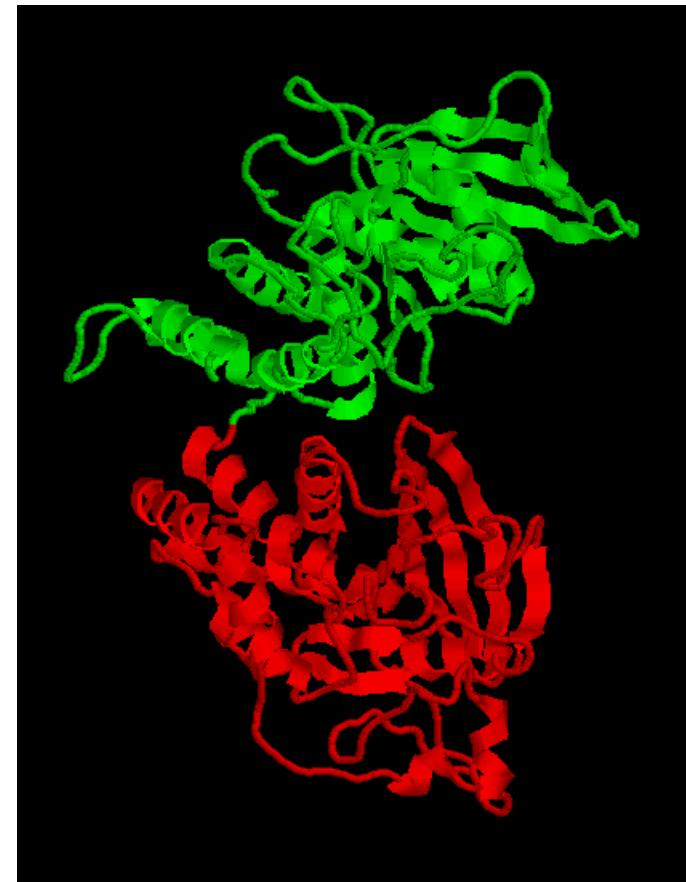


Domains

Protein domains are structural units (average 160 aa) that share:

Function
Folding
Evolution

Proteins normally are multidomain (average 300 aa)

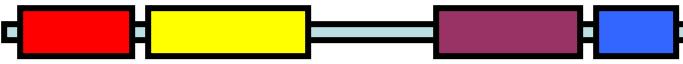


Domains

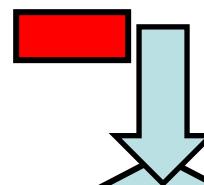
Query Sequence



Predict domains



Cut



Similar to PDB
sequence?

No

2D Prediction
3D Ab initio
3D Threading

Yes

3D Modeling by homology

3D structure prediction Ab initio

Explore conformational space

Limit the number of atoms

Break the problem into fragments of sequence

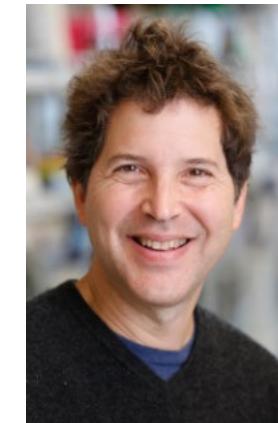
Optimize hydrophobic residue burial and pairing of beta-strands

Limited success... Until last week!

3D structure prediction Ab initio

trRosetta: David Baker

Needs large multiple sequence alignments to predict contacts



Predictions available since last week for all PFAM domains
Check: <https://www.ebi.ac.uk/interpro/entry/pfam/PF07887/model/>

2d best

AlphaFold2: Google's DeepMind /
Apparently Better... but not open

Yang et al (2020) *PNAS*

<https://yanglab.nankai.edu.cn/trRosetta/>

3D structure prediction Threading

I-Tasser: Yang Zhang &
Jeffrey Skolnick



Fold 66% sequences <200 aa long of low homology to PDB

Just submit your sequence and wait... (some days)

Output are predicted structures (PDB format)

Lee and Skolnick (2008) *Biophysical Journal*

Roy et al (2010) *Nature Methods*

Yang et al (2015) *Nature Methods*

3D structure prediction I-Tasser

