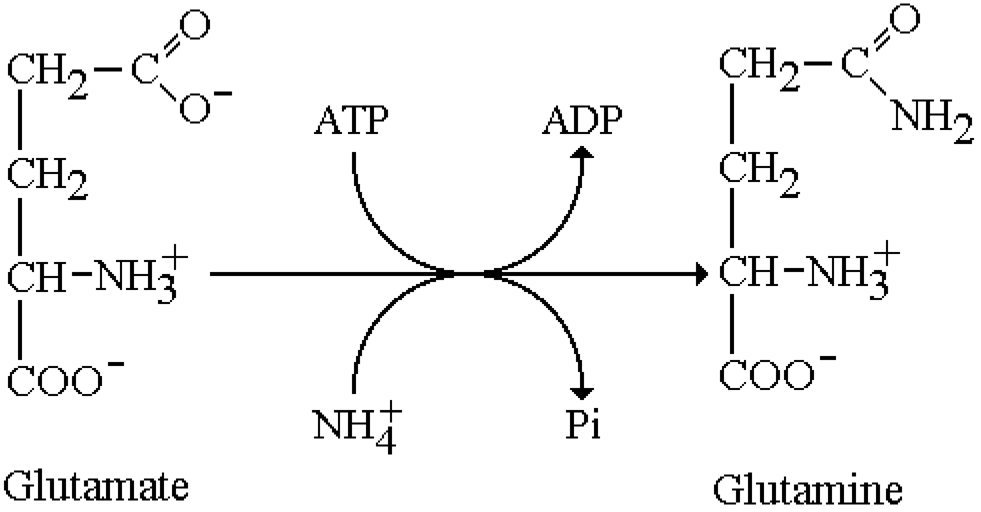
Protein structure prediction

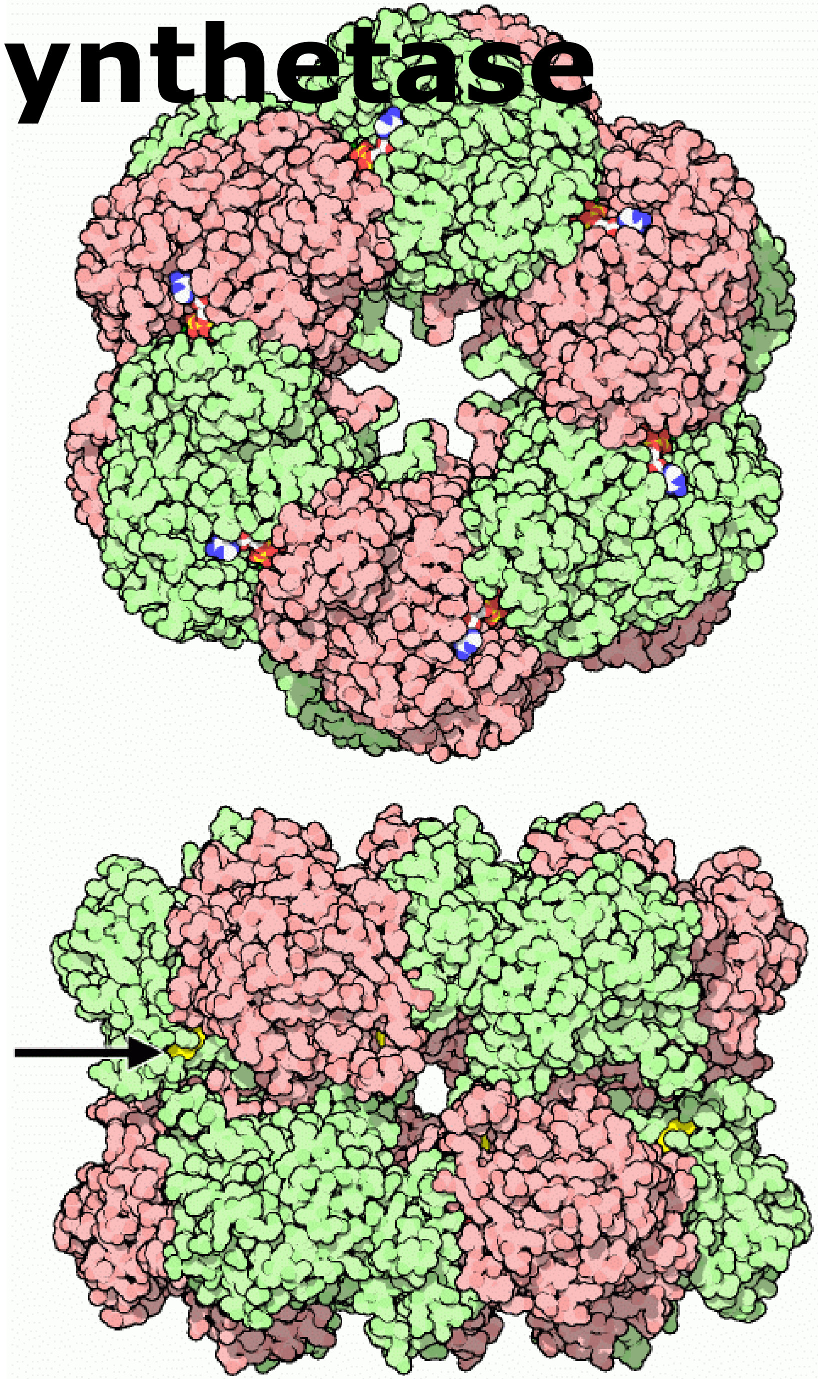
Miguel Andrade
Faculty of Biology,
Institute of Organismic and Molecular Evolution
Johannes Gutenberg University
Mainz, Germany
andrade@uni-mainz.de

