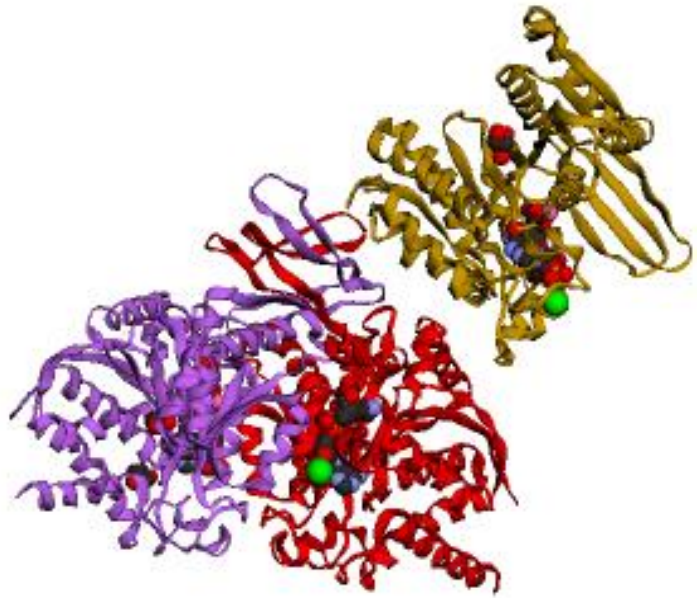
Bioinformatic insights into Isocitrate dehydrogenase 1 (IDH1) R132H mutation

Presented by: Pragathi Prasad

Guided by: Dr. Jhinuk Chatterjee



Content

- About the protein
- BLAST results
- MSA
- Phylogenetic analysis
- Secondary structure prediction
- Tertiary structure prediction
- Conserved domain
- Prosite,

About the protein

- **Isocitrate dehydrogenase 1** is one of three isocitrate dehydrogenase isozymes, the other two being IDH2 and IDH3.
- IDH1 catalyzes the **reversible oxidative decarboxylation** of **isocitrate to yield α-ketoglutarate (α-KG)** also known as **2-oxoglutarate** as part of the **TCA cycle** in glucose metabolism.
- **IDH1 (NADP+)**, soluble is a homodimer enzyme which is encoded by the IDH1 gene on chromosome 2 in humans
- The protein is found predominantly in the **cytosol and peroxisomes**.
- The best-studied mutation in IDH1 is **R132H**, which has been shown to act as a **tumor suppressor**. R132H mutation is the most important **prognostic factor** for the survival of glioma patients.

Mutation-disease association

- IDH1 mutations are **heterozygous**, typically involving an **amino acid substitution** in the active site of the enzyme in **codon 132**. These mutations are **somatic**, they primarily occur in cells involved brain and bone tumors.
- Mutations result in a loss of normal enzymatic function and the abnormal production of **2-hydroxyglutarate (2-HG)** , that has been found to inhibit enzymatic function of many alpha-ketoglutarate dependent dioxygenases, causing widespread changes in **histone and DNA methylation** and potentially promoting **tumorigenesis**.
- IDH1 has also been shown to harbor mutations in **diffused gliomas** and **human acute myeloid leukemia**.

PDB structure information

- PDB Id: **3INM**
- Name: **Crystal structure of human cytosolic NADP(+)-dependent isocitrate dehydrogenase R132H mutant in complex with NADPH, ALPHA-KETOGLUTARATE and CALCIUM(2+)**
- Resolution: **2.10 Å**
- Method: **X-Ray diffraction**
- Sequence Length: **425**
- Organism: **Homo sapiens**
- Date of release: **24-11-2009**

FASTA sequence

>3INM_1|Chains A, B, C| Isocitrate dehydrogenase [NADP] cytoplasmic| Homo sapiens (9606)

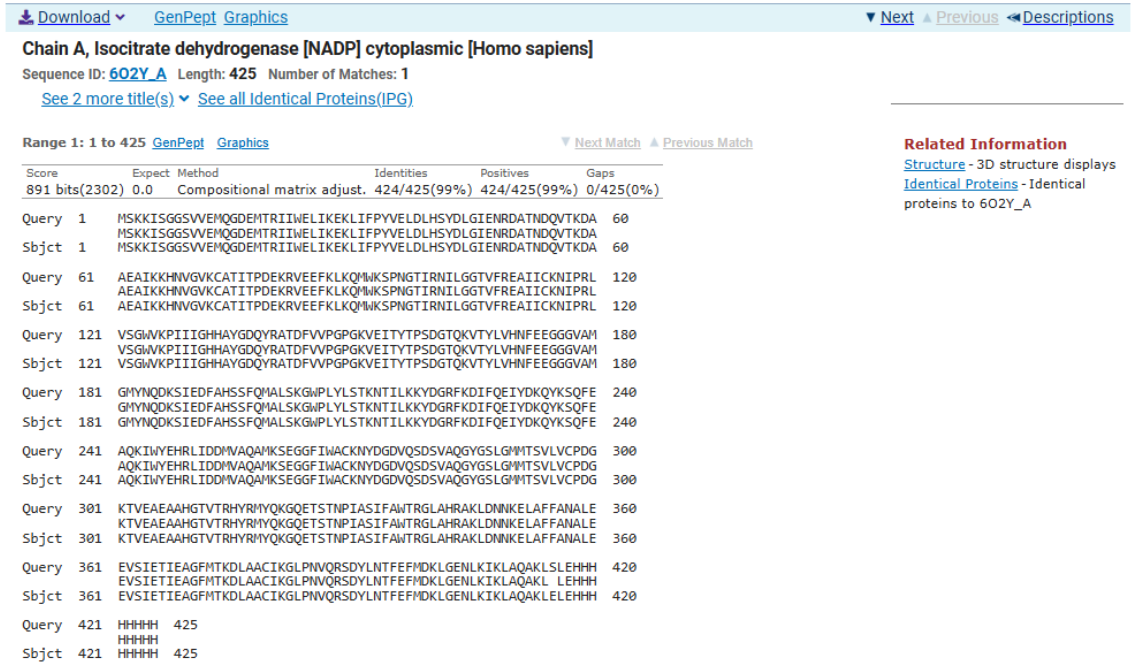
MSKKISGGSVVEMQGDEMTRIIWELIKEKLIFPYVELDLHSYDLGIENRDATNDQVTKDAAEAIKKHNV
GVKCATITPDEKRVEEFKLKQMWKSPNGTIRNILGGTVFREAIICKNIPRLVSGWVKPIIIGHHAYGDQY
RATDFVVPGPGKVEITYTPSDGTQKVTYLVHNFEEGGGVAMGMYNQDKSIEDFAHSSFQMALSKGWP
LYLSTKNTILKKYDGRFKDIFQEIYDKQYKSQFEAQKIWYEHRLIDDMVAQAMKSEGGFIWACKNYDGD
VQSDSVAQGYGSLGMMTSVLVCPDGKTVEAEAAHGTVTRHYRMYQKGQETSTNPIASIFAWTRGLAH
RAKLDNNKELAFFANALEEVSIETIEAGFMTKDLAACIKGLPNVQRSDYLNTFEFMDKLGENLKIKLA
QAKLSLEHHHHHHHH

BLASTP search

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
☐	Chain A, Isocitrate dehydrogenase [NADP] cytoplasmic [Homo sapiens]	Homo sapiens	892	892	100%	0.0	100.00%	425	3INM_A
☑	Chain A, Isocitrate dehydrogenase [NADP] cytoplasmic [Homo sapiens]	Homo sapiens	891	891	100%	0.0	99.76%	425	6O2Y_A
☐	Chain A, Isocitrate dehydrogenase [NADP] cytoplasmic [Homo sapiens]	Homo sapiens	890	890	100%	0.0	99.76%	425	6BKX_A
☐	Chain A, Isocitrate dehydrogenase [NADP] cytoplasmic [Homo sapiens]	Homo sapiens	888	888	100%	0.0	99.53%	425	4L04_A
☐	Chain A, Isocitrate dehydrogenase [NADP] cytoplasmic [Homo sapiens]	Homo sapiens	887	887	100%	0.0	99.53%	425	4L03_A
☐	Chain A, Isocitrate dehydrogenase [NADP] cytoplasmic [Homo sapiens]	Homo sapiens	887	887	100%	0.0	99.53%	425	4L06_A
☐	Chain A, Isocitrate dehydrogenase [NADP] cytoplasmic [Homo sapiens]	Homo sapiens	885	885	100%	0.0	99.76%	424	8T7N_A
☐	Chain A, Isocitrate dehydrogenase [NADP] cytoplasmic [Homo sapiens]	Homo sapiens	883	883	100%	0.0	99.53%	424	8T7D_A
☐	Chain A, Isocitrate dehydrogenase [NADP] cytoplasmic [Homo sapiens]	Homo sapiens	880	880	99%	0.0	99.76%	422	3MAP_A
☐	Chain A, Isocitrate dehydrogenase [NADP] cytoplasmic [Homo sapiens]	Homo sapiens	874	874	99%	0.0	99.29%	422	7PJM_A
☐	Chain B, Isocitrate dehydrogenase [NADP] cytoplasmic [Homo sapiens]	Homo sapiens	870	870	99%	0.0	99.05%	422	7PJN_B
☐	Chain A, Isocitrate dehydrogenase [NADP] cytoplasmic [Homo sapiens]	Homo sapiens	867	867	97%	0.0	100.00%	414	4I3K_A
☐	Chain A, Isocitrate dehydrogenase [NADP] cytoplasmic [Homo sapiens]	Homo sapiens	865	865	97%	0.0	100.00%	417	5SUN_A
☐	Chain A, Isocitrate dehydrogenase [NADP] cytoplasmic [Homo sapiens]	Homo sapiens	864	864	97%	0.0	99.76%	421	5YFN_A
☐	isocitrate dehydrogenase [NADP] cytoplasmic [Homo sapiens]	Homo sapiens	864	864	97%	0.0	99.76%	414	NP_001269315.1
☑	Chain B, Isocitrate dehydrogenase [NADP] cytoplasmic [Homo sapiens]	Homo sapiens	864	864	97%	0.0	99.76%	419	3MAS_B

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
isocitrate dehydrogenase [NADP] cytoplasmic [Symphalangus syndactylus]	Symphalangus syndactylus	862	862	97%	0.0	99.28%	414	XP_055146401.1
isocitrate dehydrogenase [NADP] cytoplasmic [Nomascus leucogenys]	Nomascus leucogenys	862	862	97%	0.0	99.52%	414	XP_030658594.1
unnamed protein product [Homo sapiens]	Homo sapiens	861	861	97%	0.0	99.52%	414	BAG35175.1
isocitrate dehydrogenase [NADP] cytoplasmic [Pan troglodytes]	Pan troglodytes	860	860	97%	0.0	99.28%	414	XP_001141655.1
isocitrate dehydrogenase 1 (NADP+), soluble [synthetic construct]	synthetic construct	860	860	97%	0.0	99.28%	414	BAI46991.1
isocitrate dehydrogenase [NADP] cytoplasmic [Gorilla gorilla gorilla]	Gorilla gorilla gorilla	860	860	97%	0.0	99.03%	414	XP_004033179.2
isocitrate dehydrogenase [NADP] cytoplasmic [Pongo pygmaeus]	Pongo pygmaeus	860	860	97%	0.0	99.03%	414	XP_054334237.1
isocitrate dehydrogenase [NADP] cytoplasmic [Macaca mulatta]	Macaca mulatta	856	856	97%	0.0	98.79%	414	NP_001252708.1
PREDICTED: isocitrate dehydrogenase [NADP] cytoplasmic isoform X1 [Cercocebus atys]	Cercocebus atys	855	855	97%	0.0	98.79%	414	XP_011903123.1
hypothetical protein [Pongo abelii]	Pongo abelii	854	854	97%	0.0	98.55%	414	CAH89719.1
isocitrate dehydrogenase [NADP] cytoplasmic [Rattus rattus]	Rattus rattus	840	840	97%	0.0	96.14%	414	XP_032757136.1
isocitrate dehydrogenase [NADP] cytoplasmic [Saccopteryx bilineata]	Saccopteryx bilineata	840	840	97%	0.0	96.38%	414	XP_066135898.1
isocitrate dehydrogenase [NADP] cytoplasmic [Molossus molossus]	Molossus molossus	840	840	97%	0.0	96.38%	414	XP_036110230.1
isocitrate dehydrogenase [NADP] cytoplasmic [Equus caballus]	Equus caballus	840	840	97%	0.0	96.38%	414	NP_001182456.1
isocitrate dehydrogenase [NADP] cytoplasmic [Marmota flaviventris]	Marmota flaviventris	840	840	97%	0.0	96.38%	414	XP_027797639.1
isocitrate dehydrogenase [NADP] cytoplasmic [Marmota marmota marmota]	Marmota marmota marmota	840	840	97%	0.0	96.38%	414	XP_015336803.1
isocitrate dehydrogenase [NADP] cytoplasmic [Marmota monax]	Marmota monax	840	840	97%	0.0	96.14%	414	KAF7474283.1
isocitrate dehydrogenase [NADP] cytoplasmic [Peromyscus leucopus]	Peromyscus leucopus	839	839	97%	0.0	96.14%	414	XP_028742540.1
isocitrate dehydrogenase [NADP] cytoplasmic [Jaculus jaculus]	Jaculus jaculus	839	839	97%	0.0	95.65%	414	XP_004653586.1
isocitrate dehydrogenase [NADP] cytoplasmic [Neogale vison]	Neogale vison	839	839	97%	0.0	96.38%	414	XP_044097392.1
Isocitrate dehydrogenase [NADP] cytoplasmic [Heterocephalus glaber]	Heterocephalus glaber	839	839	97%	0.0	96.38%	414	EHB04859.1
isocitrate dehydrogenase [NADP] cytoplasmic [Ailuropoda melanoleuca]	Ailuropoda melanoleuca	839	839	97%	0.0	96.14%	414	XP_002918143.1
isocitrate dehydrogenase [NADP] cytoplasmic isoform X1 [Mastomys coucha]	Mastomys coucha	839	839	97%	0.0	95.65%	420	XP_031223625.1
isocitrate dehydrogenase [NADP] cytoplasmic [Nycticebus coucang]	Nycticebus coucang	839	839	97%	0.0	96.62%	414	XP_053453443.1
isocitrate dehydrogenase [NADP] cytoplasmic [Rattus norvegicus]	Rattus norvegicus	839	839	97%	0.0	96.14%	414	NP_113698.1
isocitrate dehydrogenase [NADP] cytoplasmic [Pteropus alecto]	Pteropus alecto	839	839	97%	0.0	96.38%	414	XP_006921466.1
isocitrate dehydrogenase [NADP] cytoplasmic isoform X2 [Mastomys coucha]	Mastomys coucha	838	838	97%	0.0	95.65%	414	XP_031223629.1
isocitrate dehydrogenase [NADP] cytoplasmic [Mesocricetus auratus]	Mesocricetus auratus	838	838	97%	0.0	96.62%	414	XP_005070614.2
isocitrate dehydrogenase [NADP] cytoplasmic [Apodemus sylvaticus]	Apodemus sylvaticus	838	838	97%	0.0	95.65%	414	XP_052049147.1
isocitrate dehydrogenase [NADP] cytoplasmic [Cricetulus griseus]	Cricetulus griseus	838	838	97%	0.0	96.14%	414	XP_003506959.1