Mount Everest

Age: 60M years

Glutamine synthetase

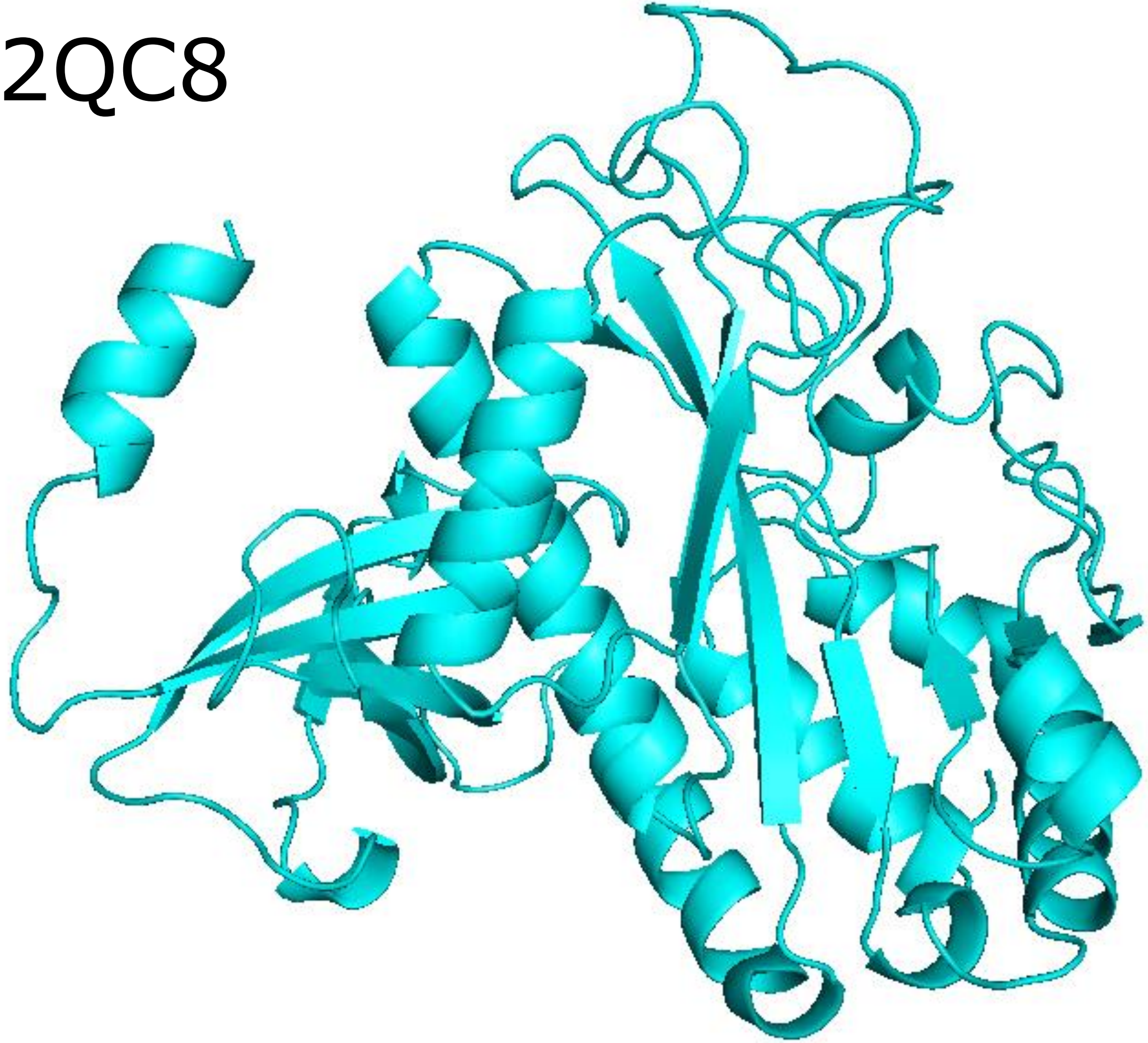




Age: +3500M years

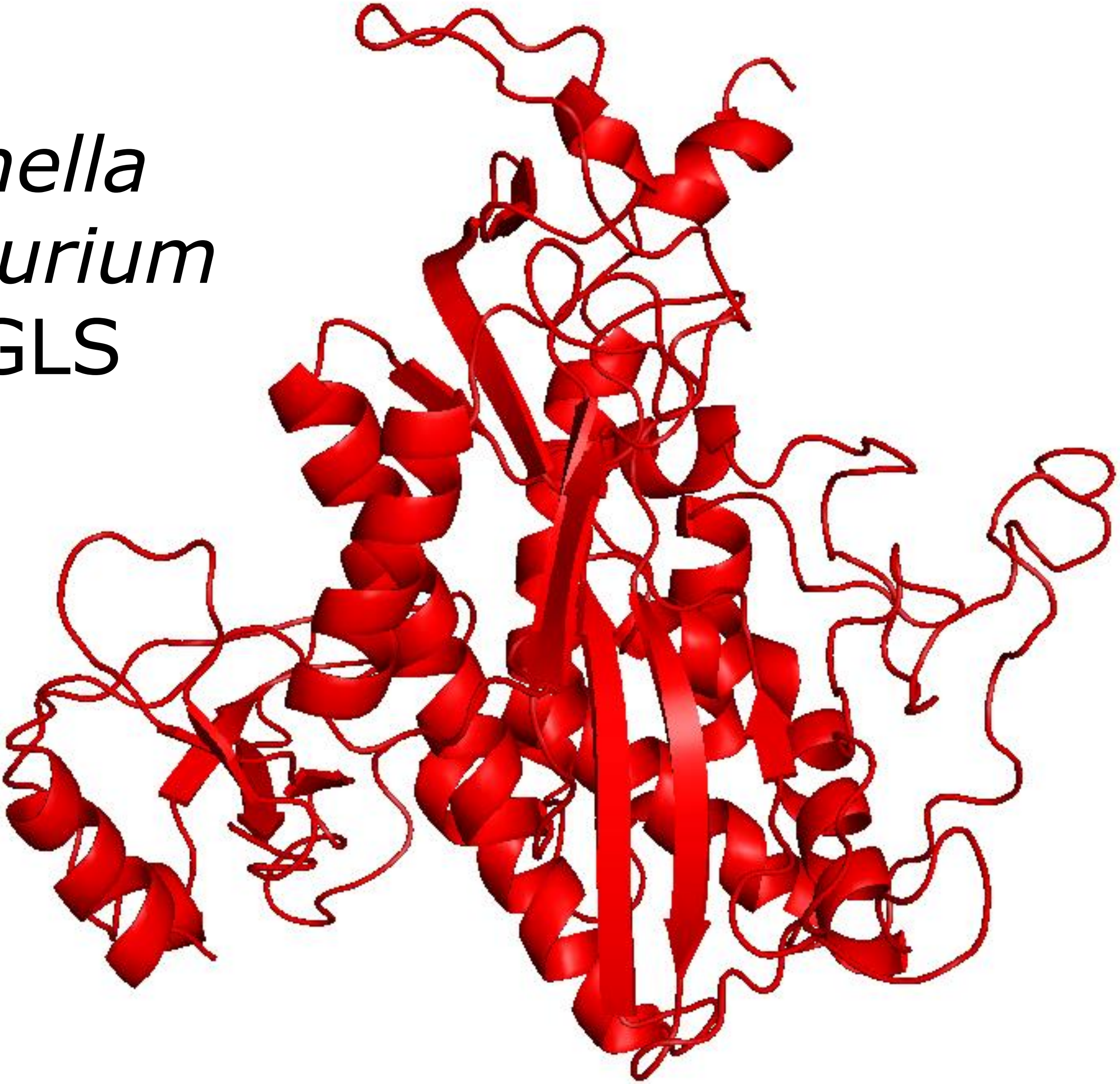
Glutamine synthetase

Human
PDB:2QC8

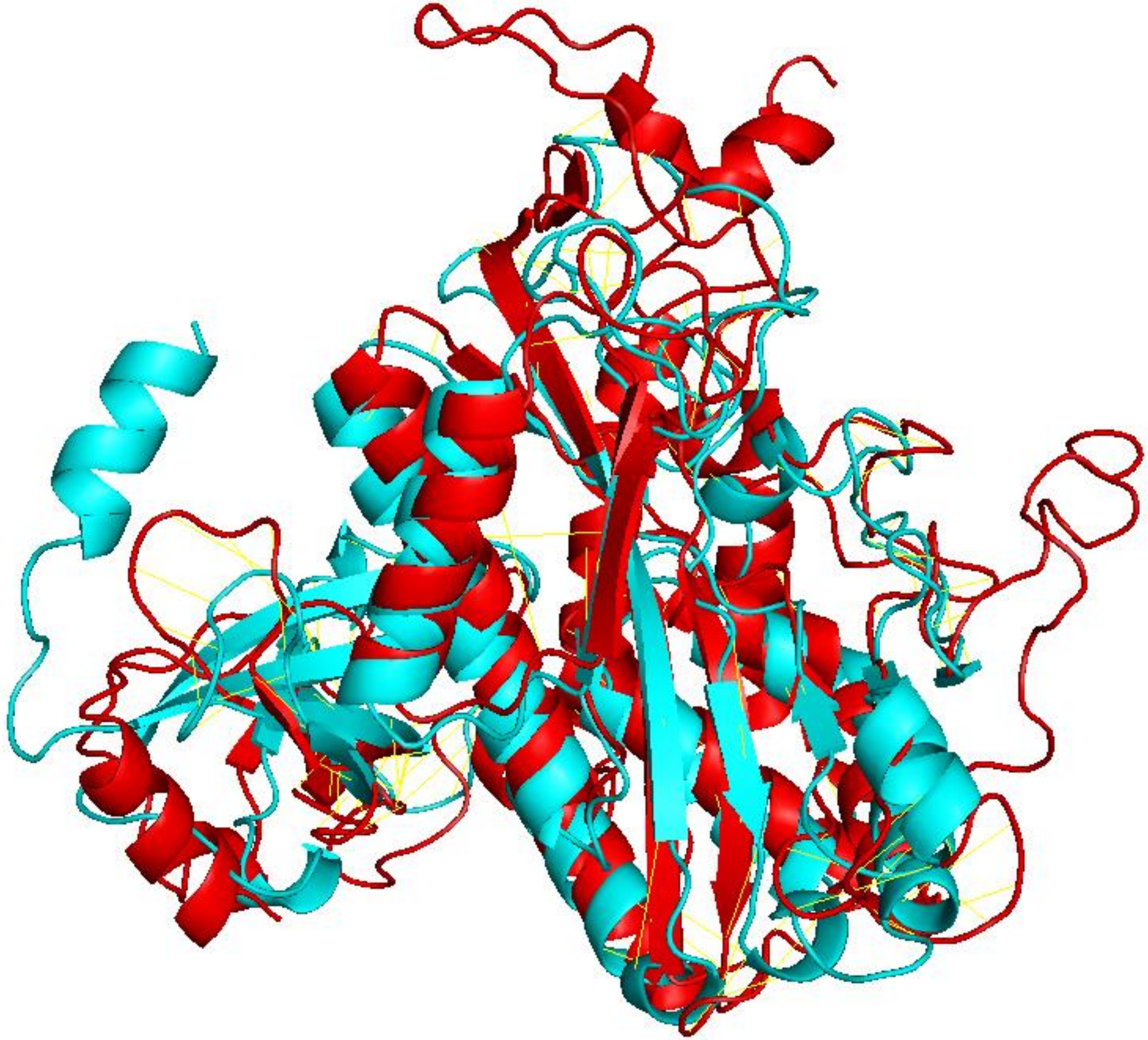


Glutamine synthetase

Salmonella typhimurium
PDB:2GLS



Glutamine synthetase



Time line

Earth: 4.6 By

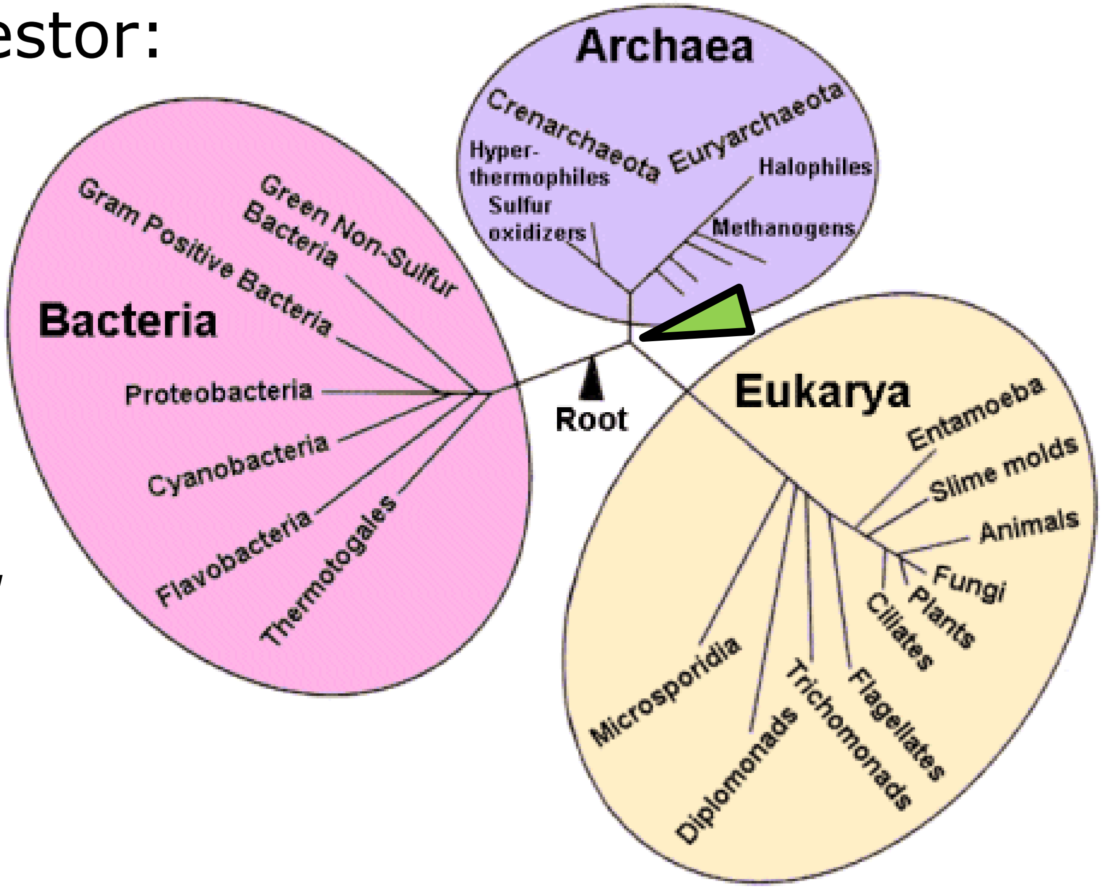
Origin of life: 3.9 By – 3.5 By

Last Common Ancestor:
3.5 – 3.8 By

Glansdorff & Labedan
(2008) *Biology Direct*

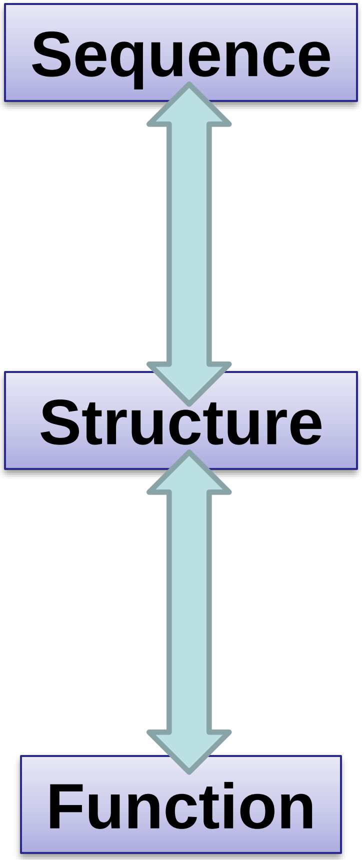
4.29 By

Sheridan *et al.* (2003)
Geomicrobiology Journal



Sequence and function

Evolutionary constraints



MTQDELKKAVGWAALQYVQ

PG DA

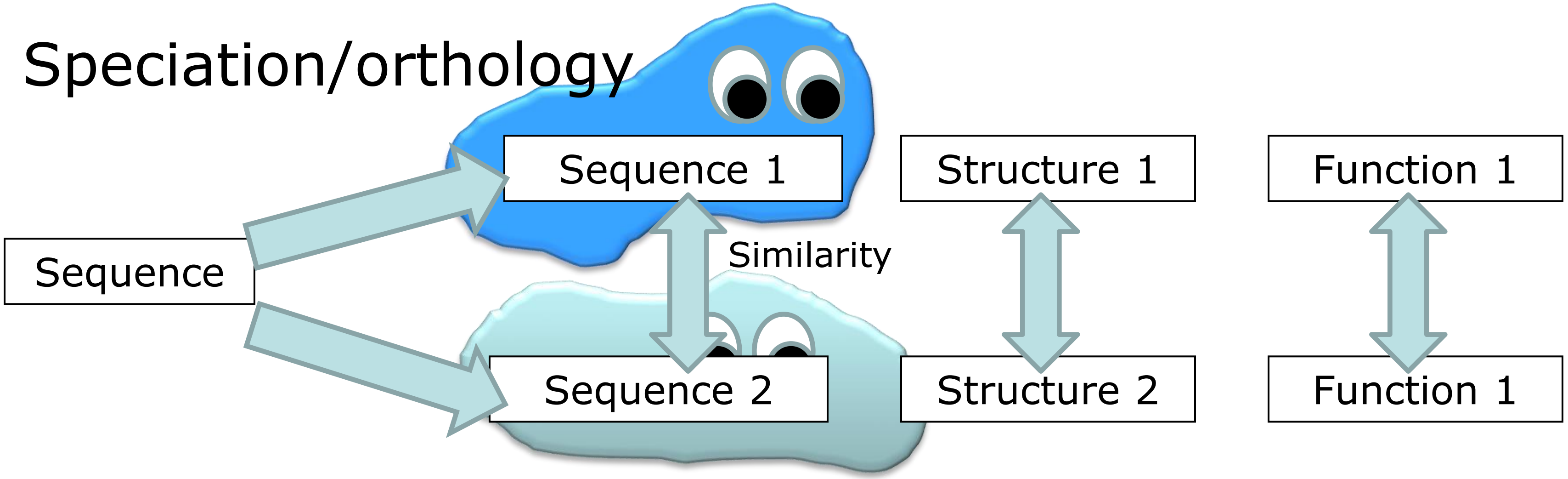
LG ST

EK

Sequence and function

Evolutionary constraints

Speciation/orthology



Sequence

Sequence 1

Sequence 2

Similarity

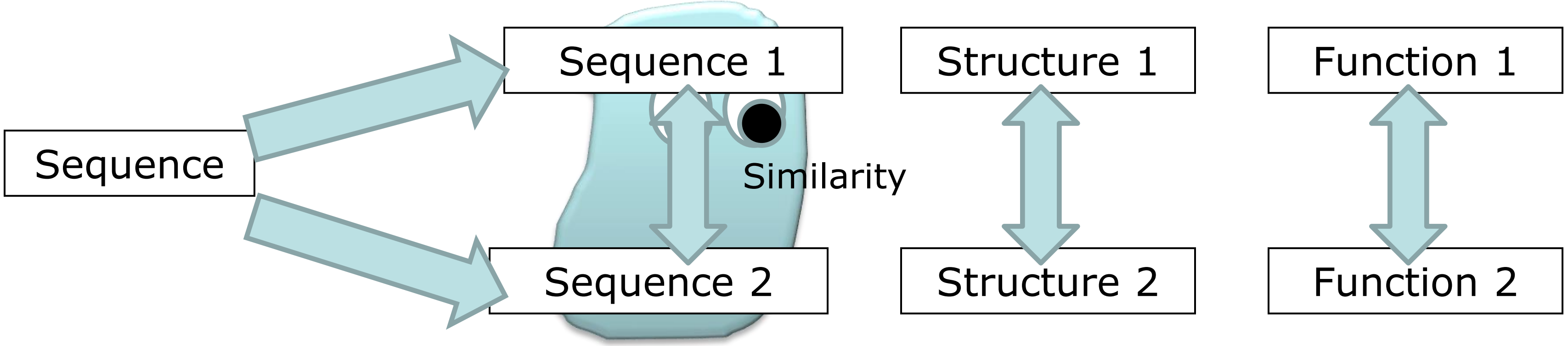
Structure 1

Structure 2

Function 1

Function 1

Gene duplication/paralogy



Sequence

Sequence 1

Sequence 2

Similarity

Structure 1

Structure 2

Function 1

Function 2

Sequence pairwise alignment

```
>gs_human gi|74271837|ref|NP_001028216.1| glutamine synthetase [Homo sapiens]
MTTSASSHLNKGIKQVYMSLPQGEKVQAMYIWIDGTGEGLRCKTRTLDSEPKCVEELPEWNFDGSSTLQS
EGSNSDMYLVPAAMFRDPFRKDPNKLVLCEVFKYNRRPAETNLRHTCKRIMDMVSNQHPWFGMEQEYTLM
GTDGHPFGWPSNGFPGPQGPYYCGVGADRAYGRDIVEAHYRACLYAGVKIAGTNAEVMPAQWEFQIGPCE
GISMGDHLWVARFILHRVCEDFGVIATFDPKPIPGNWNGAGCHTNFSTKAMREENGLKYIEEAIEKLSKR
HQYHIRAYDPKGGLDNARRLTGFHETSNINDFSAGVANRSASIRIPRTVGQEKKGYFEDRRPSANCDPFS
VTEALIRTCLLNETGDEPFQYKN
```

```
>gs_salmonella gi|16767272|ref|NP_462887.1| glutamine synthetase [Salmonella
enterica subsp. enterica serovar Typhimurium str. LT2]
MSAEHVLTMLNEHEVKFVDLRFTDTKGKEQHVTIPAHQVNAEFFEEGKMFDGSSIGGWKGINESDMVLMP
DASTAVIDPFFADSTLIIRCDILEPGTLQGYDRDPRSIAKRAEDYLRATGIADTVLFGPEPEFFLFDDIR
FGASISGSHVAIDDIEGAWNSSTKYEGGNKGHRPGVKGGYFPVPPVDSAQDIRSEMCLVMEQMGLVVEAH
HHEVATAGQNEVATRFNTMTKKADEIQIYKYVVHNVAHRFGKTATFMPKPMFGDNGSGMHCHMSLAKNGT
NLFSGDKYAGLSEQALYYIGGVIKHAKAINALANPTTNSYKRLVPGYEAPVMLAYSARNRSASIRIPVVA
SPKARRIEVRFPDPAANPYLCFAALLMAGLDGIKNKIHPGEAMDKNLYDLPPEEAKEIPQVAGSLEEALN
ALDLDREFLKAGGVFTDEAIDAYIALRREEDDRVRMTPHPVEFELYYSV
```

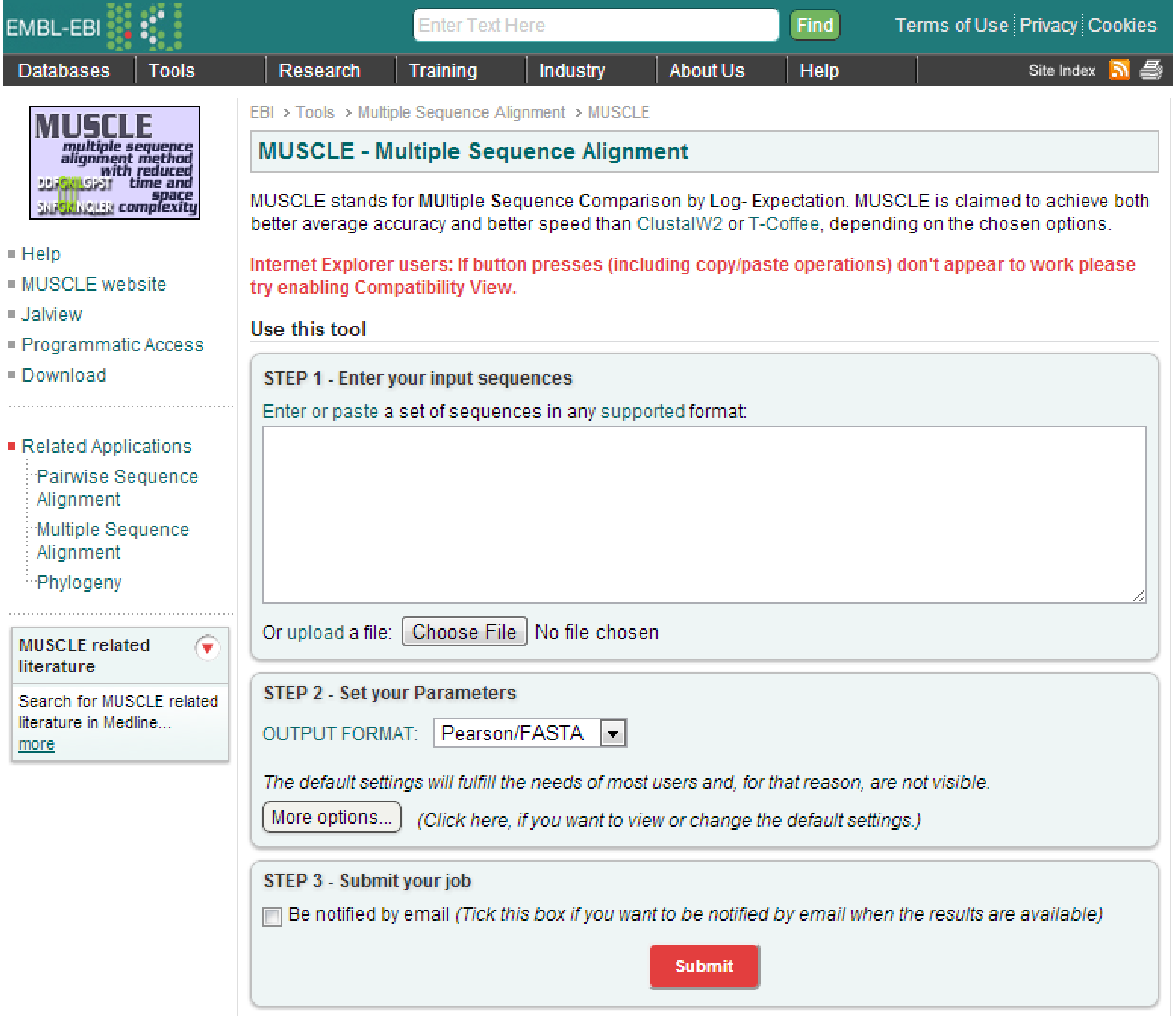
Sequence pairwise alignment

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-SDMYLVPAA--MFRDPFRKDPNKLVLCEVFK------YNRRP----  108
            FDGSS    +G N SDM L+P A     DPF  D   ++ C++ +      Y+R P
Sbjct  50   FDGSSIGGWKGINESDMVLMPDASTAVIDPFFADSTLIIRCDILEPGTLQGYDRDPRSIA  109

Query  109  --AETNLRHTCKRIMDMVSNQHPWFGMEQEYTLMGTDGHPFGWPSNGF------------  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--QDIRSEMCLVMEQMGLVVEAHHHEVATAG  218

Query  188  VKIAGTNAEVMPAQWEFQIGPCEGISMGDHLWVARFILHRVCEDFGVIATFDPKPIPG-N  246
                 T    M  +              D + + ++++H V   FG  ATF PKP+ G N
Sbjct  219  QNEVATRFNTMTKK-------------ADEIQIYKYVVHNVAHRFGKTATFMPKPMFGDN  265

Query  247  WNGAGCHTNFSTKAMREENGLKYIEEAIEKLSKRHQYHIRAYDPKGGLDNA---------  297
             +G  CH + +       +G KY       LS++  Y+I          NA
Sbjct  266  GSGMHCHMSLAKNGTNLFSGDKY-----AGLSEQALYYIGGVIKHAKAINALANPTTNSY  320

Query  298  RRLTGFHETSNINDFSAGVANRSASIRIPRTVGQEKKGYFEDRRPSANCDPFSVTEALIR  357