Pairwise alignment



Download GenPept Graphics ▼ Next ▲ Previous ◀◀ Descriptions

Chain A, Isocitrate dehydrogenase [NADP] cytoplasmic [Homo sapiens]

Sequence ID: 6O2Y_A Length: 425 Number of Matches: 1

See 2 more title(s) See all Identical Proteins(IPG)

Range 1: 1 to 425 GenPept Graphics ▼ Next Match ▲ Previous Match

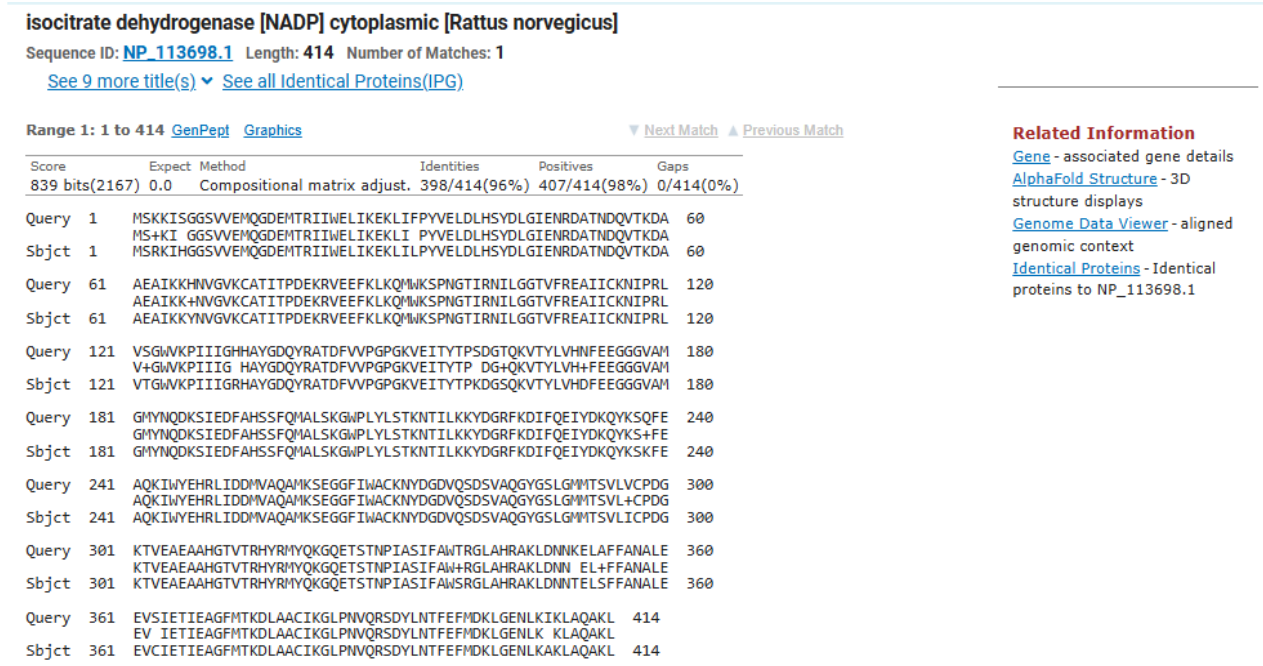
Score	Expect	Method	Identities	Positives	Gaps
891 bits(2302)	0.0	Compositional matrix adjust.	424/425(99%)	424/425(99%)	0/425(0%)

```
Query  1    MSKKISGGSVVEMQGDEMTRIIWELIKEKLIFPYVELDLHSYDLGIENRDATNDQVTKDA  60
            MSKKISGGSVVEMQGDEMTRIIWELIKEKLIFPYVELDLHSYDLGIENRDATNDQVTKDA
Sbjct  1    MSKKISGGSVVEMQGDEMTRIIWELIKEKLIFPYVELDLHSYDLGIENRDATNDQVTKDA  60

Query  61   AEAIKKHNVGVKCATITPDEKRVEEFKLKQMWKSPNGTIRNILGGTVFREAIICKNIPRL  120
            AEAIKKHNVGVKCATITPDEKRVEEFKLKQMWKSPNGTIRNILGGTVFREAIICKNIPRL
Sbjct  61   AEAIKKHNVGVKCATITPDEKRVEEFKLKQMWKSPNGTIRNILGGTVFREAIICKNIPRL  120

Query  121  VSGWVKPIIIGHHAYGDQYRATDFVVPGPGKVEITYTPSDGTQKVTYLVHNFEEGGGVAM  180
            VSGWVKPIIIGHHAYGDQYRATDFVVPGPGKVEITYTPSDGTQKVTYLVHNFEEGGGVAM
Sbjct  121  VSGWVKPIIIGHHAYGDQYRATDFVVPGPGKVEITYTPSDGTQKVTYLVHNFEEGGGVAM  180

Query  181  GMYNQDKSIEDFAHSSFQMALSKGWPLYLSTKNTILKKYDGRFKDIFQEIYDKQYKSQFE  240
            GMYNQDKSIEDFAHSSFQMALSKGWPLYLSTKNTILKKYDGRFKDIFQEIYDKQYKSQFE
Sbjct  181  GMYNQDKSIEDFAHSSFQMALSKGWPLYLSTKNTILKKYDGRFKDIFQEIYDKQYKSQFE  240

Query  241  AQKIWYEHRLIDDMVAQAMKSEGGFIWACKNYDGDVQSDSVAQGYGSLGMMTSVLVCPDG  300
            AQKIWYEHRLIDDMVAQAMKSEGGFIWACKNYDGDVQSDSVAQGYGSLGMMTSVLVCPDG
Sbjct  241  AQKIWYEHRLIDDMVAQAMKSEGGFIWACKNYDGDVQSDSVAQGYGSLGMMTSVLVCPDG  300

Query  301  KTVEAEAAHGTVTRHYRMYQKGQETSTNPIASIFAWTRGLAHRAKLDNNKELAFFANALE  360
            KTVEAEAAHGTVTRHYRMYQKGQETSTNPIASIFAWTRGLAHRAKLDNNKELAFFANALE
Sbjct  301  KTVEAEAAHGTVTRHYRMYQKGQETSTNPIASIFAWTRGLAHRAKLDNNKELAFFANALE  360

Query  361  EVSIETIEAGFMTKDLAACIKGLPNVQRSDYLNTFEFMDKLGENLKIKLAQAKLSLEHHH  420
            EVSIETIEAGFMTKDLAACIKGLPNVQRSDYLNTFEFMDKLGENLKIKLAQAKL LEHHH
Sbjct  361  EVSIETIEAGFMTKDLAACIKGLPNVQRSDYLNTFEFMDKLGENLKIKLAQAKLELEHHH  420

Query  421  HHHHH  425
            HHHHH
Sbjct  421  HHHHH  425
```

Related Information

Structure - 3D structure displays

Identical Proteins - Identical proteins to 6O2Y_A



isocitrate dehydrogenase [NADP] cytoplasmic [Rattus norvegicus]

Sequence ID: NP_113698.1 Length: 414 Number of Matches: 1

See 9 more title(s) See all Identical Proteins(IPG)

Range 1: 1 to 414 GenPept Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
839 bits(2167)	0.0	Compositional matrix adjust.	398/414(96%)	407/414(98%)	0/414(0%)

```
Query  1    MSKKISGGSVVEMQGDEMTRIIWELIKEKLIFPYVELDLHSYDLGIENRDATNDQVTKDA  60
            MS+KI GGSVVEMQGDEMTRIIWELIKEKLI PYVELDLHSYDLGIENRDATNDQVTKDA
Sbjct  1    MSRKIHGGSVVEMQGDEMTRIIWELIKEKLILPYVELDLHSYDLGIENRDATNDQVTKDA  60

Query  61   AEAIKKHNVGVKCATITPDEKRVEEFKLKQMWKSPNGTIRNILGGTVFREAIICKNIPRL  120
            AEAIKK+NVGVKCATITPDEKRVEEFKLKQMWKSPNGTIRNILGGTVFREAIICKNIPRL
Sbjct  61   AEAIKKYNVGVKCATITPDEKRVEEFKLKQMWKSPNGTIRNILGGTVFREAIICKNIPRL  120

Query  121  VSGWVKPIIIGHHAYGDQYRATDFVVPGPGKVEITYTPSDGTQKVTYLVHNFEEGGGVAM  180
            V+GWVKPIIIG HAYGDQYRATDFVVPGPGKVEITYTP DG+QKVTYLVH+FEEGGGVAM
Sbjct  121  VTGWVKPIIIGRHAYGDQYRATDFVVPGPGKVEITYTPKDGSQKVTYLVHDFEEGGGVAM  180

Query  181  GMYNQDKSIEDFAHSSFQMALSKGWPLYLSTKNTILKKYDGRFKDIFQEIYDKQYKSQFE  240
            GMYNQDKSIEDFAHSSFQMALSKGWPLYLSTKNTILKKYDGRFKDIFQEIYDKQYKS+FE
Sbjct  181  GMYNQDKSIEDFAHSSFQMALSKGWPLYLSTKNTILKKYDGRFKDIFQEIYDKQYKSKFE  240

Query  241  AQKIWYEHRLIDDMVAQAMKSEGGFIWACKNYDGDVQSDSVAQGYGSLGMMTSVLVCPDG  300
            AQKIWYEHRLIDDMVAQAMKSEGGFIWACKNYDGDVQSDSVAQGYGSLGMMTSVL+CPDG
Sbjct  241  AQKIWYEHRLIDDMVAQAMKSEGGFIWACKNYDGDVQSDSVAQGYGSLGMMTSVLICPDG  300

Query  301  KTVEAEAAHGTVTRHYRMYQKGQETSTNPIASIFAWTRGLAHRAKLDNNKELAFFANALE  360
            KTVEAEAAHGTVTRHYRMYQKGQETSTNPIASIFAW+RGLAHRAKLDNN EL+FFANALE
Sbjct  301  KTVEAEAAHGTVTRHYRMYQKGQETSTNPIASIFAWSRGLAHRAKLDNNTELSFFANALE  360

Query  361  EVSIETIEAGFMTKDLAACIKGLPNVQRSDYLNTFEFMDKLGENLKIKLAQAKL  414
            EV IETIEAGFMTKDLAACIKGLPNVQRSDYLNTFEFMDKLGENLK KLAQAKL
Sbjct  361  EVCIETIEAGFMTKDLAACIKGLPNVQRSDYLNTFEFMDKLGENLKAKLAQAKL  414
```