            +RL   +E   +  +SA   NRSASIRIP  V   K    E R P    +P+    AL+
Sbjct  321  KRLVPGYEAPVMLAYSA--RNRSASIRIP-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]
MTTSASSHLNKGIKQVYMSLPQGEKVQAMYIWIDGTGEGLRCKTRTLDSEPKCVEELPEWNFDGSSTLQS
EGSNSDMYLVPAAMFRDPFRKDPNKLVLCEVFKYNRRPAETNLRHTCKRIMDMVSNQHPWFGMEQEYTLM
GTDGHPFGWPSNGFPGPQGPYYCGVGADRAYGRDIVEAHYRACLYAGVKIAGTNAEVMPAQWEFQIGPCE
GISMGDHLWVARFILHRVCEDFGVIATFDPKPIPGNWNGAGCHTNFSTKAMREENGLKYIEEAIEKLSKR
HQYHIRAYDPKGGLDNARRLTGFHETSNINDFSAGVANRSASIRIPRTVGQEKKGYFEDRRPSANCDPFS
VTEALIRTCLLNETGDEPFQYKN
>gs_vulca gi|307594850|ref|YP_003901167.1| glutamine synthetase [Vulcanisaeta
distributa DSM 14429]
MPTRNLEIEPADLWRILKASGIKYVKFIIVDINGAPRSEIVPIDMAKDLFIDGMPFDASSIPSYSTVNKS
DFVAYVDPRAVYVEYWQDGKVADVFTMVSDIADKPSPLDPRRVLNDALEQARSKGYEFLMGVEVEFFVIK
EDGGKPVFADPGIYFDGWNVTVQSQFMKELITAIADAGINYTKTHHEVAPSQYEVNIGATDPLRLADQIV
YFKIMAKDIARKYGLVATFMPKPFWGVNGSGAHTHISVWKDGKNLFQSSTGKITEECGYAISAILSNARA
LSSFVAPLVNSYKRLVPHYEAPTRIVWGYANRSAMIRIPQYKMRINRIEYRHPDPSMNPYLAFTAIIKTM
IRGLEEKKEPPPPTEEVAYELANALETPATLEDTLKELSKSFLATELPSELVNAYIKIKQNEWEDYLTNV
GPWEKTWNIITQWEYNKYLVTA
>gs_salmonella gi|16767272|ref|NP_462887.1| glutamine synthetase [Salmonella
enterica subsp. enterica serovar Typhimurium str. LT2]
MSAEHVLTMLNEHEVKFVDLRFTDTKGKEQHVTIPAHQVNAEFFEEGKMFDGSSIGGWKGINESDMVLMP
DASTAVIDPFFADSTLIIRCDILEPGTLQGYDRDPRSIAKRAEDYLRATGIADTVLFGPEPEFFLFDDIR
FGASISGSHVAIDDIEGAWNSSTKYEGGNKGHRPGVKGGYFPVPPVDSAQDIRSEMCLVMEQMGLVVEAH
HHEVATAGQNEVATRFNTMTKKADEIQIYKYVVHNVAHRFGKTATFMPKPMFGDNGSGMHCHMSLAKNGT
NLFSGDKYAGLSEQALYYIGGVIKHAKAINALANPTTNSYKRLVPGYEAPVMLAYSARNRSASIRIPVVA
SPKARRIEVRFPDPAANPYLCFAALLMAGLDGIKNKIHPGEAMDKNLYDLPPEEAKEIPQVAGSLEEALN
ALDLDREFLKAGGVFTDEAIDAYIALRREEDDRVRMTPHPVEFELYYSV
>gs_yeast gi|330443748|ref|NP_015360.2| Gln1p [Saccharomyces cerevisiae S288c]
MAEASIEKTQILQKYLELDQRGRIIAEYVWIDGTGNLRSKGRTLKKRITSIDQLPEWNFDGSSTNQAPGH
DSDIYLKPVAYYPDPFRRGDNIVVLAACYNNDGTPNKFNHRHEAAKLFAAHKDEEIWFGLEQEYTLFDMY
DDVYGWPKGGYPAPQGPYYCGVGAGKVYARDMIEAHYRACLYAGLEISGINAEVMPSQWEFQVGPCTGID
MGDQLWMARYFLHRVAEEFGIKISFHPKPLKGDWNGAGCHTNVSTKEMRQPGGMKYIEQAIEKLSKRHAE
HIKLYGSDNDMRLTGRHETASMTAFSSGVANRGSSIRIPRSVAKEGYGYFEDRRPASNIDPYLVTGIMCE
TVCGAIDNADMTKEFERESS
```


EMBL-EBI
Enter Text Here
Find
Terms of Use | Privacy | Cookies
Databases
Tools
Research
Training
Industry
About Us
Help
Site Index
MUSCLE multiple sequence alignment method with reduced time and space complexity

- Help
- MUSCLE website
- Jalview
- Programmatic Access
- Download

- Related Applications
  - Pairwise Sequence Alignment
  - Multiple Sequence Alignment
  - Phylogeny

MUSCLE related literature

Search for MUSCLE related literature in Medline... more

EBI > Tools > Multiple Sequence Alignment > MUSCLE

MUSCLE - Multiple Sequence Alignment

MUSCLE stands for **MU**ltiple **S**equence **C**omparison by **L**og- **E**xpectation. 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: Choose File No file chosen

STEP 2 - Set your Parameters

OUTPUT FORMAT: Pearson/FASTA

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

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

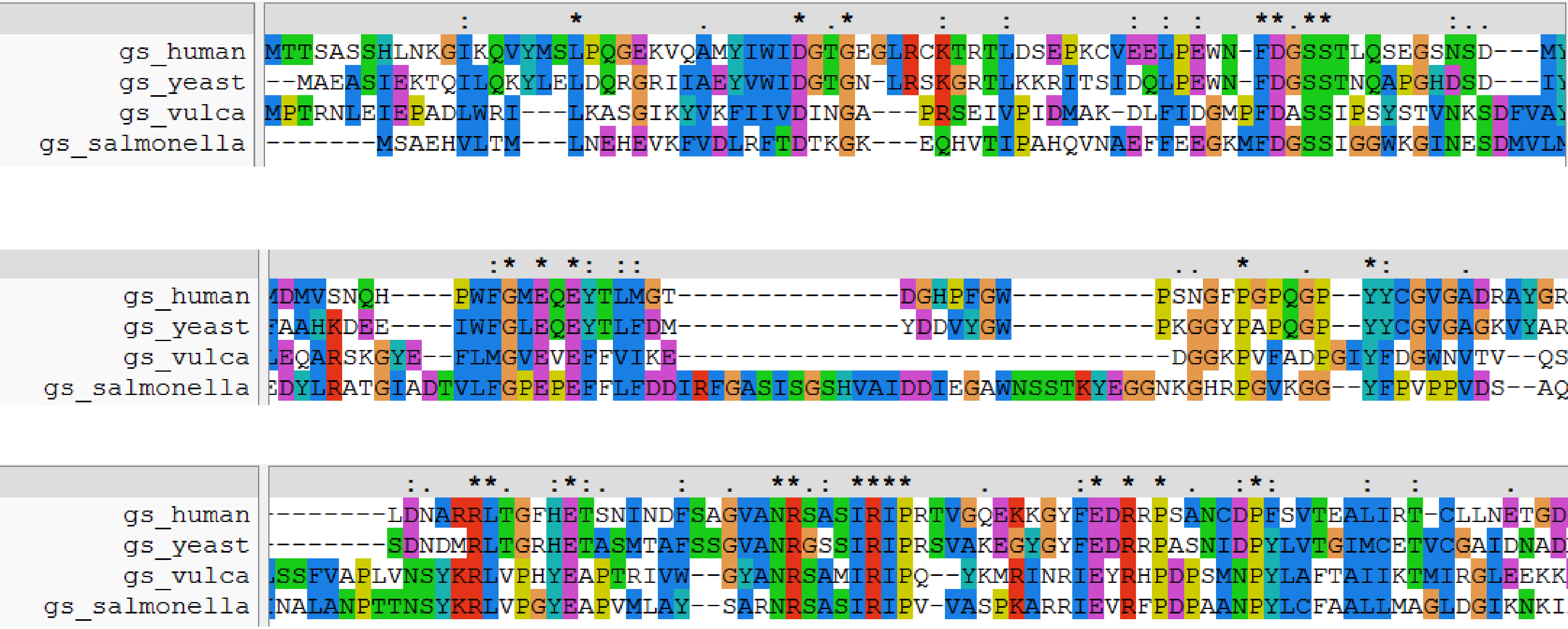
STEP 3 - Submit your job

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

Submit

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