Related Information

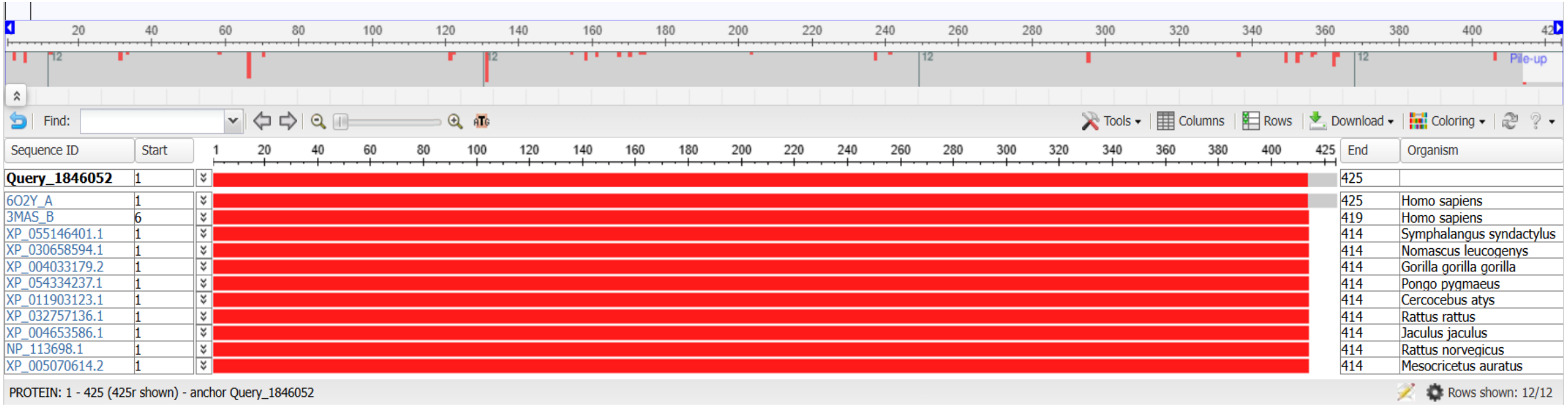
Gene - associated gene details

AlphaFold Structure - 3D structure displays

Genome Data Viewer - aligned genomic context

Identical Proteins - Identical proteins to NP_113698.1

Multiple sequence alignment



Multiple sequence alignment

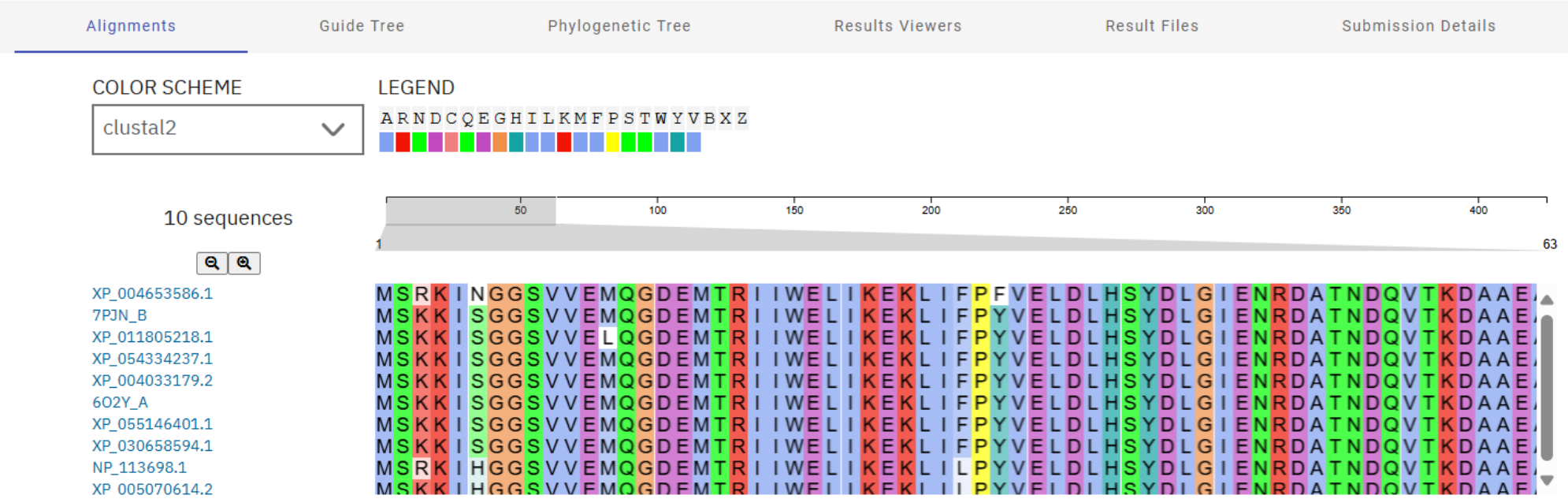
Clustal omega results

```
XP_004653586.1      MSRKINGGSVVEMQGDEMTRIIWELIKEKLIFPFVELDLHSYDLGIENRDATNDQVTKDA      60
7PJN_B              MSKKISGGSVVEMQGDEMTRIIWELIKEKLIFPYVELDLHSYDLGIENRDATNDQVTKDA      60
XP_011805218.1      MSKKISGGSVVELQGDEMTRIIWELIKEKLIFPYVELDLHSYDLGIENRDATNDQVTKDA      60
XP_054334237.1      MSKKISGGSVVEMQGDEMTRIIWELIKEKLIFPYVELDLHSYDLGIENRDATNDQVTKDA      60
XP_004033179.2      MSKKISGGSVVEMQGDEMTRIIWELIKEKLIFPYVELDLHSYDLGIENRDATNDQVTKDA      60
6O2Y_A              MSKKISGGSVVEMQGDEMTRIIWELIKEKLIFPYVELDLHSYDLGIENRDATNDQVTKDA      60
XP_055146401.1      MSKKISGGSVVEMQGDEMTRIIWELIKEKLIFPYVELDLHSYDLGIENRDATNDQVTKDA      60
XP_030658594.1      MSKKISGGSVVEMQGDEMTRIIWELIKEKLIFPYVELDLHSYDLGIENRDATNDQVTKDA      60
NP_113698.1         MSRKIHGGSVVEMQGDEMTRIIWELIKEKLILPYVELDLHSYDLGIENRDATNDQVTKDA      60
XP_005070614.2      MSKKIHGGSVVEMQGDEMTRIIWELIKEKLILPYVELDLHSYDLGIENRDATNDQVTKDA      60
                    **:** ******:******************:*:**************************

XP_004653586.1      AEAIKKYNVGIKCATITPDEKRVEEFKLKQMWKSPNGTIRNILGGTVFREAIICKNIPRL      120
7PJN_B              AEAIKKHNVGVKCATITPDEKRVEEFKLKQMWKSPNGTIRNILGGTVFREAIICKNIPRL      120
XP_011805218.1      AEAIKKYNVGVKCATITPDEKRVEEFKLKQMWKSPNGTIRNILGGTVFREAIICKNIPRL      120
XP_054334237.1      AEAIKKYNVGVKCATITPDEKRVEEFKLKQMWKSPNGTIRNILGGTVFREAIICKNIPRL      120
XP_004033179.2      AEAIKKYNVGVKCATITPDEKRVEEFKLKQMWKSPNGTIRNILGGTVFREAIICKNIPRL      120
6O2Y_A              AEAIKKHNVGVKCATITPDEKRVEEFKLKQMWKSPNGTIRNILGGTVFREAIICKNIPRL      120
XP_055146401.1      AEAIKKYNVGIKCATITPDEKRVEEFKLKQMWKSPNGTIRNILGGTVFREAIICKNIPRL      120
XP_030658594.1      AEAIKKYNVGVKCATITPDEKRVEEFKLKQMWKSPNGTIRNILGGTVFREAIICKNIPRL      120
NP_113698.1         AEAIKKYNVGVKCATITPDEKRVEEFKLKQMWKSPNGTIRNILGGTVFREAIICKNIPRL      120
XP_005070614.2      AEAIKKYNVGVKCATITPDEKRVEEFKLKQMWKSPNGTIRNILGGTVFREAIICKNIPRL      120
                    ******:***:**************************************************
```

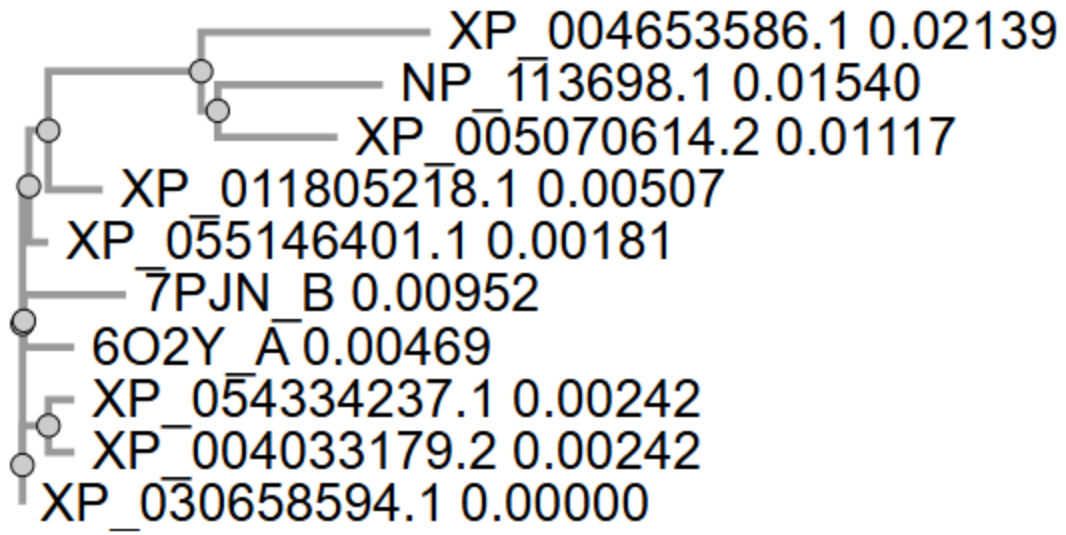
```
XP_004653586.1      EVCIETIEAGFMTKDLAACIKGLPNVQRSDYLNTFEFMDKLGENLKIKLAQAKL------      414
7PJN_B              EVSIETIEAGFMTKDLAAXIKGLPNVQRSDYLNTFEFMDKLGENLKIKLAQAKLLEHHHH      420
XP_011805218.1      EVCIETIEAGFMTKDLAACIKGLPNVQRSDYLNTFEFMDKLGENLKIKLAQAKL------      414
XP_054334237.1      EVSVETIEAGFMTKDLAACIKGLPNVQRSDYLNTFEFMDKLGENLKIKLAQAKL------      414
XP_004033179.2      EVSVETIEAGFMTKDLAACIKGLPNVQRSDYLNTFEFMDKLGENLKIKLAQAKL------      414
6O2Y_A              EVSIETIEAGFMTKDLAACIKGLPNVQRSDYLNTFEFMDKLGENLKIKLAQAKLELEHHH      420
XP_055146401.1      EVSIETIEAGFMTKDLAACIKGLPNVQRSDYLNTFEFMDKLGENLKIKLAQAKL------      414
XP_030658594.1      EVSIETIEAGFMTKDLAACIKGLPNVQRSDYLNTFEFMDKLGENLKIKLAQAKL------      414
NP_113698.1         EVCIETIEAGFMTKDLAACIKGLPNVQRSDYLNTFEFMDKLGENLKAKLAQAKL------      414
XP_005070614.2      EVCIETIEAGFMTKDLAACIKGLPNVQRSDYLNTFEFMDKLGENLKTKLAQAKL------      414
                    **.:************** ************************** *******

XP_004653586.1      -----       414
7PJN_B              HH---       422
XP_011805218.1      -----       414
XP_054334237.1      -----       414
XP_004033179.2      -----       414
6O2Y_A              HHHHH       425
XP_055146401.1      -----       414
XP_030658594.1      -----       414
NP_113698.1         -----       414
XP_005070614.2      -----       414
```

Multiple sequence alignment

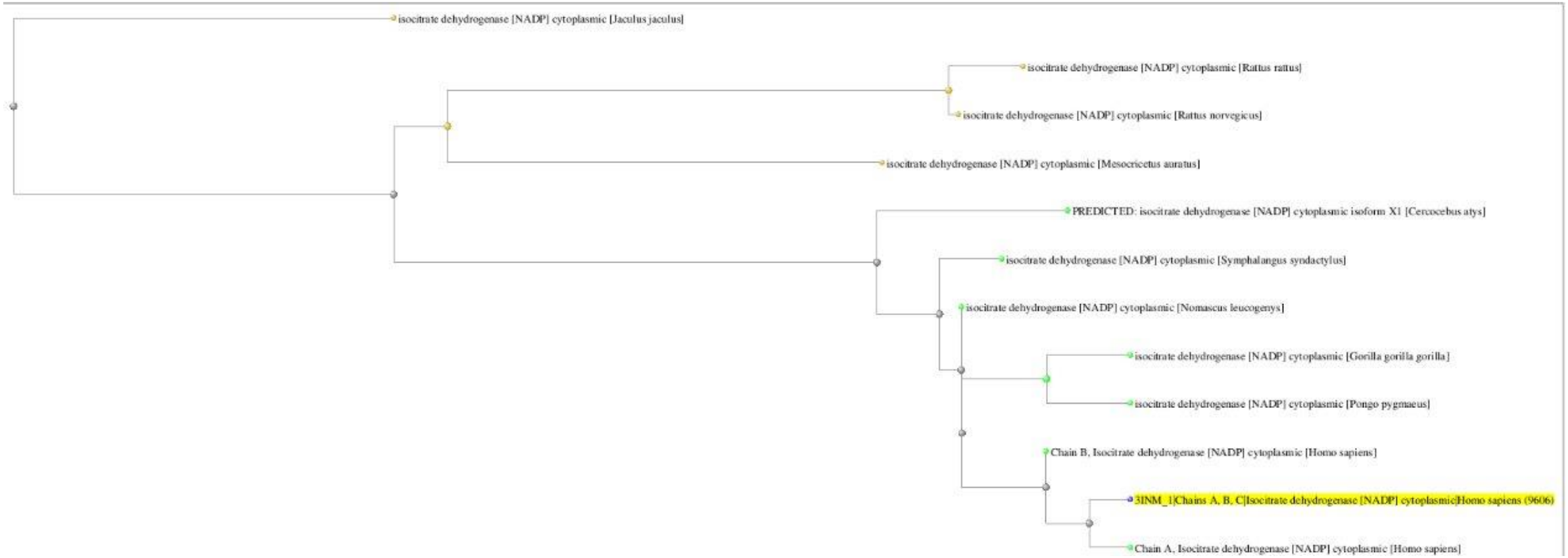
Clustal omega results



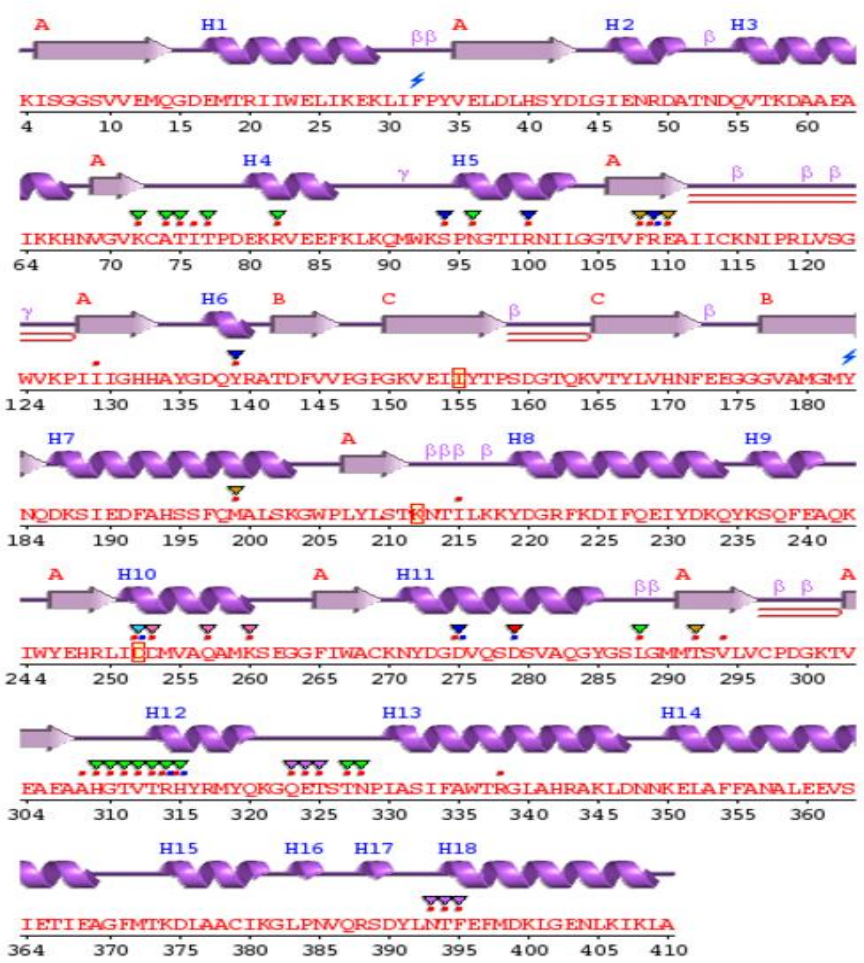
Phylogenetic tree result: ClustalW



Phylogenetic tree result: Blast tree viewer

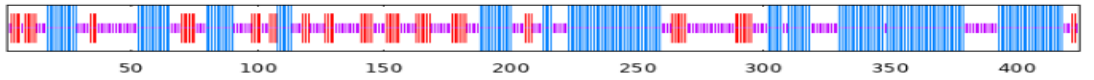


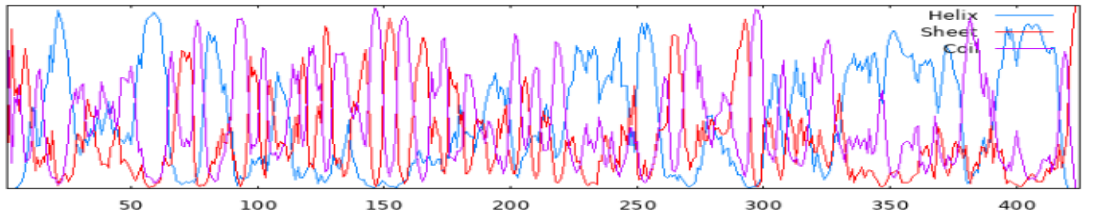
Secondary structure – PDBSum and GOR4



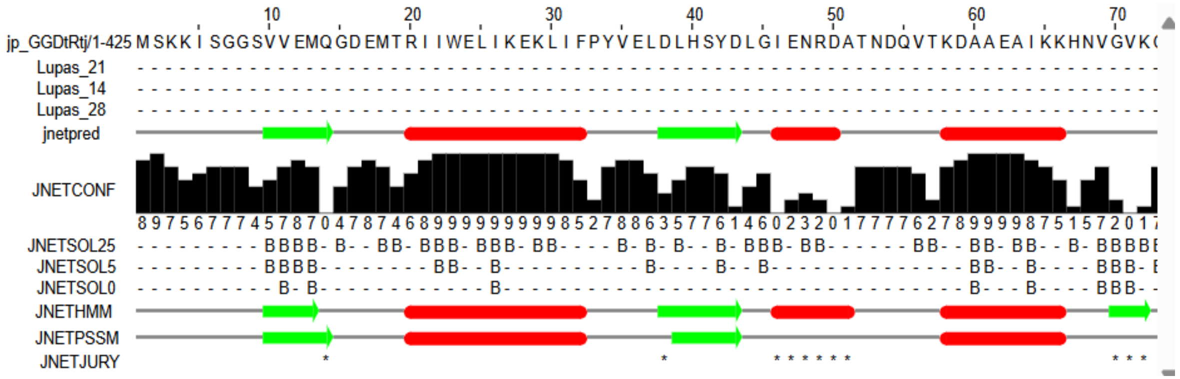

```
         10        20        30        40        50        60
          |         |         |         |         |         |
MSKKISGGSVVEMQGDEMTRIIWELIKEKLIFPYVELDLHSYDLGIENRDATNDQVTKDA
ccceeeeceeeeecccchhhhhhhhhhhcccccceeecccccccccccccccchhhhhhh
AEAIKKHNVGVKCATITPDEKRVEEFKLKQMWKSPNGTIRNILGGTVFREAIICKNIPRL
hhhhhhccccceeeeeccccchhhhhhhhhhhccccccceeeeccceeehhhhhhccccee
VSGWVKPIIIGHHAYGDQYRATDFVVPGPGKVEIITYTPSDGTQKVTYLVHNFEEGGGVAM
ecccccceeeeccccccccccceeeeecccccceeeeeecccccceeeeeecccccccceee
GMYNQDKSIEDFAHSSFQMALSKGWPLYLSTKNTILKKYDGRFKDIFQEIYDKQYKSQFE
eeecccccchhhhhhhhhhhhccccceeecccchhhhcccccchhhhhhhhhhhhhhhhhh
AQKIWYEHRLIDDMVAQAMKSEGGFIWACKNYDGDVQSDSVAQGYGSLGMMTSVLVCPDG
hhhhhhhhhhhhhhhhhhhcccceeeeeecccccccccccccccccceeeeeeecccc
KTVEAEAAHGTVTRHYRMYQKGQETSTNPIASIFAWTRGLAHRAKLDNNKELAFFANALE
cchhhhhccchhhhhhhhhcccccccccchhhhhhhhhhhhhhhhhchhhhhhhhhhh
EVSIETIEAGFMTKDLAACIKGLPNVQRSDYLNTFEFMDKLGENLKIKLAQAKLSLEHHH
hhhhhhhhhhhhhhhhhhhhcccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhc
HHHHH
cceec
```

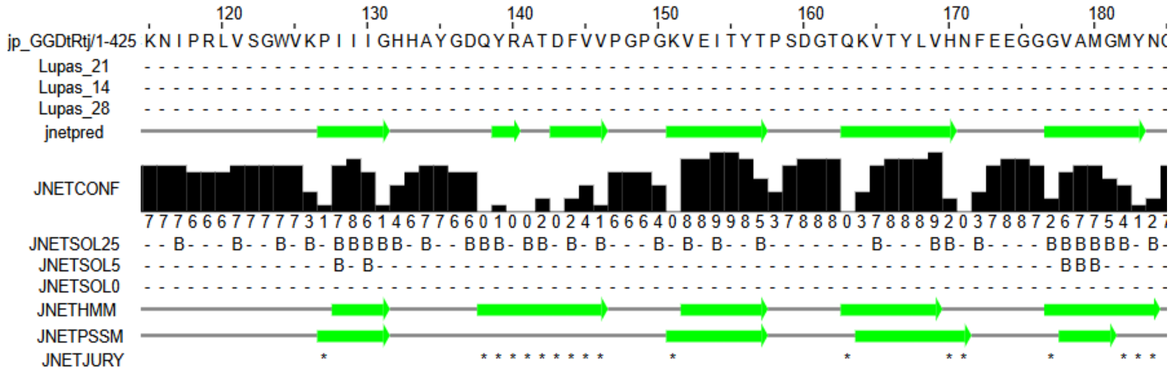
```
Sequence length :   425

GOR4 :
   Alpha helix      (Hh) :   186 is  43.76%
   3₁₀ helix        (Gg) :     0 is   0.00%
   Pi helix         (Ii) :     0 is   0.00%
   Beta bridge      (Bb) :     0 is   0.00%
   Extended strand  (Ee) :    73 is  17.18%
   Beta turn        (Tt) :     0 is   0.00%
   Bend region      (Ss) :     0 is   0.00%
   Random coil      (Cc) :   166 is  39.06%
   Ambiguous states (?)  :     0 is   0.00%
   Other states          :     0 is   0.00%
```

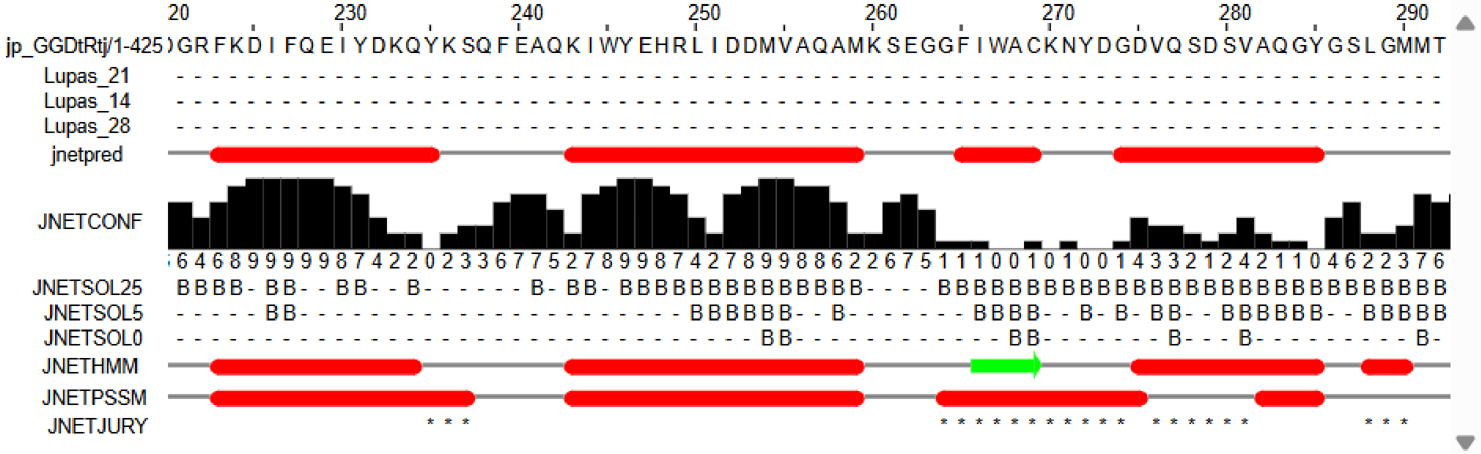




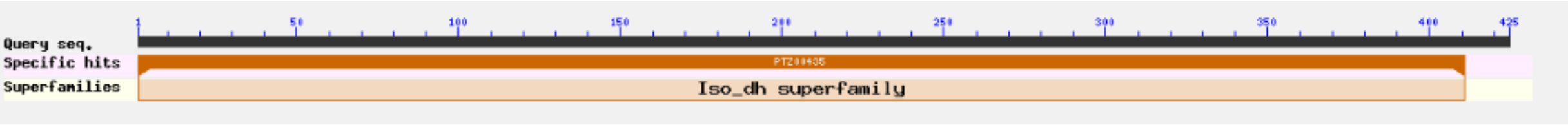
Secondary structure – Jpred



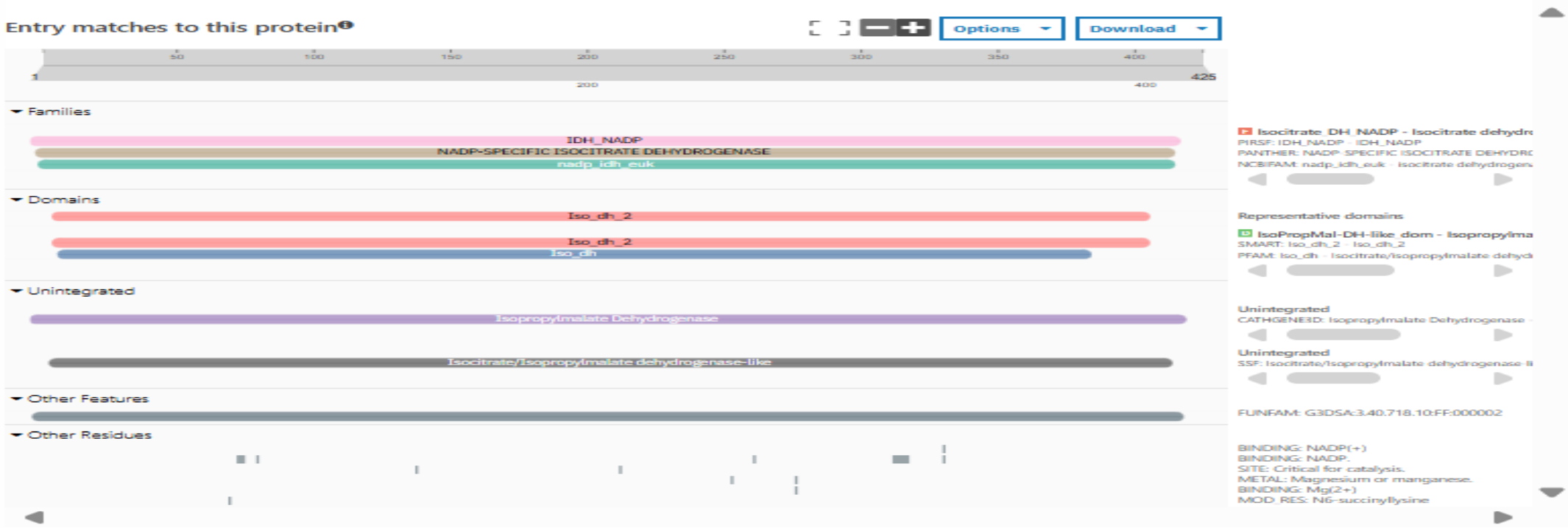




Conserved domains



InterPro scan results



InterPro GO terms

Biological Process

- isocitrate metabolic process (GO:0006102)