```
>gs_human gi|74271837|ref|NP_001028216.1| glutamine synthetase [Homo sapiens]
MTTSASSHLNKGIKQVYMSLPQGEKVQAMYIWIDGTGEGLRCKTRTLDSEPKCVEELPEW
N-FDGSSTLQSEGSNSD---MYLVPAAMFRDPFRKDPNKLVLCEVFKYNRRPA-ETNLRH
TCKRIMDMVSNQH----PWFGMEQEYTLMGT--------------DGHPFGW--------
-PSNGFPGPQGP--YYCGVGADRAYGRDIVEAHYRACLYAGVKIAGTNAEVMPA-QWEFQ
IGPCEGISMGDHLWVARFILHRVCEDFGVIATFDPKPIPGNWNGAGCHTNFSTKAMREEN
GLKYIEEAIEKLSKRHQYHIRAYDPKGG---------LDNARRLTGFHETSNINDFSAGV
ANRSASIRIPRTVGQEKKGYFEDRRPSANCDPFSVTEALIRT-CLLNETGDEP-------
------------------------------------------------------------
--------------FQYKN-----

>gs_yeast gi|330443748|ref|NP_015360.2| Gln1p [Saccharomyces cerevisiae S288c]
--MAEASIEKTQILQKYLELDQRGRIIAEYVWIDGTGN-LRSKGRTLKKRITSIDQLPEW
N-FDGSSTNQAPGHDSD---IYLKPVAYYPDPFRRGDNIVVLAACYNNDGTPN-KFNHRH
EAAKLFAAHKDEE----IWFGLEQEYTLFDM--------------YDDVYGW--------
-PKGGYPAPQGP--YYCGVGAGKVYARDMIEAHYRACLYAGLEISGINAEVMPS-QWEFQ
VGPCTGIDMGDQLWMARYFLHRVAEEFGIKISFHPKPLKGDWNGAGCHTNVSTKEMRQPG
GMKYIEQAIEKLSKRHAEHIKLYG-------------SDNDMRLTGRHETASMTAFSSGV
ANRGSSIRIPRSVAKEGYGYFEDRRPASNIDPYLVTGIMCETVCGAIDNADMT-------
------------------------------------------------------------
--------------KEFERESS—

>gs_vulca gi|307594850|ref|YP_003901167.1| glutamine synthetase [Vulcanisaeta distributa DSM 14429]
MPTRNLEIEPADLWRI---LKASGIKYVKFIIVDINGA---PRSEIVPIDMAK-DLFIDG
MPFDASSIPSYSTVNKSDFVAYVDPRAVYVEYWQDGKVADVFTMVSDIADKPS-PLDPRR
VLNDALEQARSKGYE--FLMGVEVEFFVIKE-----------------------------
--DGGKPVFADPGIYFDGWNVTV--QSQFMKELITAIADAGINYTKTHHEVAPS-QYEVN
IGATDPLRLADQIVYFKIMAKDIARKYGLVATFMPKPFWGV-NGSGAHTHIS---VWKDG
KNLF-QSSTGKITEECGYAISAILSNARALSSFVAPLVNSYKRLVPHYEAPTRIVW--GY
ANRSAMIRIPQ--YKMRINRIEYRHPDPSMNPYLAFTAIIKTMIRGLEEKKEPPPPTEEV
AYELA--NALETP---ATLEDTLK--ELSKSFLATE--LPSELVNAYIKIKQNEWEDYLT
NVGPWEKTWNIITQWEYNKYLVTA

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



```
                           :      *       .     * .*     :   :       : : :   **.**       :..
     gs_human  MTTSASSHLNKGIKQVYMSLPQGEKVQAMYIWIDGTGEGLRCKTRTLDSEPKCVEELPEWN-FDGSSTLQSEGSNSD---MY
     gs_yeast  --MAEASIEKTQILQKYLELDQRGRIIAEYVWIDGTGN-LRSKGRTLKKRITSIDQLPEWN-FDGSSTNQAPGHDSD---IY
     gs_vulca  MPTRNLEIEPADLWRI---LKASGIKYVKFIIVDINGA---PRSEIVPIDMAK-DLFIDGMPFDASSIPSYSTVNKSDFVAY
gs_salmonella  -------MSAEHVLTM---LNEHEVKFVDLRFTDTKGK---EQHVTIPAHQVNAEFFEEGKMFDGSSIGGWKGINESDMVLM

                             :* * *: ::                                 ..  *   .   *:    .
     gs_human  MDMVSNQH----PWFGMEQEYTLMGT--------------DGHPFGW---------PSNGFPGPQGP--YYCGVGADRAYGR
     gs_yeast  FAAHKDEE----IWFGLEQEYTLFDM--------------YDDVYGW---------PKGGYPAPQGP--YYCGVGAGKVYAR
     gs_vulca  LEQARSKGYE--FLMGVEVEFFVIKE-------------------------------DGGKPVFADPGIYFDGWNVTV--QS
gs_salmonella  EDYLRATGIADTVLFGPEPEFFLFDDIRFGASISGSHVAIDDIEGAWNSSTKYEGGNKGHRPGVKGG--YFPVPPVDS--AQ

                        :.  **.  :*:.    :  .  **.: ****    .     :* * * .  :*:     :  :     .
     gs_human  --------LDNARRLTGFHETSNINDFSAGVANRSASIRIPRTVGQEKKGYFEDRRPSANCDPFSVTEALIRT-CLLNETGD
     gs_yeast  --------SDNDMRLTGRHETASMTAFSSGVANRGSSIRIPRSVAKEGYGYFEDRRPASNIDPYLVTGIMCETVCGAIDNAD
     gs_vulca  LSSFVAPLVNSYKRLVPHYEAPTRIVW--GYANRSAMIRIPQ--YKMRINRIEYRHPDPSMNPYLAFTAIIKTMIRGLEEKK
gs_salmonella  INALANPTTNSYKRLVPGYEAPVMLAY--SARNRSASIRIPV-VASPKARRIEVRFPDPAANPYLCFAALLMAGLDGIKNKI
```

ClustalW, JalView

Determination of protein structure

X-ray crystallography (144K in PDB)

- need crystals

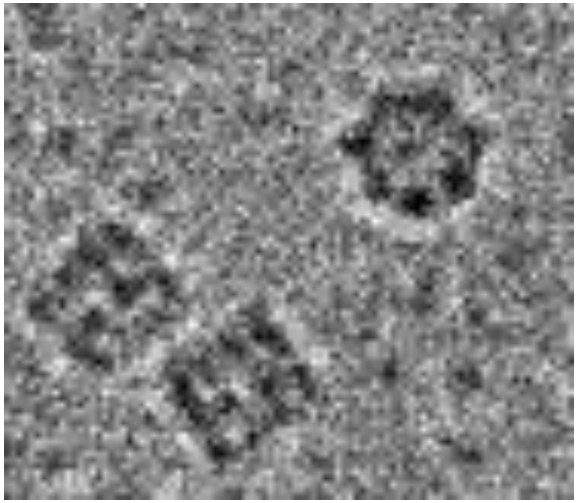
Nuclear Magnetic Resonance (NMR) (12K)

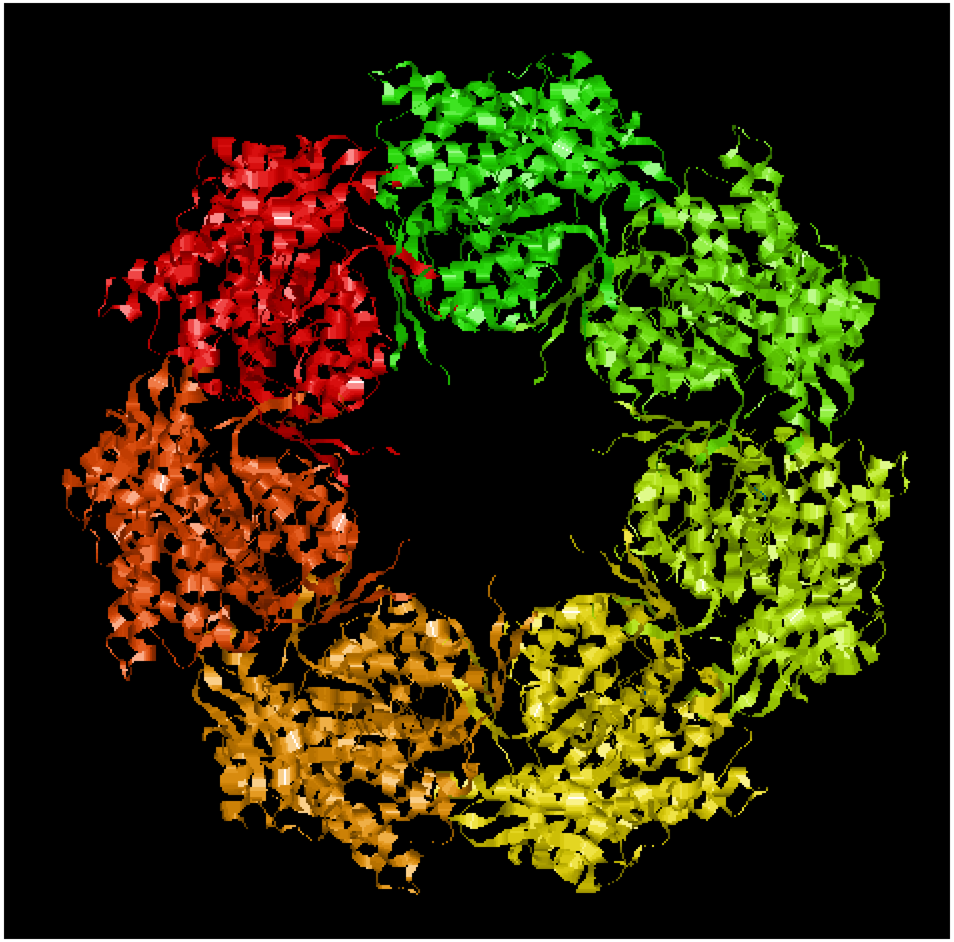
- proteins in solution
- lower size limit (600 aa)