Molecular Function

- isocitrate dehydrogenase (NADP+) activity (GO:0004450)
- magnesium ion binding (GO:0000287)
- oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor (GO:0016616)
- NAD binding (GO:0051287)

Cellular Component

None

PANTHER GO terms

Biological Process

- isocitrate metabolic process (GO:0006102)
- NADP metabolic process (GO:0006739)

Molecular Function

- isocitrate dehydrogenase (NADP+) activity (GO:0004450)

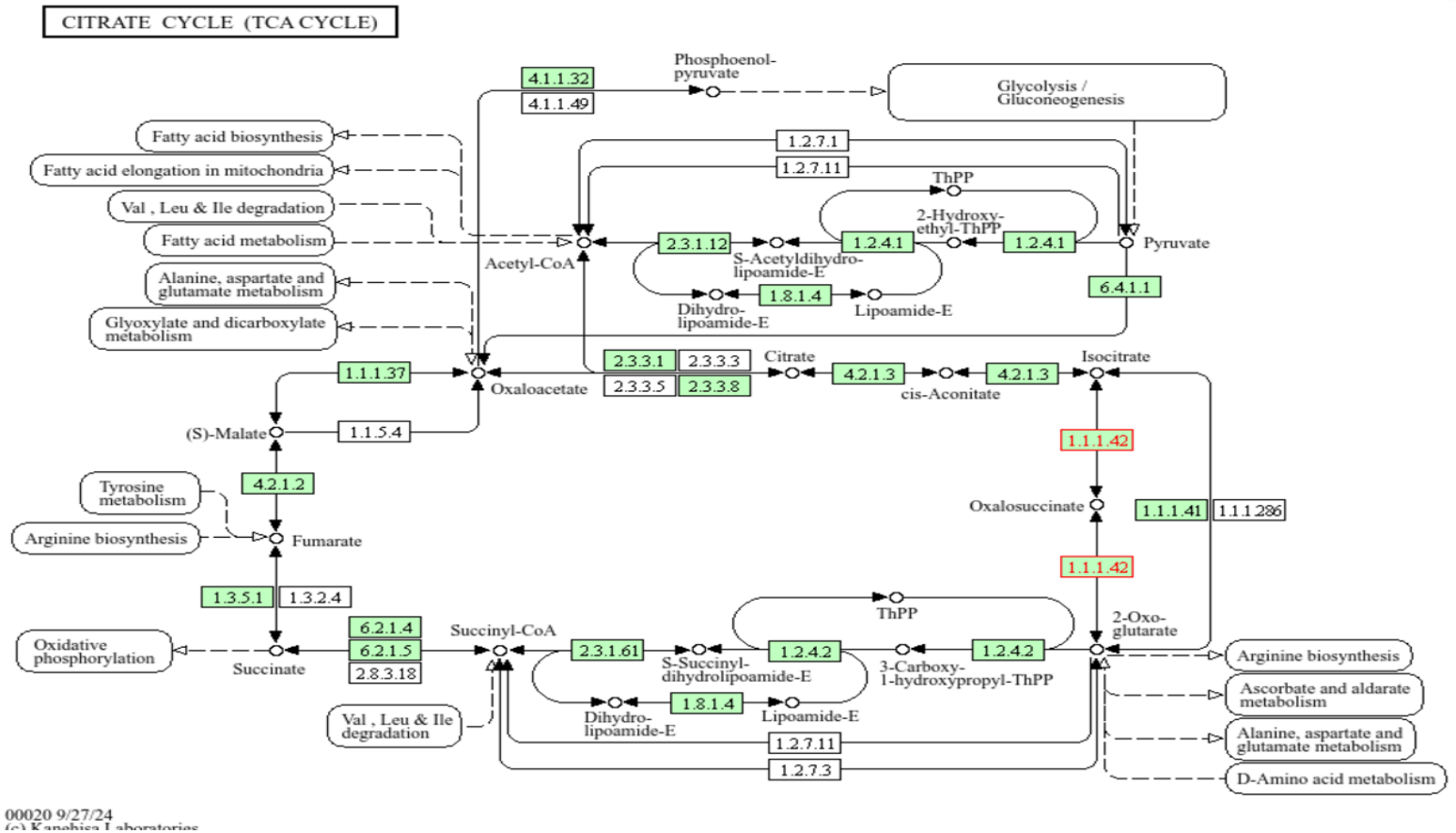
Cellular Component

- mitochondrion (GO:0005739)
- peroxisome (GO:0005777)
- cytosol (GO:0005829)

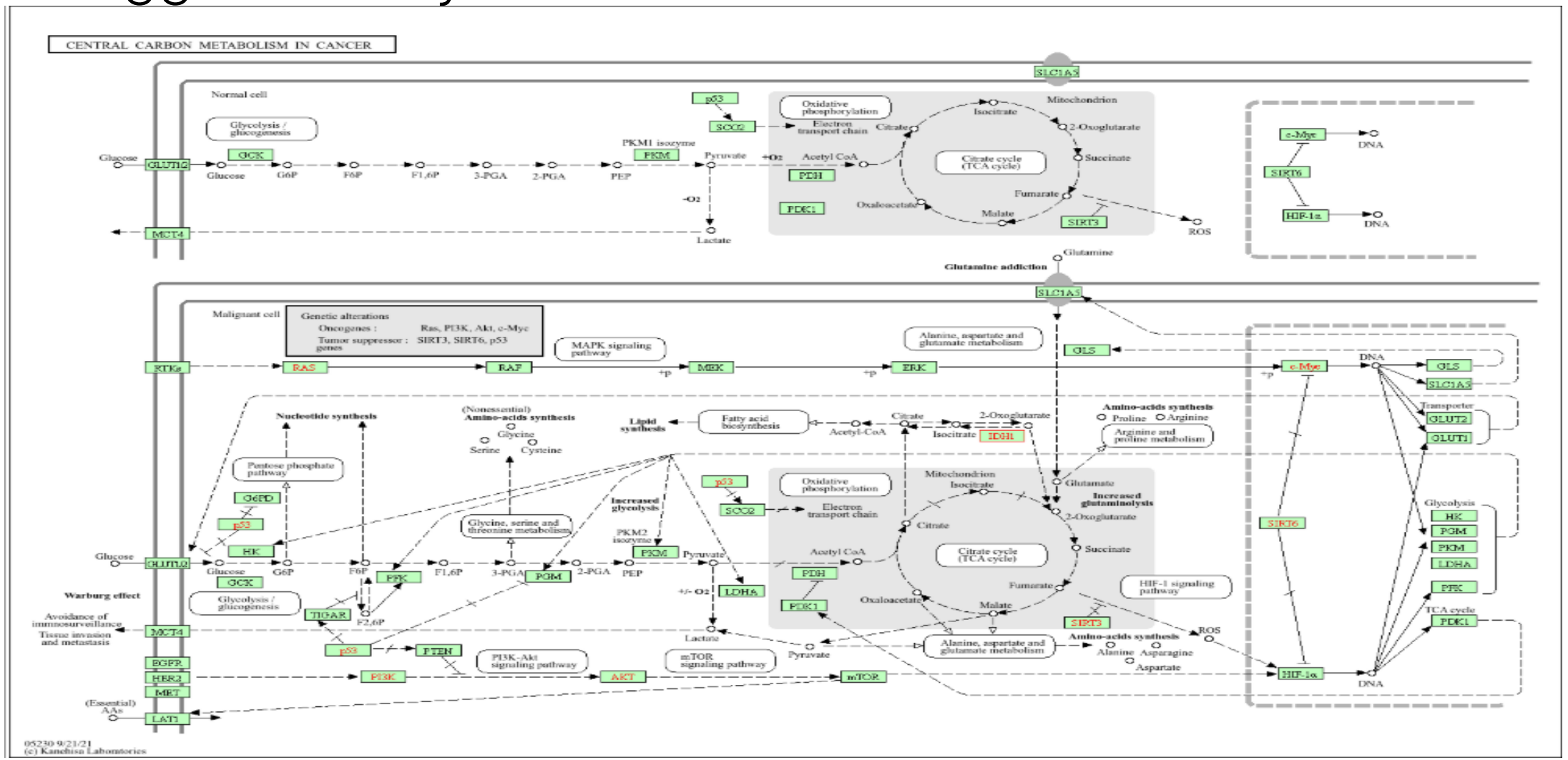
Kegg pathways

Pathway	bta00020	Citrate cycle (TCA cycle)
	bta00480	Glutathione metabolism
	bta01100	Metabolic pathways
	bta01200	Carbon metabolism
	bta01210	2-Oxocarboxylic acid metabolism
	bta01230	Biosynthesis of amino acids
	bta04146	Peroxisome
	bta05230	Central carbon metabolism in cancer

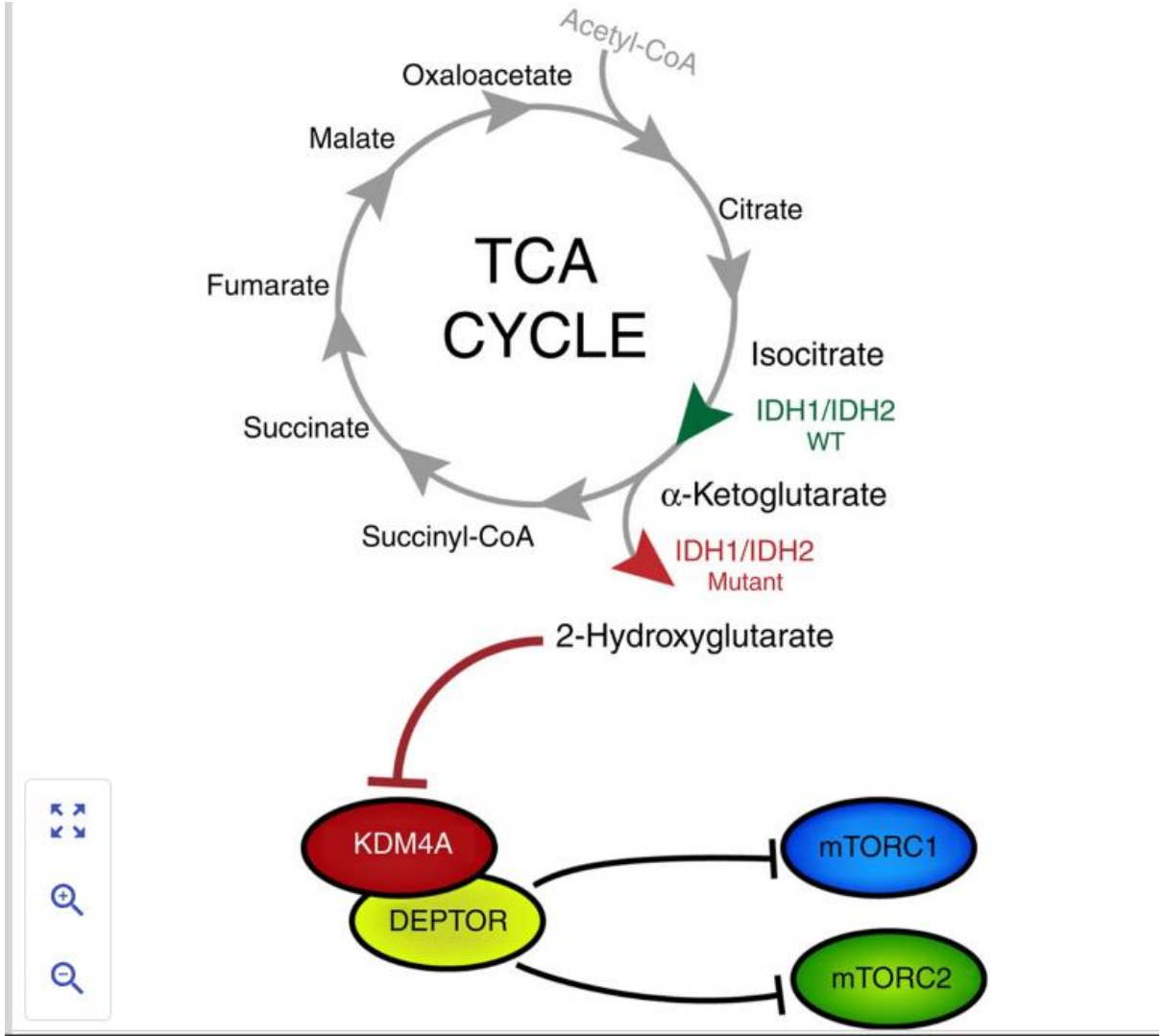
Kegg Pathway



Kegg Pathway



NDEx Pathway figure results



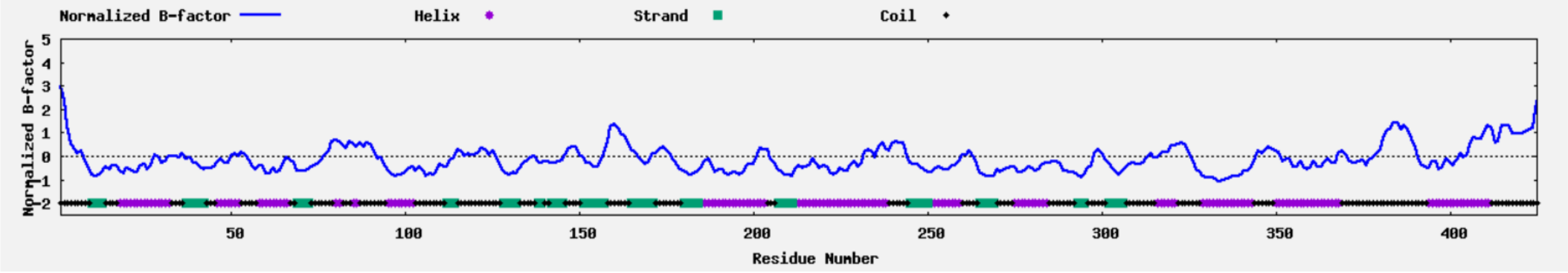
I-TASSER Results

Predicted Secondary Structure

```
                    20        40        60        80        100       120       140       160
                    |         |         |         |         |         |         |         |
Sequence    MSKKISGGSVVEMQGDEMTRIIWELIKEKLIFPYVELDLHSYDLGIENRDATNDQVTKDAAEAIKKHNVGVKCATITPDEKRVEEFKLKQMWKSPNGTIRNILGGTVFREAIICKNIPRLVSGWVKPIIIGHHAYGDQYRATDFVVPGPGKVEITYTPSD
Prediction  CCCCCCCCCSSSSCCCCHHHHHHHHHHHHHHHCCCCSSSSSSCCCHHHHHHHCCCCCHHHHHHHHHCCSSSSCCCCCCCHHCCCHHCCCCCCCCHHHHHHHHCCCCCCCCCSSSCCCCCCCCCCCCCSSSSSCCCCCSSCCSSSSCCCCCSSSSSSSCCC
Conf.Score  99753589879956887469999999999964689706999637566575218958699999998639564236669402101022026467782699998658710137775079998777887987999506788765624551789727999844784

            H:Helix; S:Strand; C:Coil
```