Electron microscopy (7K)

- Low resolution (>5A)

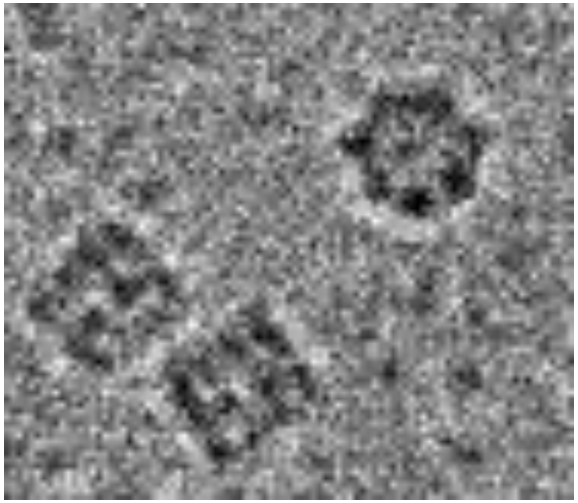
Determination of protein structure

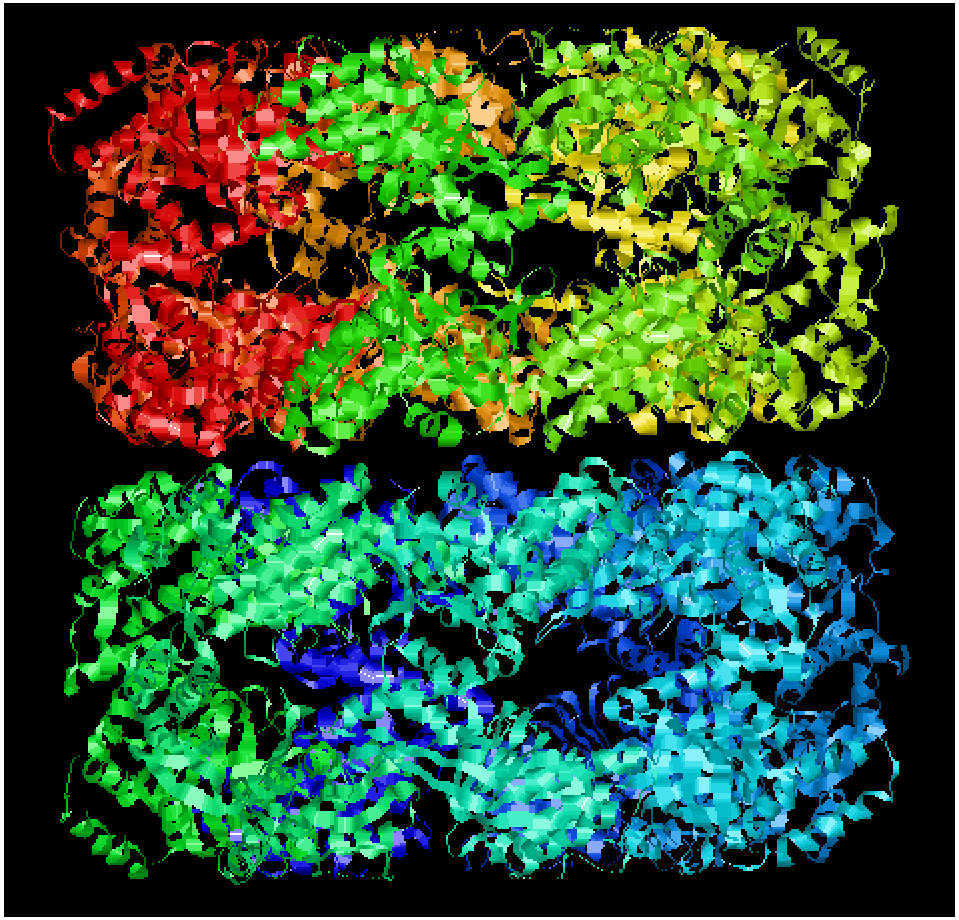




resolution 2.4 A

Determination of protein structure



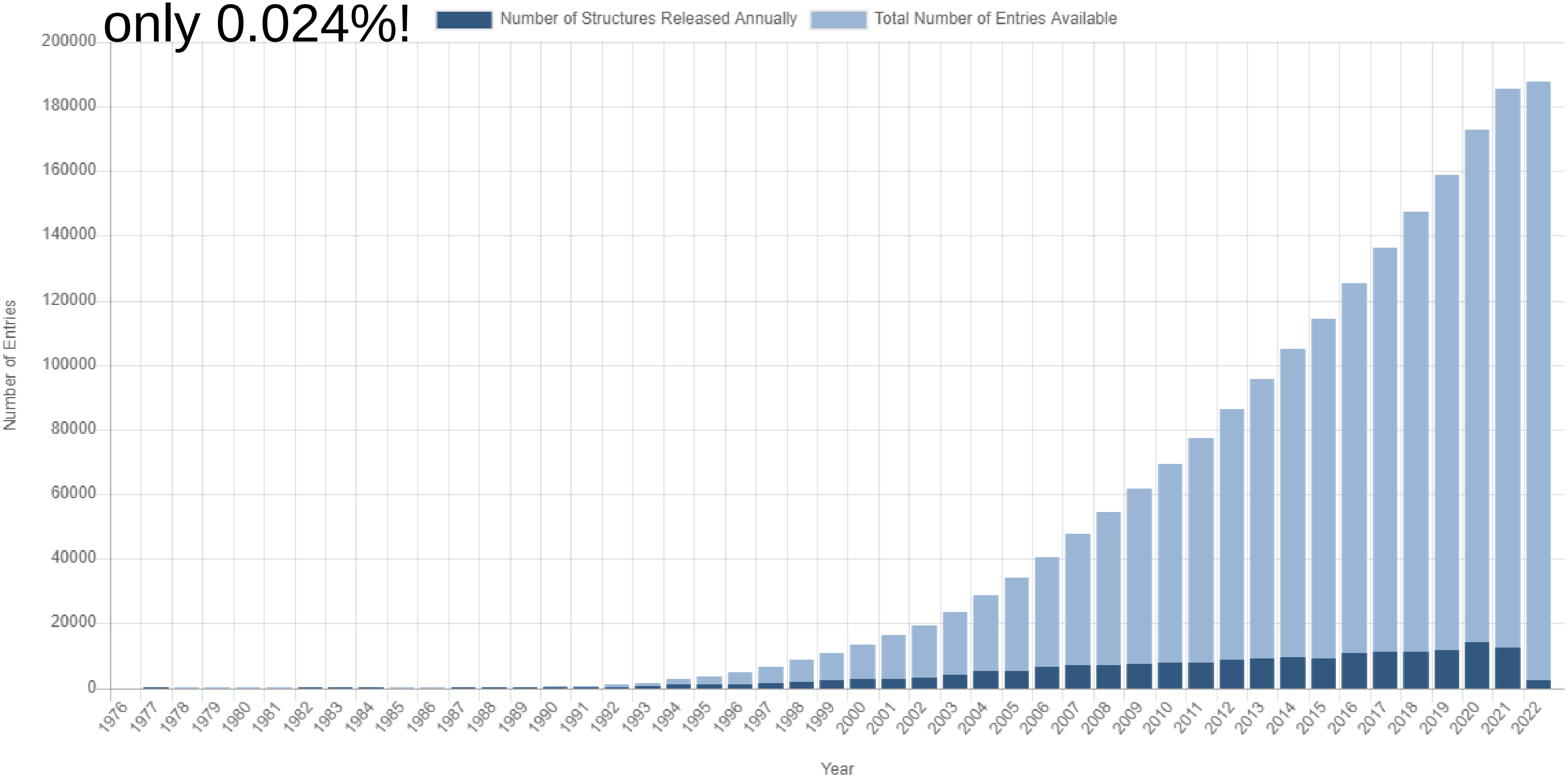


resolution 2.4 A

Structural genomics

Currently: 187K protein 3D structures
from around 55K sequences in UniProt (how do I know?)
226M sequences in UniProt

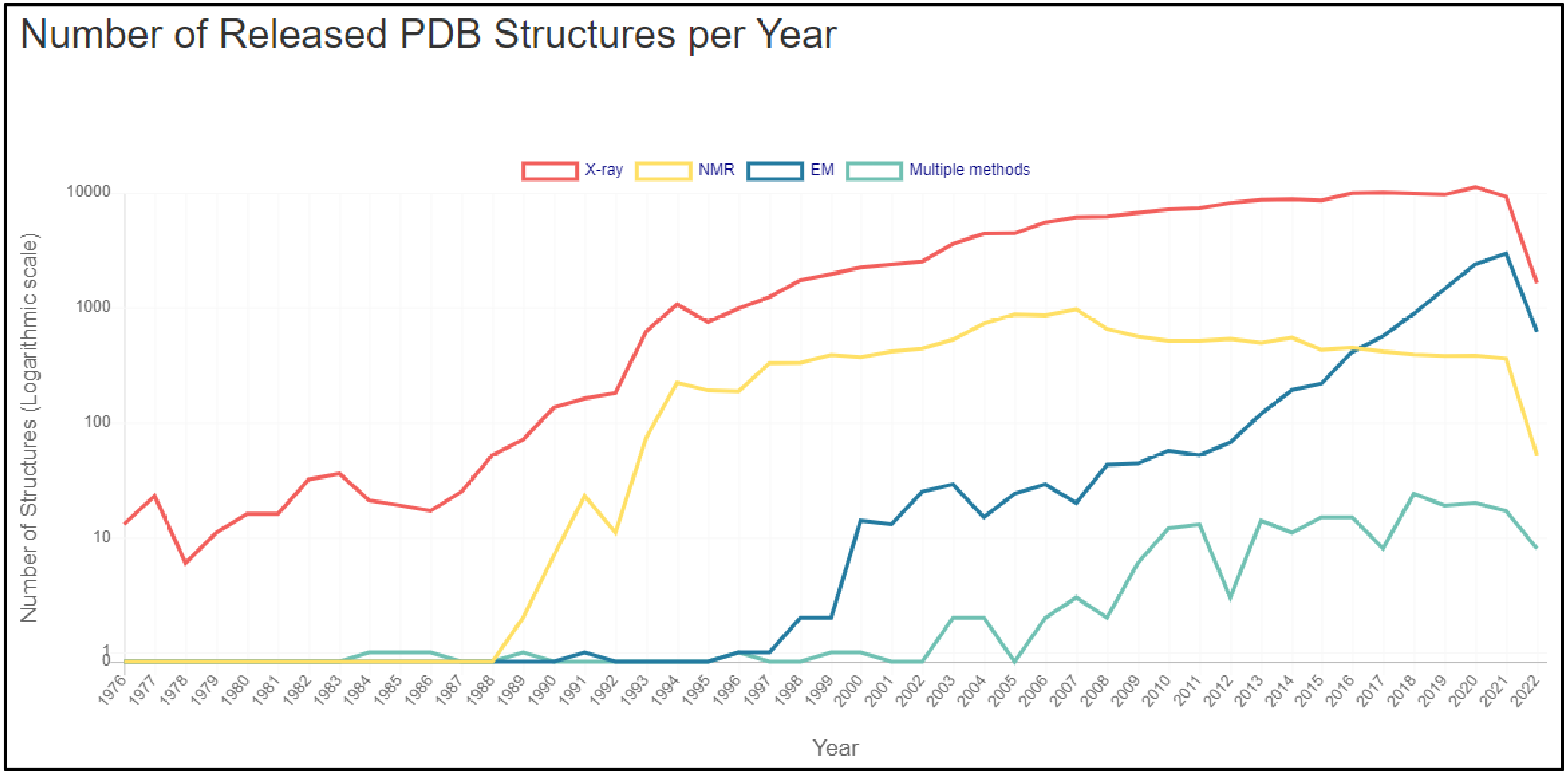
only 0.024%!