Predicted normalized B-factor

(B-factor is a value to indicate the extent of the inherent thermal mobility of residues/atoms in proteins. In I-TASSER, this value is deduced from threading template proteins from the PDB in combination with the sequence profiles derived from sequence databases. The reported B-factor profile in the figure below corresponds to the normalized B-factor of the target protein, defined by B=(B'-u)/s, where B' is the raw B-factor value, u and s are respectively the mean and standard deviation of the raw B-factors along the sequence. Click here to read more about predicted normalized B-factor)

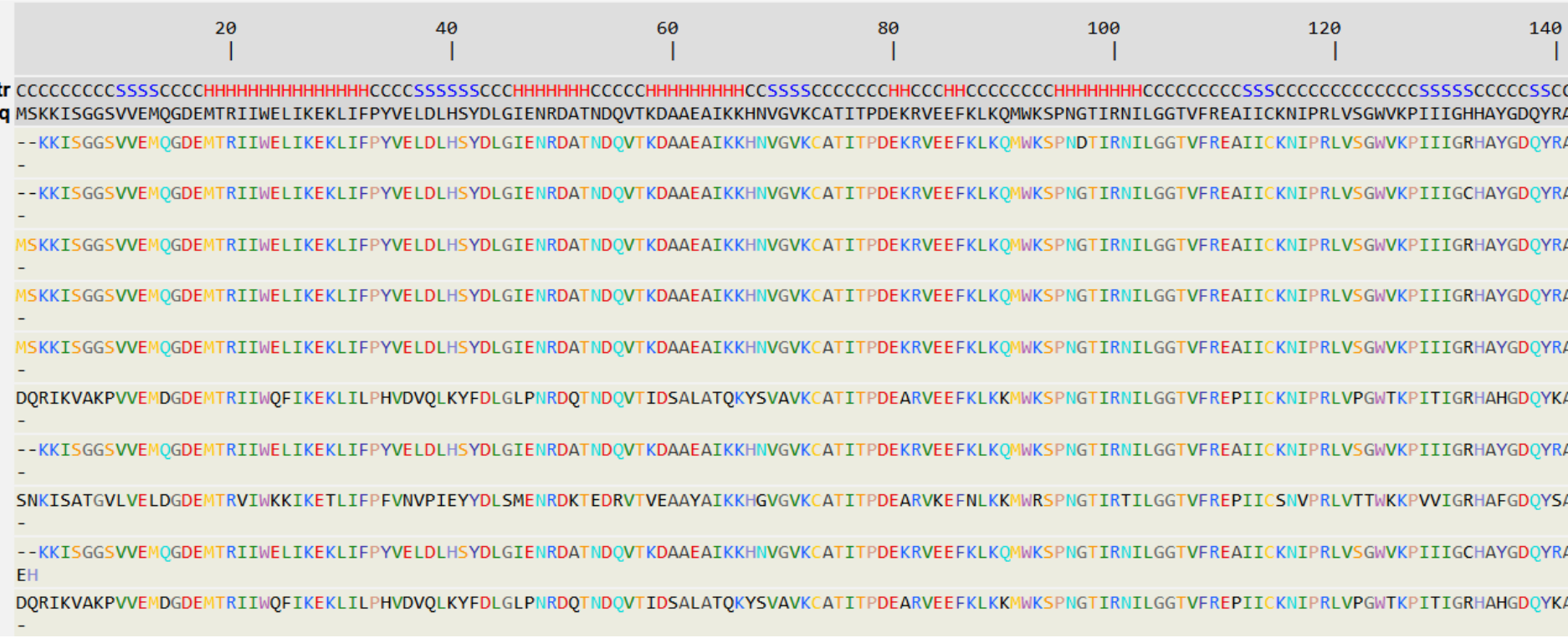


I-TASSER Results

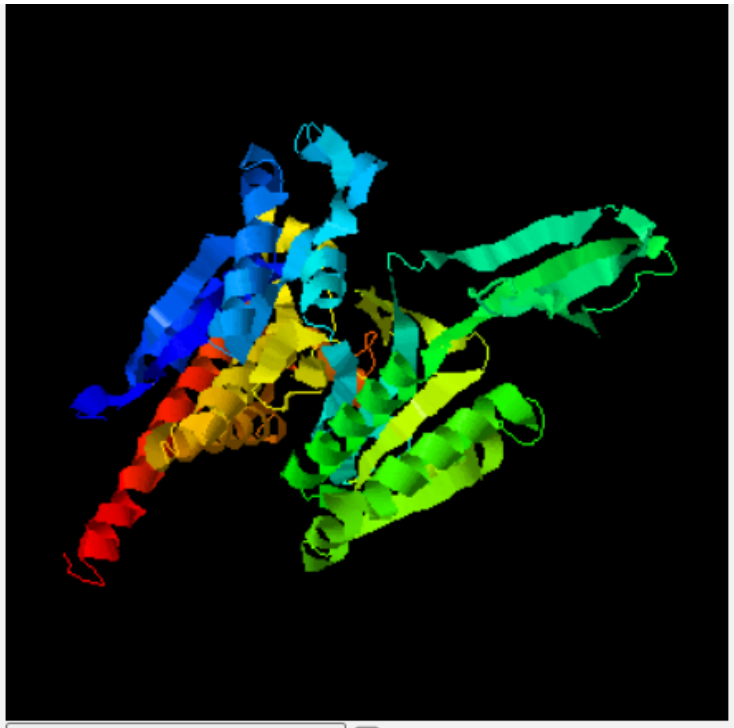
Top 10 threading templates used by I-TASSER

(I-TASSER modeling starts from the structure templates identified by LOMETS from the PDB library. LOMETS is a meta-server threading approach containing multiple threading programs, where each threading program can generate tens of thousands of templat
I-TASSER only uses the templates of the highest significance in the threading alignments, the significance of which are measured by the Z-score, i.e. the difference between the raw and average scores in the unit of standard deviation. The templates in this section a
templates selected from the LOMETS threading programs. Usually, one template of the highest Z-score is selected from each threading program, where the threading programs are sorted by the average performance in the large-scale benchmark test experiments.)

Rank	PDB Hit	Iden1	Iden2	Cov	Norm. Z-score	Download Align.
1	4l03A	1.00	0.98	0.98	5.97	Download
2	7pjmA	0.99	0.97	0.98	4.43	Download
3	1t09	1.00	0.97	0.97	1.85	Download
4	1t09	0.99	0.97	0.97	1.30	Download
5	1t09A	1.00	0.97	0.97	4.55	Download
6	1lwd	0.69	0.67	0.96	1.96	Download
7	6bkxA	1.00	0.97	0.97	8.41	Download
8	6aj6A	0.66	0.65	0.97	3.54	Download
9	7pjmA	0.99	0.97	0.98	4.21	Download
10	1lwdA	0.69	0.67	0.97	7.66	Download



I-TASSER Results – Top predicted structures



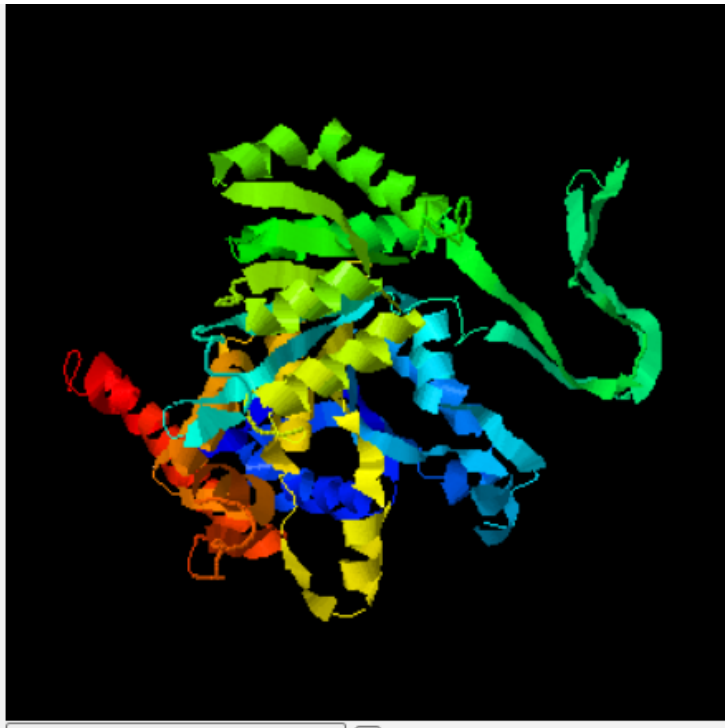
Reset to initial orientation ☑ Spin On/Off

[Download Model 1]

C-score=0.78 ([Read more about C-score])

Estimated TM-score = 0.82±0.08

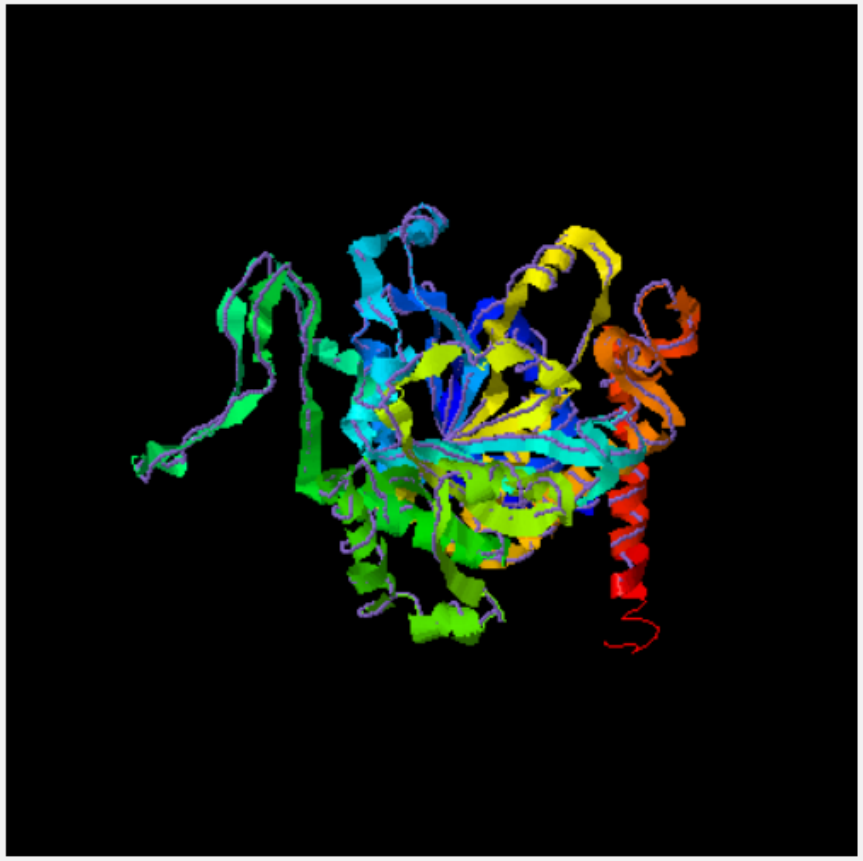
Estimated RMSD = 5.3±3.4Å



Reset to initial orientation ☐ Spin On/Off

- [Download Model 2]
- C-score = 0.70

I-TASSER Results



Top 10 Identified stuctural analogs in PDB

Click to view	Rank	PDB Hit	TM-score	RMSD[a]	IDEN[a]	Cov	Alignment
◉	1	1t09A	0.969	0.54	0.995	0.974	Download
○	2	6aj6A	0.921	2.06	0.643	0.967	Download
○	3	1zorA	0.880	2.30	0.516	0.934	Download
○	4	2qfwD	0.851	2.86	0.610	0.951	Download
○	5	1lwdA	0.831	3.28	0.649	0.955	Download
○	6	4aoyA	0.822	2.42	0.547	0.880	Download
○	7	3us8A	0.801	3.18	0.621	0.915	Download
○	8	2uxqA	0.799	3.46	0.476	0.934	Download
○	9	6lkzC	0.784	3.84	0.286	0.929	Download
○	10	2b0tA	0.778	3.46	0.121	0.896	Download

(a) Query structure is shown in cartoon, while the structural analog is displayed using backbone trace.
(b) Ranking of proteins is based on TM-score of the structural alignment between the query structure and known structures in the PDB library.
(c) RMSD[a] is the RMSD between residues that are structurally aligned by TM-align.
(d) IDEN[a] is the percentage sequence identity in the structurally aligned region.
(e) Cov represents the coverage of the alignment by TM-align and is equal to the number of structurally aligned residues divided by length of the query protein.