Structural genomics

Currently: 187K protein 3D structures
from around 55K sequences in UniProt (how do I know?)
226M sequences in UniProt

only 0.024%!

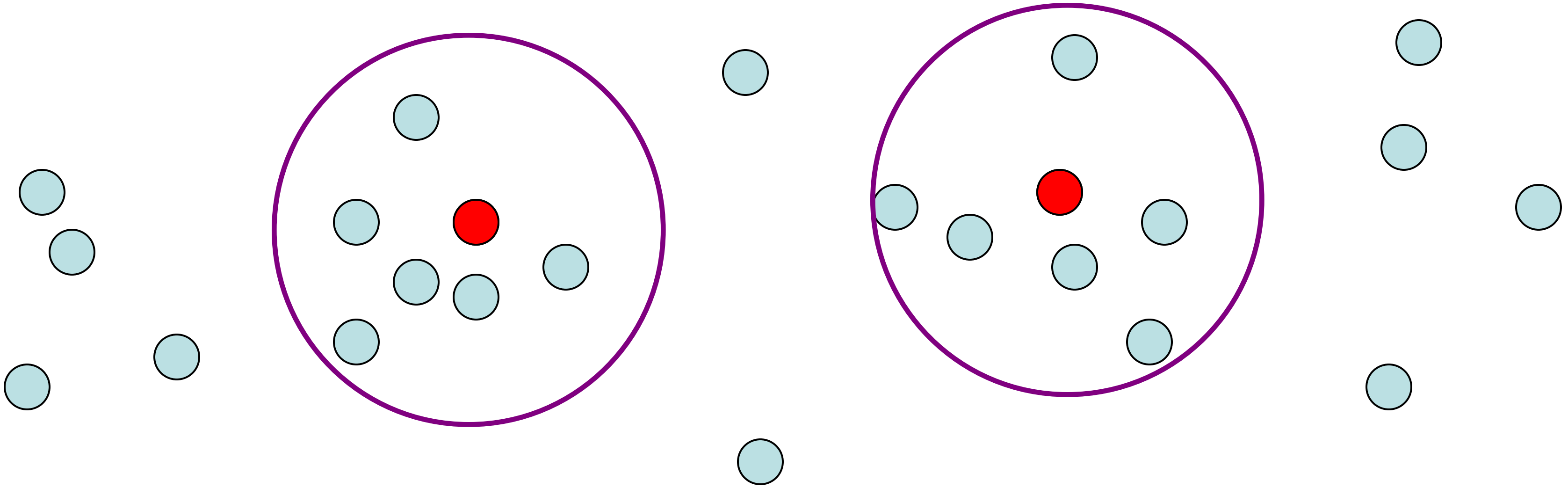


Number of Released PDB Structures per Year

Structural genomics

Currently: 187K protein 3D structures
from around 55K sequences in UniProt (how do I know?)
226M sequences in UniProt

only 0.024%!

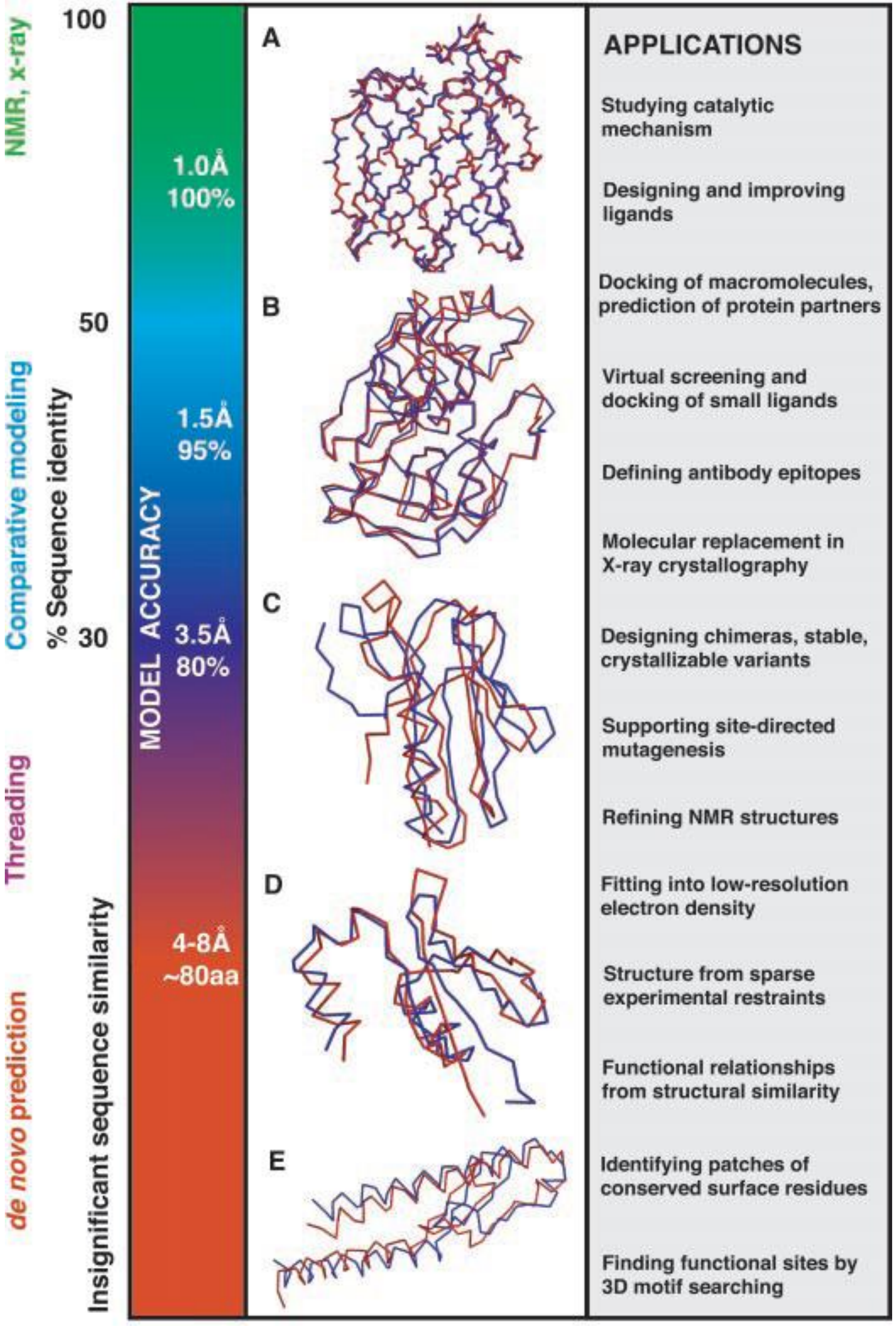


50% sequences covered (25% in 1995)

Relation between sequence identity and accuracy/applications

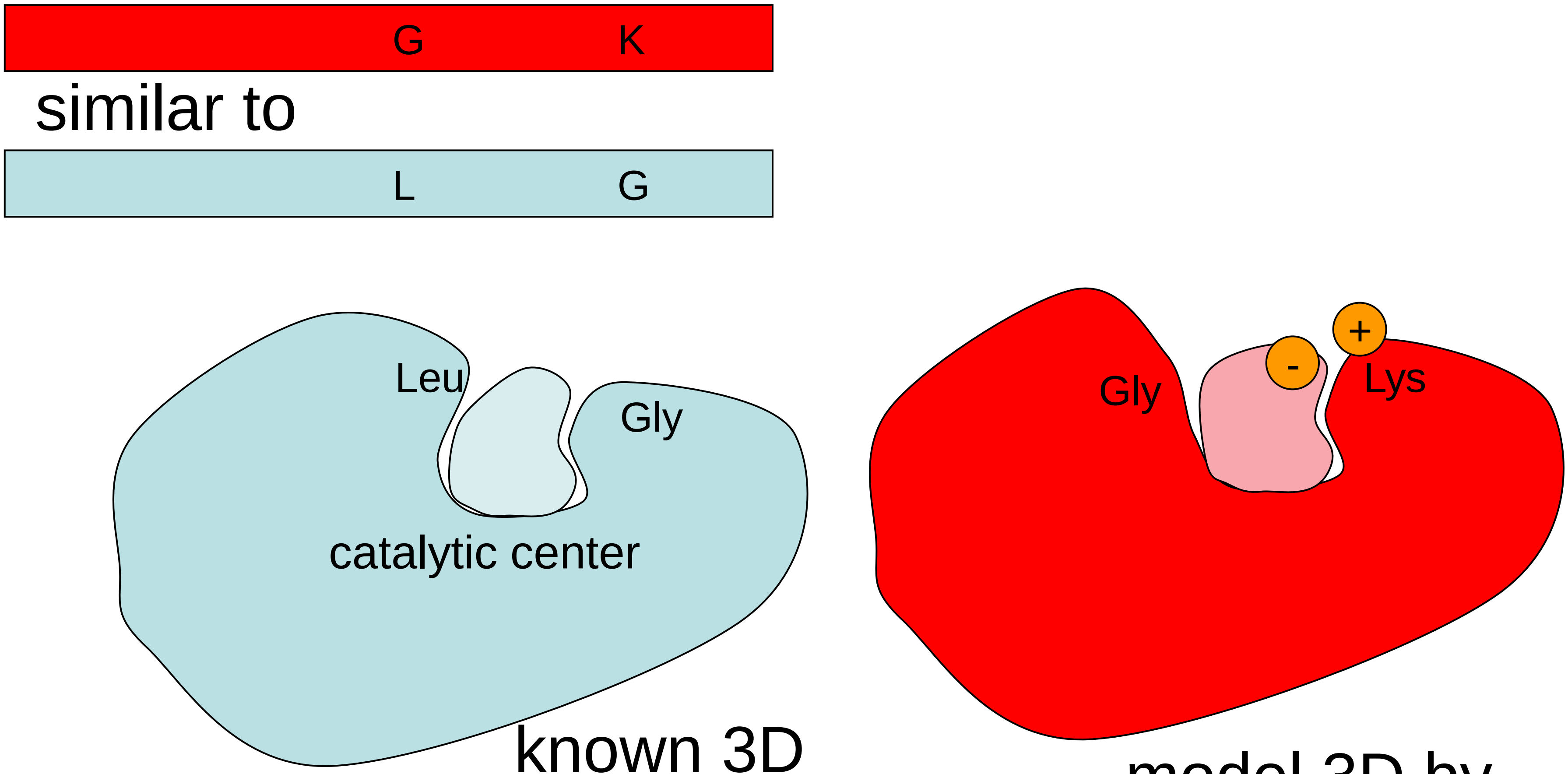
Predicted structure (red) and real (blue)

From:
Baker and Sali (2001) *Science*


NMR, x-ray
Comparative modeling
Threading
de novo prediction
100
50
30
% Sequence identity
Insignificant sequence similarity
MODEL ACCURACY
1.0Å
100%
1.5Å
95%
3.5Å
80%
4-8Å
~80aa
A
B
C
D
E
APPLICATIONS
Studying catalytic mechanism
Designing and improving ligands
Docking of macromolecules, prediction of protein partners
Virtual screening and docking of small ligands
Defining antibody epitopes
Molecular replacement in X-ray crystallography
Designing chimeras, stable, crystallizable variants
Supporting site-directed mutagenesis
Refining NMR structures
Fitting into low-resolution electron density
Structure from sparse experimental restraints
Functional relationships from structural similarity
Identifying patches of conserved surface residues
Finding functional sites by 3D motif searching