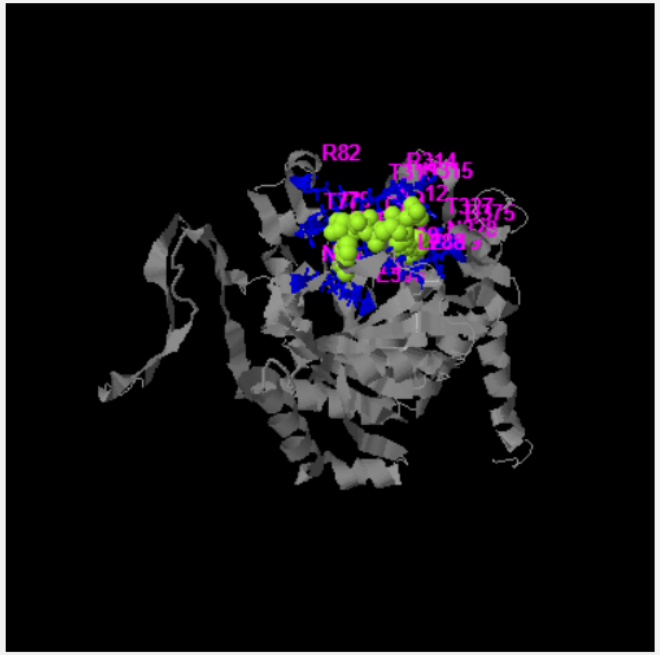
Reset to initial orientation ☐ Spin On/Off

I-TASSER Results

Predicted function using COFACTOR and COACH

(This section reports biological annotations of the target protein by COFACTOR and COACH based on the I-TASSER structure prediction. While COFACTOR deduces protein functions (ligand-binding sites, EC and GO) using structure comparison and protein-protein networks, COACH is a meta-server approach that combines multiple function annotation results (on ligand-binding sites) from the COFACTOR, TM-SITE and S-SITE programs.)

Ligand binding sites



Reset to initial orientation ☐ Spin On/Off

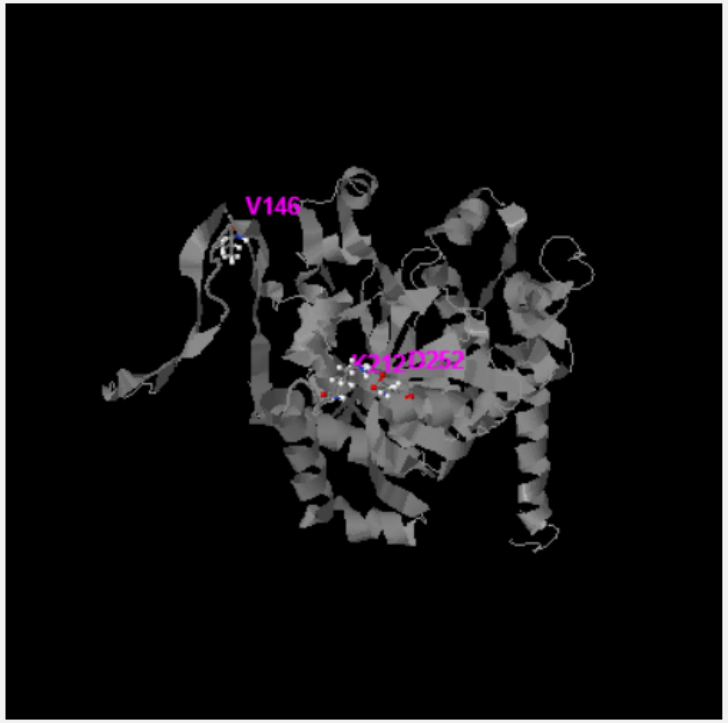
Click to view	Rank	C-score	Cluster size	PDB Hit	Lig Name	Download Complex	Ligand Binding Site Residues
○	1	0.94	166	3mapA	NAP	Rep, Mult	72,74,75,76,77,82,96,288,289,306,309,310,311,312,313,314,315,327,328,375
○	2	0.25	49	1xkdB	ICT	Rep, Mult	212,214,215,252
○	3	0.22	66	2qfwA	ICT	Rep, Mult	77,94,96,100,109,132,139
○	4	0.13	28	1isoA	NAD	Rep, Mult	309,310,311,312,314,327,328,375
○	5	0.09	29	2qfxE	CA	N/A	109,275,279,308

Download the residue-specific ligand binding probability, which is estimated by SVM.
Download the all possible binding ligands and detailed prediction summary.
Download the templates clustering results.

(a) **C-score** is the confidence score of the prediction. C-score ranges [0-1], where a higher score indicates a more reliable prediction.
(b) **Cluster size** is the total number of templates in a cluster.
(c) **Lig Name** is name of possible binding ligand. Click the name to view its information in the BioLiP database.
(d) **Rep** is a single complex structure with the most representative ligand in the cluster, i.e., the one listed in the **Lig Name** column.
Mult is the complex structures with all potential binding ligands in the cluster.

I-TASSER Results

Enzyme Commission (EC) numbers and active sites



Reset to initial orientation ☐ Spin On/Off

Click to view	Rank	Cscore^{EC}	PDB Hit	TM-score	RMSD^{a}	IDEN^{a}	Cov	EC Number	Active Site Residues
○	1	0.887	1t09A	0.969	0.54	0.995	0.974	1.1.1.42	146;212,252
○	2	0.812	1t09B	0.926	1.77	0.973	0.962	1.1.1.42	94,97,100,102,104,107,109,134,136,139,184,296,298
○	3	0.766	1lwdA	0.831	3.28	0.649	0.955	1.1.1.42	212,252
○	4	0.704	1zorA	0.880	2.30	0.516	0.934	1.1.1.42	145;212,252
○	5	0.665	2qfwD	0.851	2.86	0.610	0.951	1.1.1.42	296;212,252

Click on the radio buttons to visualize predicted active site residues.

(a) Cscore^{EC} is the confidence score for the EC number prediction. Cscore^{EC} values range in between [0-1]; where a higher score indicates a more reliable EC number prediction.

(b) TM-score is a measure of global structural similarity between query and template protein.

(c) RMSD^{a} is the RMSD between residues that are structurally aligned by TM-align.

(d) IDEN^{a} is the percentage sequence identity in the structurally aligned region.

(e) Cov represents the coverage of global structural alignment and is equal to the number of structurally aligned residues divided by length of the query protein.

I-TASSER Results

Gene Ontology (GO) terms

Top 10 homologous GO templates in PDB

Rank	Cscore^{GO}	TM-score	RMSD^{a}	IDEN^{a}	Cov	PDB Hit	Associated GO Terms
1	0.89	0.9690	0.54	0.99	0.97	1t09A	GO:0006749 GO:0005829 GO:0005625 GO:0050661 GO:0006102 GO:0005737 GO:0006103 GO:0005739 GO:0005782 GO:0042803 GO:0044255 GO:0055114 GO:0005777 GO:0000287 GO:0006740 GO:0046872 GO:0004450 GO:0006979 GO:0048545 GO:0008585 GO:0016491 GO:0006099 GO:0016616 GO:0006097 GO:0051287 GO:0005975 GO:0014070
2	0.75	0.9176	0.66	1.00	0.92	3marA	GO:0005737 GO:0006979 GO:0046872 GO:0008585 GO:0006103 GO:0016491 GO:0005777 GO:0006099 GO:0005739 GO:0016616 GO:0005829 GO:0006097 GO:0005782 GO:0051287 GO:0000287 GO:0005975 GO:0042803 GO:0014070 GO:0004450 GO:0006749 GO:0044255 GO:0005625 GO:0006740 GO:0050661 GO:0055114 GO:0006102 GO:0048545
3	0.71	0.8511	2.86	0.61	0.95	2qfwD	GO:0004450 GO:0042645 GO:0006537 GO:0006102 GO:0055114 GO:0046872 GO:0006097 GO:0016491 GO:0005739 GO:0005515 GO:0006099 GO:0000287 GO:0016616 GO:0051287
4	0.60	0.6710	3.13	0.18	0.76	1x0lA	GO:0000287 GO:0016616 GO:0051287 GO:0055114
5	0.58	0.6846	3.37	0.16	0.79	2y42D	GO:0008652 GO:0009082 GO:0000287 GO:0051287 GO:0016616 GO:0003862 GO:0055114 GO:0016491 GO:0009098 GO:0005737 GO:0046872
6	0.50	0.6864	3.47	0.15	0.80	3r8wA	GO:0000287 GO:0003862 GO:0005737 GO:0008652 GO:0009082 GO:0009098 GO:0009507 GO:0009536 GO:0009570 GO:0009941 GO:0016491 GO:0016616 GO:0046872 GO:0051287 GO:0055114
7	0.48	0.8796	2.30	0.52	0.93	1zorA	GO:0046872 GO:0006102 GO:0006099 GO:0016491 GO:0055114 GO:0004450
8	0.48	0.8314	3.28	0.65	0.96	1lwdA	GO:0006102 GO:0005739 GO:0004450 GO:0000287 GO:0006097 GO:0046872 GO:0016491 GO:0006099 GO:0006103 GO:0055114 GO:0016616 GO:0051287
9	0.39	0.7990	3.46	0.48	0.93	2uxqA	GO:0046872 GO:0004450 GO:0000287 GO:0016491 GO:0055114 GO:0006099 GO:0051287 GO:0016616 GO:0006102
10	0.38	0.6673	3.06	0.17	0.76	1wpwA	GO:0046872 GO:0009098 GO:0003862 GO:0009082 GO:0005737 GO:0008652 GO:0055114 GO:0016491 GO:0000287 GO:0016616 GO:0051287

Consensus prediction of GO terms

Molecular Function	GO:0000287	GO:0051287	GO:0004450	GO:0050661	GO:0042803	GO:0003862				
GO-Score	1.00	1.00	0.99	0.97	0.97	0.58				
Biological Process	GO:0006102	GO:0006097	GO:0006099	GO:0006979	GO:0006103	GO:0044255	GO:0006740	GO:0006749	GO:0048545	GO:0008585
GO-Score	0.99	0.99	0.99	0.97	0.97	0.97	0.97	0.97	0.97	0.97
Cellular Component	GO:0005829	GO:0005782	GO:0005625	GO:0042645						
GO-Score	0.97	0.97	0.97	0.71						

(a) Cscore^{GO} is a combined measure for evaluating global and local similarity between query and template protein. It's range is [0-1] and higher values indicate more confident predictions.

(b) TM-score is a measure of global structural similarity between query and template protein.

(c) RMSD^{a} is the RMSD between residues that are structurally aligned by TM-align.

(d) IDEN^{a} is the percentage sequence identity in the structurally aligned region.

(e) Cov represents the coverage of global structural alignment and is equal to the number of structurally aligned residues divided by length of the query protein.

(f) The second table shows a consensus GO terms amongst the top scoring templates. The GO-Score associated with each prediction is defined as the average weight of the GO term, where the weights are assigned based on Cscore^{GO} of the template.

Link: I-TASSER results

I-TASSER Results

Gene Ontology (GO) terms

Top 10 homologous GO templates in PDB

Rank	Cscore^{GO}	TM-score	RMSD^{a}	IDEN^{a}	Cov	PDB Hit	Associated GO Terms
1	0.89	0.9690	0.54	0.99	0.97	1t09A	GO:0006749 GO:0005829 GO:0005625 GO:0050661 GO:0006102 GO:0005737 GO:0006103 GO:0005739 GO:0005782 GO:0042803 GO:0044255 GO:0055114 GO:0005777 GO:0000287 GO:0006740 GO:0046872 GO:0004450 GO:0006979 GO:0048545 GO:0008585 GO:0016491 GO:0006099 GO:0016616 GO:0006097 GO:0051287 GO:0005975 GO:0014070
2	0.75	0.9176	0.66	1.00	0.92	3marA	GO:0005737 GO:0006979 GO:0046872 GO:0008585 GO:0006103 GO:0016491 GO:0005777 GO:0006099 GO:0005739 GO:0016616 GO:0005829 GO:0006097 GO:0005782 GO:0051287 GO:0000287 GO:0005975 GO:0042803 GO:0014070 GO:0004450 GO:0006749 GO:0044255 GO:0005625 GO:0006740 GO:0050661 GO:0055114 GO:0006102 GO:0048545
3	0.71	0.8511	2.86	0.61	0.95	2qfwD	GO:0004450 GO:0042645 GO:0006537 GO:0006102 GO:0055114 GO:0046872 GO:0006097 GO:0016491 GO:0005739 GO:0005515 GO:0006099 GO:0000287 GO:0016616 GO:0051287
4	0.60	0.6710	3.13	0.18	0.76	1x0lA	GO:0000287 GO:0016616 GO:0051287 GO:0055114
5	0.58	0.6846	3.37	0.16	0.79	2y42D	GO:0008652 GO:0009082 GO:0000287 GO:0051287 GO:0016616 GO:0003862 GO:0055114 GO:0016491 GO:0009098 GO:0005737 GO:0046872
6	0.50	0.6864	3.47	0.15	0.80	3r8wA	GO:0000287 GO:0003862 GO:0005737 GO:0008652 GO:0009082 GO:0009098 GO:0009507 GO:0009536 GO:0009570 GO:0009941 GO:0016491 GO:0016616 GO:0046872 GO:0051287 GO:0055114
7	0.48	0.8796	2.30	0.52	0.93	1zorA	GO:0046872 GO:0006102 GO:0006099 GO:0016491 GO:0055114 GO:0004450
8	0.48	0.8314	3.28	0.65	0.96	1lwdA	GO:0006102 GO:0005739 GO:0004450 GO:0000287 GO:0006097 GO:0046872 GO:0016491 GO:0006099 GO:0006103 GO:0055114 GO:0016616 GO:0051287
9	0.39	0.7990	3.46	0.48	0.93	2uxqA	GO:0046872 GO:0004450 GO:0000287 GO:0016491 GO:0055114 GO:0006099 GO:0051287 GO:0016616 GO:0006102
10	0.38	0.6673	3.06	0.17	0.76	1wpwA	GO:0046872 GO:0009098 GO:0003862 GO:0009082 GO:0005737 GO:0008652 GO:0055114 GO:0016491 GO:0000287 GO:0016616 GO:0051287

Consensus prediction of GO terms

Molecular Function	GO:0000287	GO:0051287	GO:0004450	GO:0050661	GO:0042803	GO:0003862				
GO-Score	1.00	1.00	0.99	0.97	0.97	0.58				
Biological Process	GO:0006102	GO:0006097	GO:0006099	GO:0006979	GO:0006103	GO:0044255	GO:0006740	GO:0006749	GO:0048545	GO:0008585
GO-Score	0.99	0.99	0.99	0.97	0.97	0.97	0.97	0.97	0.97	0.97
Cellular Component	GO:0005829	GO:0005782	GO:0005625	GO:0042645						
GO-Score	0.97	0.97	0.97	0.71						

(a) Cscore^{GO} is a combined measure for evaluating global and local similarity between query and template protein. It's range is [0-1] and higher values indicate more confident predictions.