Homology modelling
Applications: target design

Query sequence

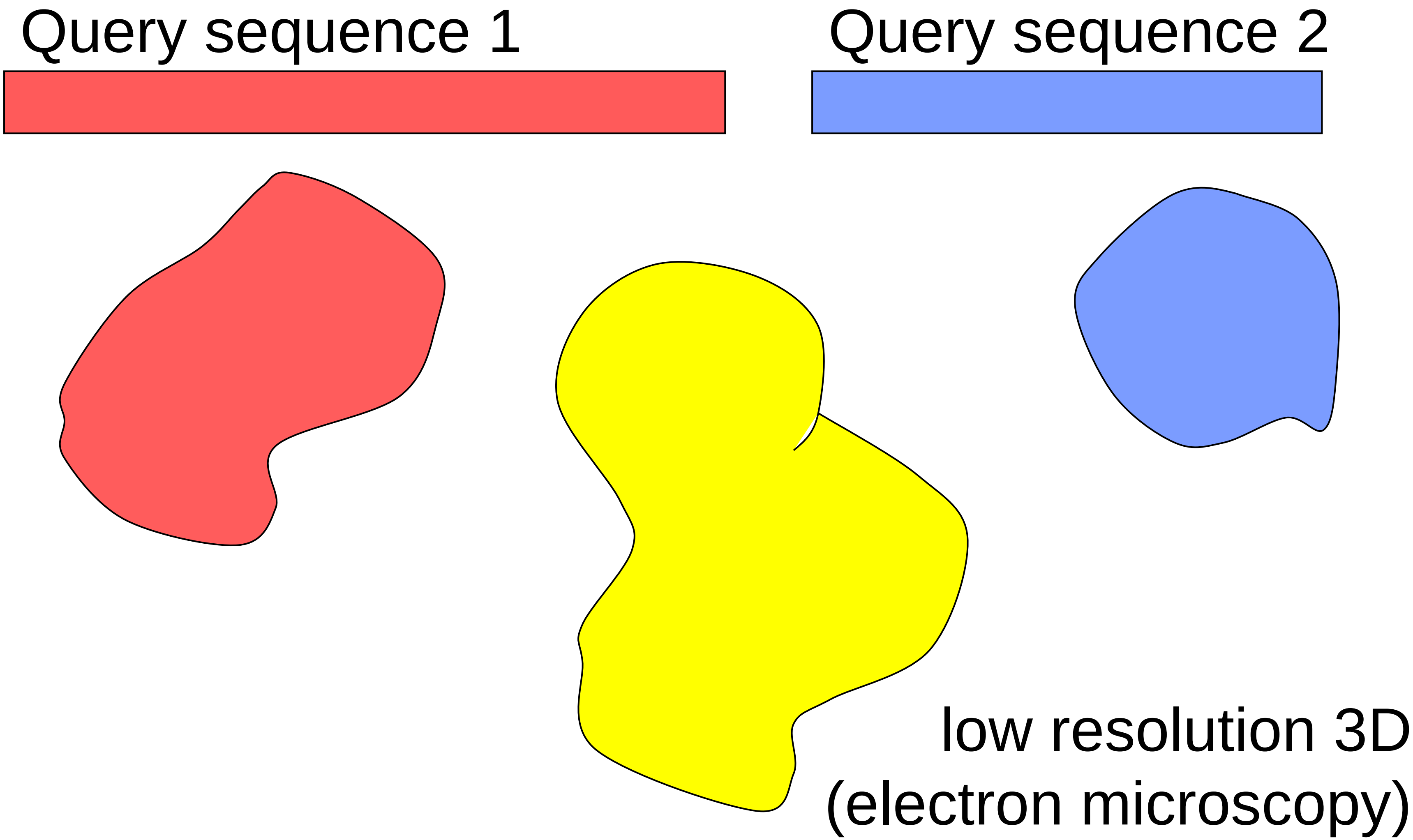


similar to

known 3D

model 3D by homology

Homology modelling
Applications: fit to low res 3D



Query sequence 1

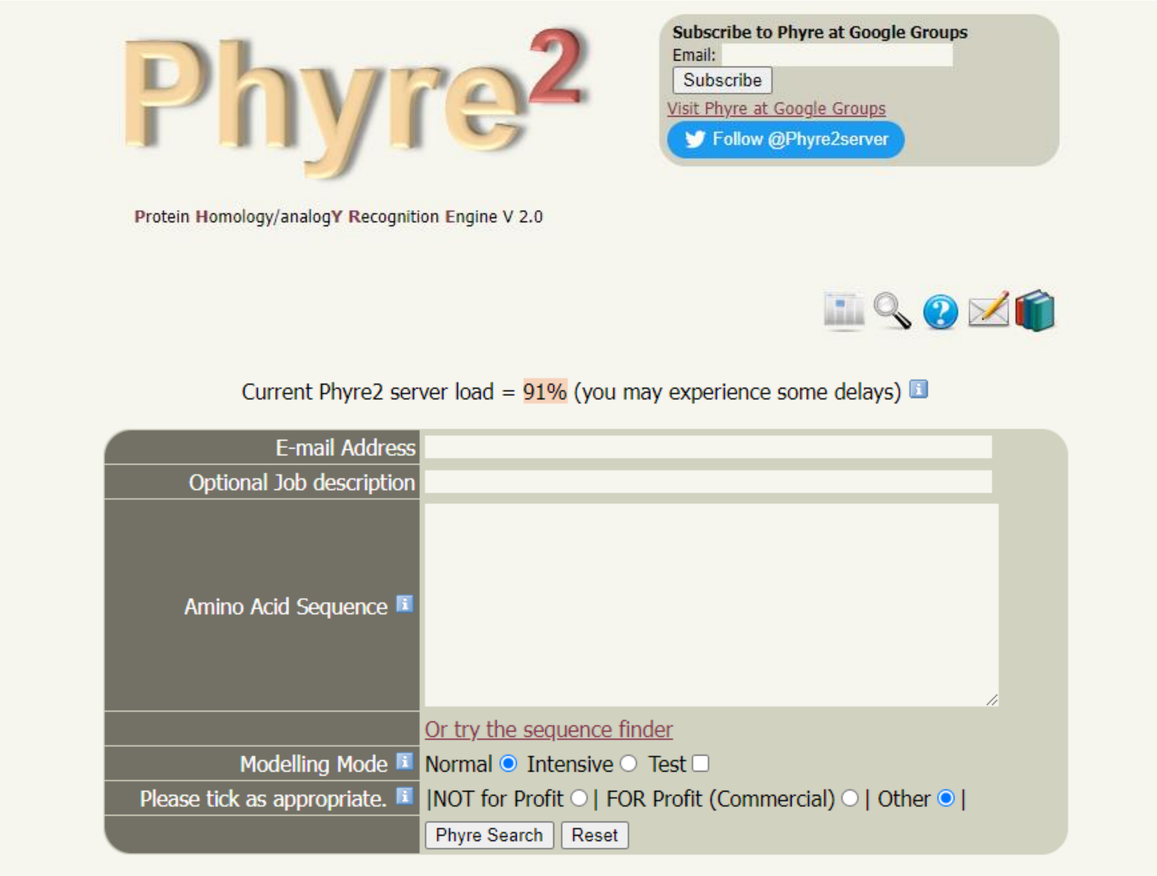
Query sequence 2

low resolution 3D
(electron microscopy)

Homology modelling
Phyre

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

Kelley et al (2000) *J Mol Biol*
Kelley et al (2015)
Nature Protocols



Phyre²

Protein Homology/analogY Recognition Engine V 2.0

Subscribe to Phyre at Google Groups
Email:
Subscribe
Visit Phyre at Google Groups
Follow @Phyre2server

Current Phyre2 server load = 91% (you may experience some delays)

E-mail Address	
Optional Job description	
Amino Acid Sequence	
	Or try the sequence finder
Modelling Mode	Normal ● Intensive ○ Test □
Please tick as appropriate.	|NOT for Profit ○ | FOR Profit (Commercial) ○ | Other ● |
	Phyre Search Reset

Processing time can be hours

Domains

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

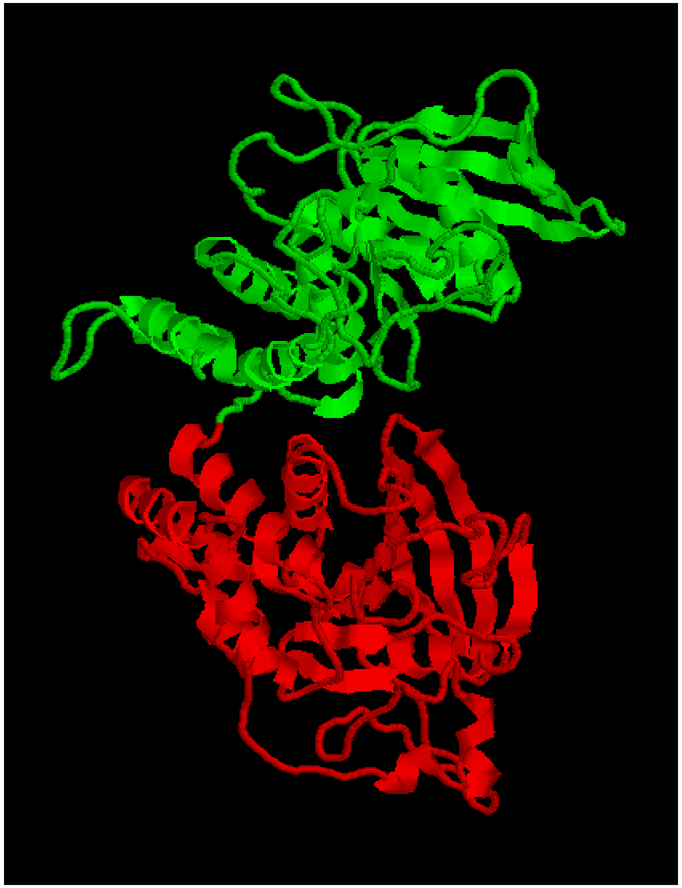
Function
Folding
Evolution

Proteins normally are multidomain (average 300 aa)

Domains

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

Function
Folding
Evolution



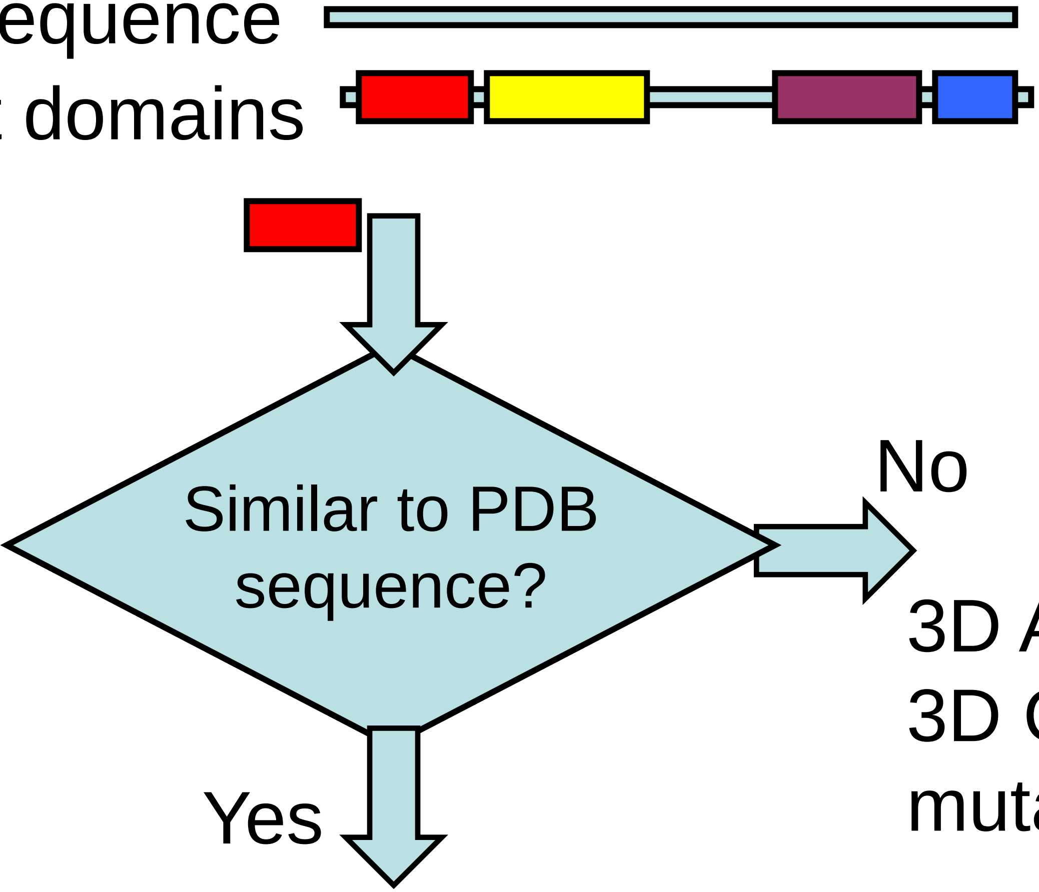
Proteins normally are multidomain (average 300 aa)

Domains

Query Sequence

Predict domains

Cut



Similar to PDB sequence?

No

3D Ab initio
3D Correlated mutations

Yes

Protein structure modeling by homology

Protein structure prediction
Ab initio

Explore conformational space

Limit the number of atoms

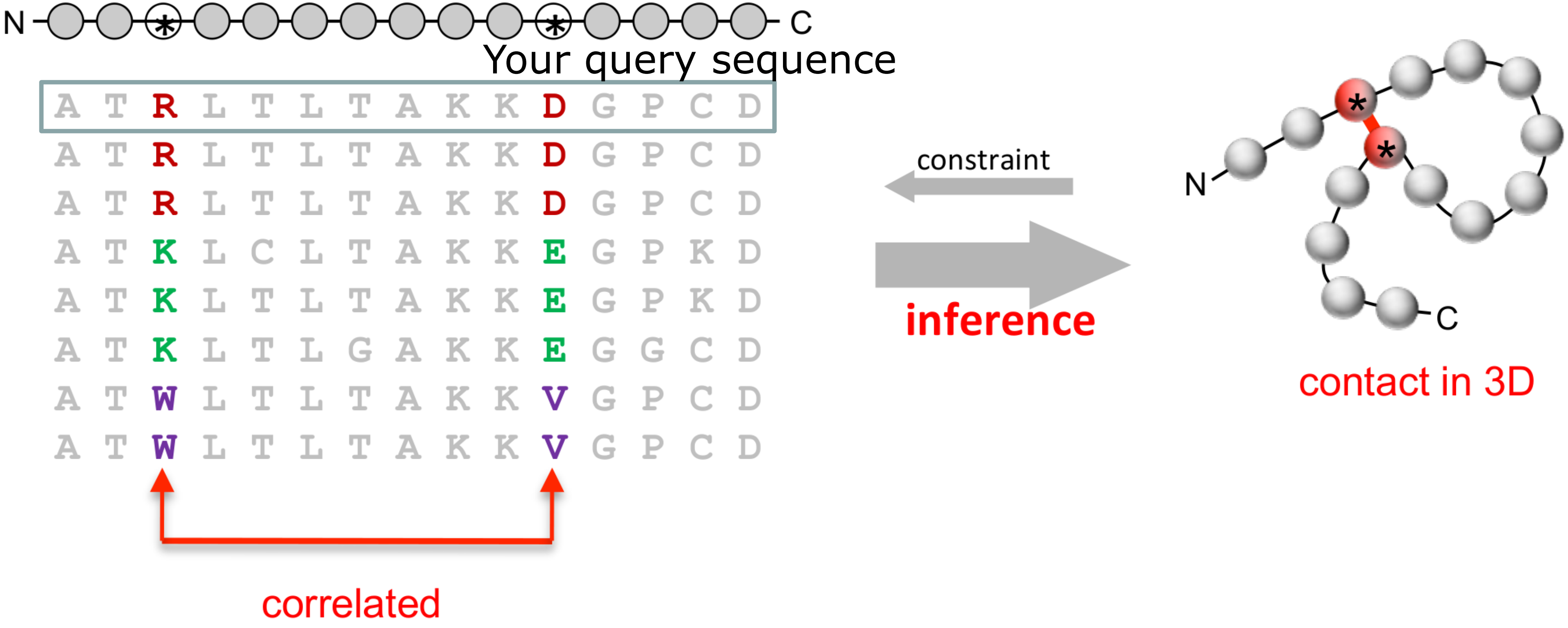
Break the problem into fragments of sequence

Optimize hydrophobic residue burial and pairing of beta-strands

Limited success…

Protein structure prediction

Correlated mutations



```
A T R L T L T A K K D G P C D
A T R L T L T A K K D G P C D
A T R L T L T A K K D G P C D
A T K L C L T A K K E G P K D
A T K L T L T A K K E G P K D
A T K L T L G A K K E G G C D
A T W L T L T A K K V G P C D
A T W L T L T A K K V G P C D
```

https://commons.wikimedia.org/wiki/File:Correlated_mutation.png

Protein structure prediction
Combined

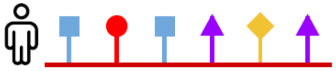
Homology to solved structures

Correlated sequence variation in homologs

Generation of a structure following physical constraints

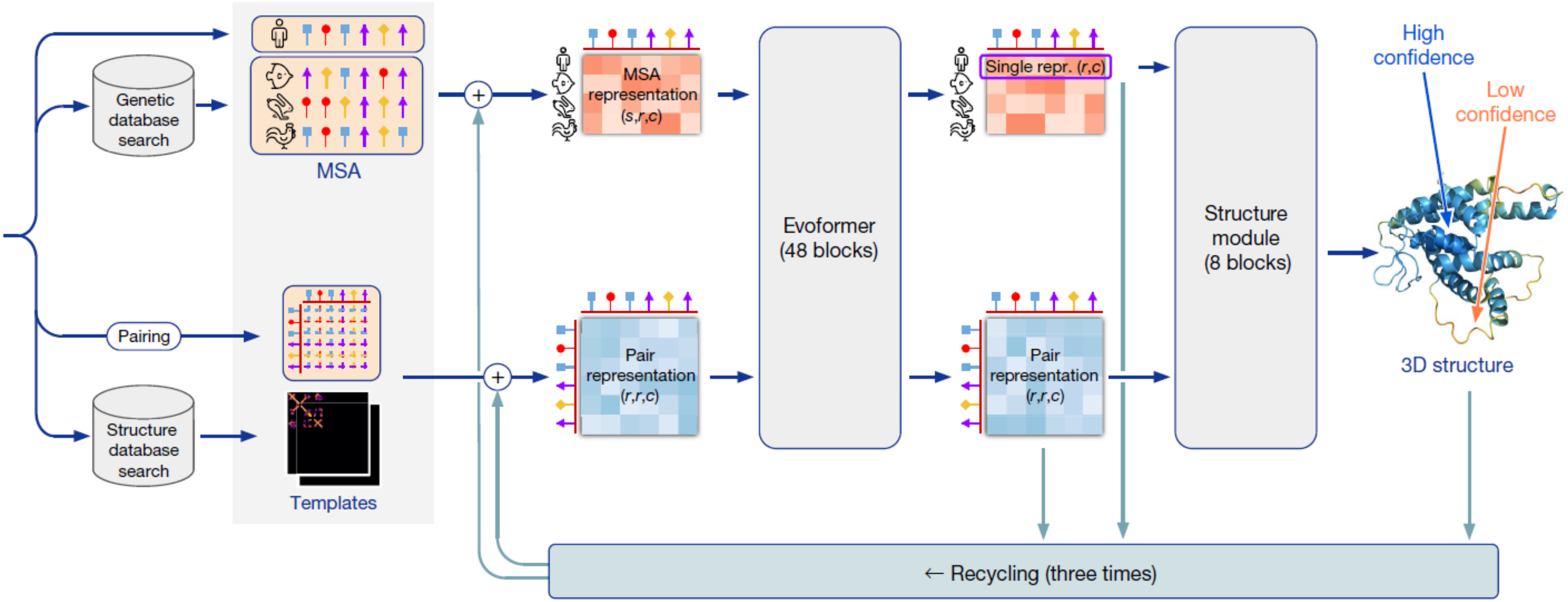
Protein structure prediction

AlphaFold: DeepMind, Google



Input sequence

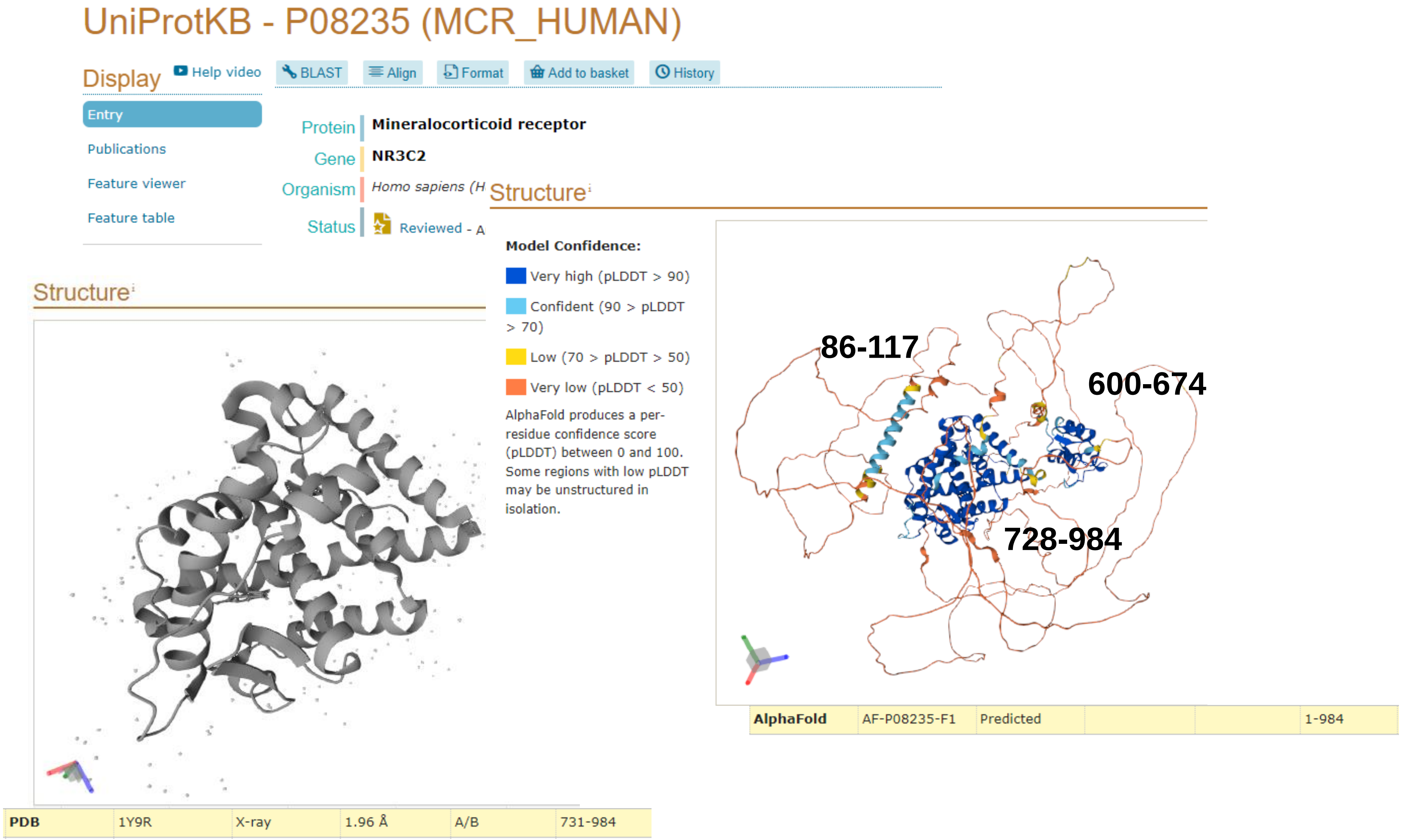
Demis Hassabis


Genetic database search
MSA
Pairing
Structure database search
Templates
MSA representation (s,r,c)
Pair representation (r,r,c)
Evoformer (48 blocks)
Single repr. (r,c)
Pair representation (r,r,c)
Structure module (8 blocks)
High confidence
Low confidence
3D structure
← Recycling (three times)

Jumper et al (2021) Nature

Protein structure prediction

AlphaFold: DeepMind, Google



UniProtKB - P08235 (MCR_HUMAN)

Protein **Mineralocorticoid receptor**

Gene **NR3C2**

Organism *Homo sapiens (H*

Status Reviewed - A

Structure

PDB	1Y9R	X-ray	1.96 Å	A/B	731-984

Structure

Model Confidence:

- 🟦 Very high (pLDDT > 90)
- 🟦 Confident (90 > pLDDT > 70)
- 🟨 Low (70 > pLDDT > 50)
- 🟧 Very low (pLDDT < 50)

AlphaFold produces a per-residue confidence score (pLDDT) between 0 and 100. Some regions with low pLDDT may be unstructured in isolation.

AlphaFold	AF-P08235-F1	Predicted			1-984

Protein structure prediction

AlphaFold: DeepMind, Google

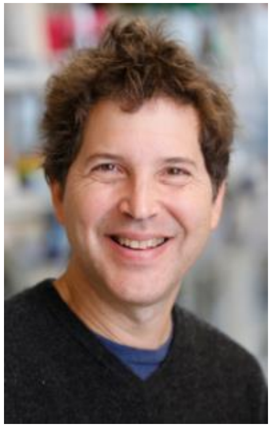
Precomputed models:
UniProt
https://alphafold.ebi.ac.uk/
(limited to model organisms)

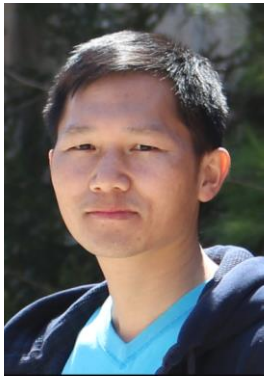
Colab notebook (simplified version / limited server / takes hours)

Source code (Needs 3 Tb disk space)

Protein structure prediction

trRosetta: David Baker & Jianyi Yang





Needs large multiple sequence alignments to predict contacts

Predictions available for all PFAM domains
Example:
https://www.ebi.ac.uk/interpro/entry/pfam/PF07887/rosettafold/

Run online at https://yanglab.nankai.edu.cn/trRosetta/

Du et al (2021) *Nature Protocols*

Protein structure prediction

C-I-Tasser: Yang Zhang

Run online at
https://zhanggroup.org/C-I-TASSER/

Zheng et al (2021) *Cell Reports Methods*