(b) TM-score is a measure of global structural similarity between query and template protein.

(c) RMSD^{a} is the RMSD between residues that are structurally aligned by TM-align.

(d) IDEN^{a} is the percentage sequence identity in the structurally aligned region.

(e) Cov represents the coverage of global structural alignment and is equal to the number of structurally aligned residues divided by length of the query protein.

(f) The second table shows a consensus GO terms amongst the top scoring templates. The GO-Score associated with each prediction is defined as the average weight of the GO term, where the weights are assigned based on Cscore^{GO} of the template.

Link: I-TASSER results

PeptideCutter results

Name of enzyme	No. of cleavages	Positions of cleavage sites
Arg-C proteinase	14	20 49 82 100 109 119 140 222 249 314 317 338 343 388
Asp-N endopeptidase	25	15 37 42 49 53 58 78 136 142 159 185 190 219 224 231 251 252 272 274 278 298 346 374 389 398
Asp-N endopeptidase + N-terminal Glu	55	11 15 16 23 27 35 37 42 46 49 53 58 61 78 79 83 84 109 136 142 152 159 172 173 185 189 190 219 224 228 231 239 246 251 252 261 272 274 278 298 303 305 323 346 350 359 360 364 367 374 389 395 398 402 416
BNPS-Skatole	7	23 92 124 205 245 267 336
CNBr	14	1 13 18 91 180 182 199 254 259 290 291 318 372 398
Chymotrypsin-high specificity (C-term to [FYW], not before P)	39	23 34 42 86 92 108 124 135 139 144 156 167 172 183 192 197 208 219 223 227 231 235 239 245 246 265 267 272 285 316 319 334 336 354 355 371 391 395 397
Chymotrypsin-low specificity (C-term to [FYWML], not before P)	96	1 13 18 23 25 30 34 37 39 40 42 44 67 86 88 91 92 103 108 120 124 132 133 135 139 144 156 167 168 170 172 180 183 192 194 197 199 201 207 208 209 216 219 223 227 231 235 239 245 246 248 250 254 259 265 267 272 285 288 290 291 295 309 315 316 319 334 336 340 342 346 352 354 355 359 371 372 376 391 392 395 397 398 401 405 409 414 416 418 419 420 421 422 423 424 425
Clostripain	14	20 49 82 100 109 119 140 222 249 314 317 338 343 388
Formic acid	25	16 38 43 50 54 59 79 137 143 160 186 191 220 225 232 252 253 273 275 279 299 347 375 390 399
Glutamyl endopeptidase	30	12 17 24 28 36 47 62 80 84 85 110 153 173 174 190 229 240 247 262 304 306 324 351 360 361 365 368 396 403 417
Hydroxylamine	1	96
Iodosobenzoic acid	7	23 92 124 205 245 267 336
LysC	37	3 4 27 29 58 65 66 72 81 87 89 93 115 126 151 164 187 203 212 217 218 224 233 236 243 260 270 301 321 345 350 374 381 400 406 408 413
LysN	37	2 3 26 28 57 64 65 71 80 86 88 92 114 125 150 163 186 202 211 216 217 223 232 235 242 259 269 300 320 344 349 373 380 399 405 407 412
NTCB (2-nitro-5-thiocyanobenzoic acid)	5	72 113 268 296 378
Pepsin (pH1.3)	69	24 25 30 32 36 37 38 39 43 44 85 86 87 88 103 107 108 120 143 144 167 168 171 191 192 197 200 201 206 208 209 215 216 223 227 239 264 265 287 288 294 295 333 334 339 346 351 353 354 355 358 359 370 371 375 391 392 394 395 396 397 400 401 404 405 409 413 414 416
Pepsin (pH>2)	95	23 24 25 30 32 33 36 37 38 39 41 43 44 85 86 87 88 92 103 107 108 120 123 124 138 139 143 144 155 167 168 171 182 183 191 192 197 200 201 206 208 209 215 216 218 223 227 230 231 234 239 244 246 264 265 266 267 271 284 285 287 288 294 295 315 318 333 334 335 336 339 346 351 353 354 355 358 359 370 371 375 391 392 394 395 396 397 400 401 404 405 409 413 414 416
Proline-endopeptidase [*]	1	127
Proteinase K	209	5 10 11 12 17 19 21 22 23 24 25 26 28 30 31 32 34 35 36 37 39 42 44 46 47 51 52 56 57 60 61 62 63 64 69 71 74 75 76 77 80 83 84 85 86 88 92 98 99 102 103 106 107 108 110 111 112 113 117 120 121 124 125 128 129 130 134 135 139 141 142 144 145 146 152 153 154 155 156 157 162 165 166 167 168 169 172 173 174 178 179 183 189 190 192 193 197 200 201 205 207 208 209 211 214 215 216 219 223 226 227 229 230 231 235 239 240 241 244 245 246 247 250 251 255 256 258 262 265 266 267 268 272 276 281 282 285 288 292 294 295 296 302 303 304 305 306 307 308 311 312 313 316 319 324 325 327 330 331 333 334 335 336 337 340 341 344 346 351 352 353 354 355 356 358 359 360 361 362 364 365 366 367 368 369 371 373 376 377 378 380 383 386 391 392 394 395 396 397 401 403 405 407 409 410 412 414 416 417
Staphylococcal peptidase I	27	12 17 24 28 36 47 62 80 84 110 153 173 190 229 240 247 262 304 306 324 351 360 365 368 396 403 417
Thermolysin	114	4 9 10 20 21 25 29 30 34 45 55 60 63 68 70 73 75 82 87 90 98 101 102 106 107 111 112 119 120 124 127 128 129 133 140 144 151 164 167 168 171 177 178 179 181 188 192 196 198 199 200 206 208 214 215 222 226 238 243 249 250 254 255 257 258 264 265 267 280 281 287 289 290 293 294 295 302 307 311 317 329 330 332 333 334 339 340 343 345 352 353 354 355 357 358 363 366 370 371 376 377 379 385 391 394 397 400 404 406 408 409 411 413 415
Trypsin	50	3 4 20 27 29 49 58 65 66 72 81 82 87 89 93 100 109 115 119 140 151 164 187 203 212 217 218 222 224 233 236 243 249 260 270 301 314 317 321 338 343 345 350 374 381 388 400 406 408 413

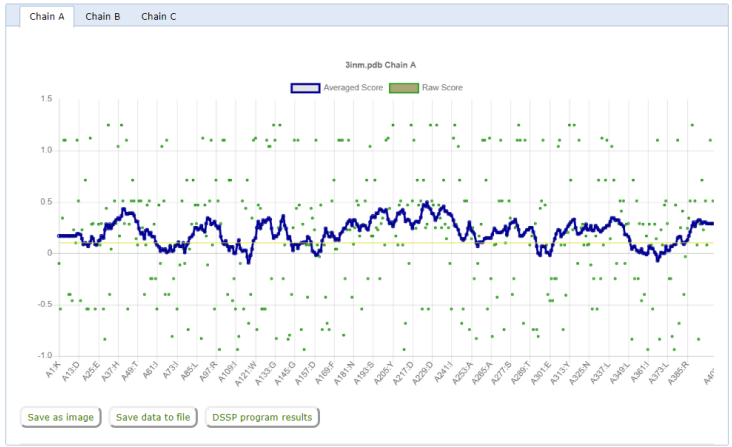
The selected enzymes do not cut:

Caspase1
Caspase10
Caspase2
Caspase3

Verify3D results

VERIFY3D
78.53% of the residues have averaged 3D-1D score >= 0.1
Fail

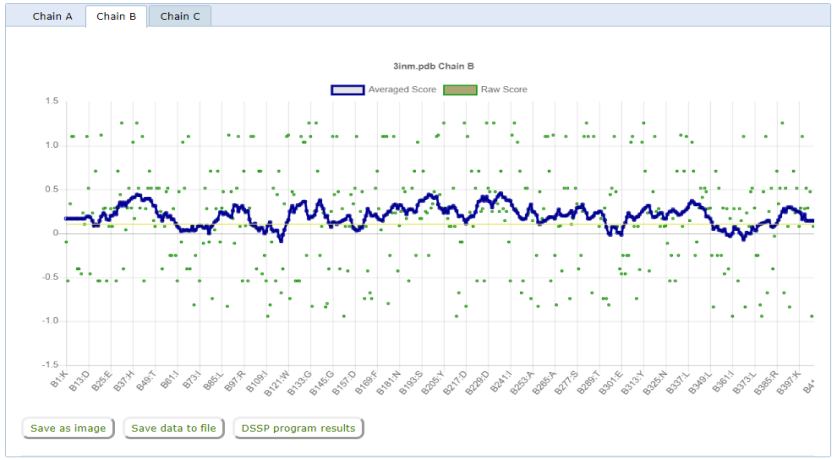
Fewer than 80% of the amino acids have scored >= 0.1 in the 3D/1D profile.



Complete. Time taken: 00:00:18

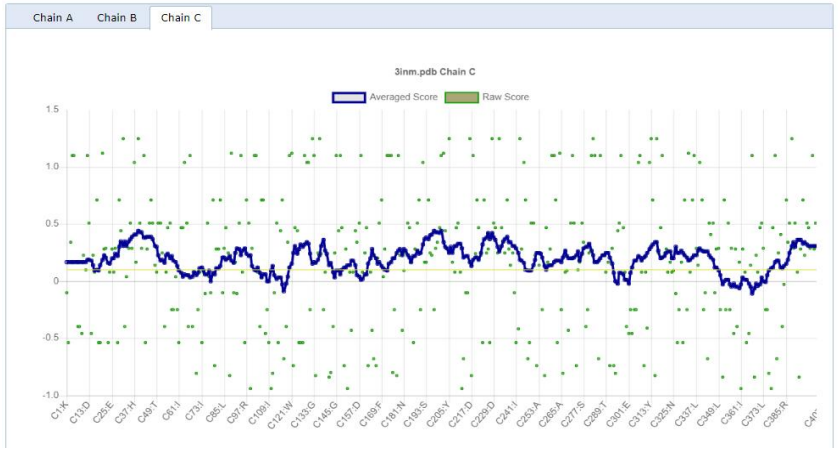
VERIFY3D
78.53% of the residues have averaged 3D-1D score >= 0.1
Fail

Fewer than 80% of the amino acids have scored >= 0.1 in the 3D/1D profile.

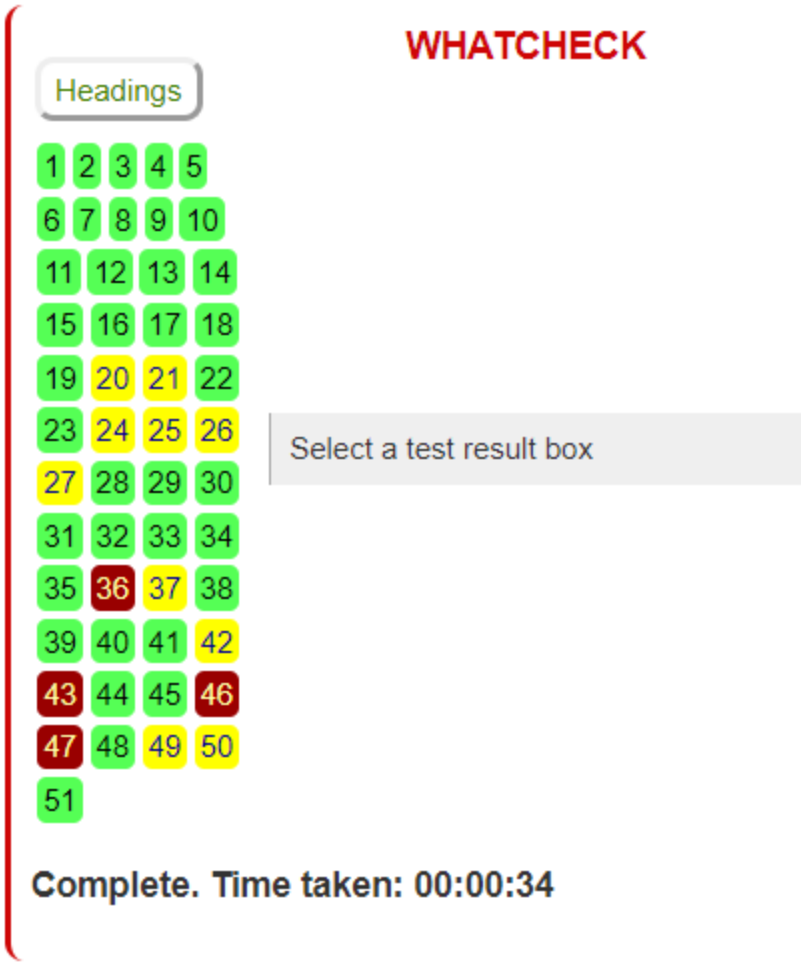


VERIFY3D
78.53% of the residues have averaged 3D-1D score >= 0.1
Fail

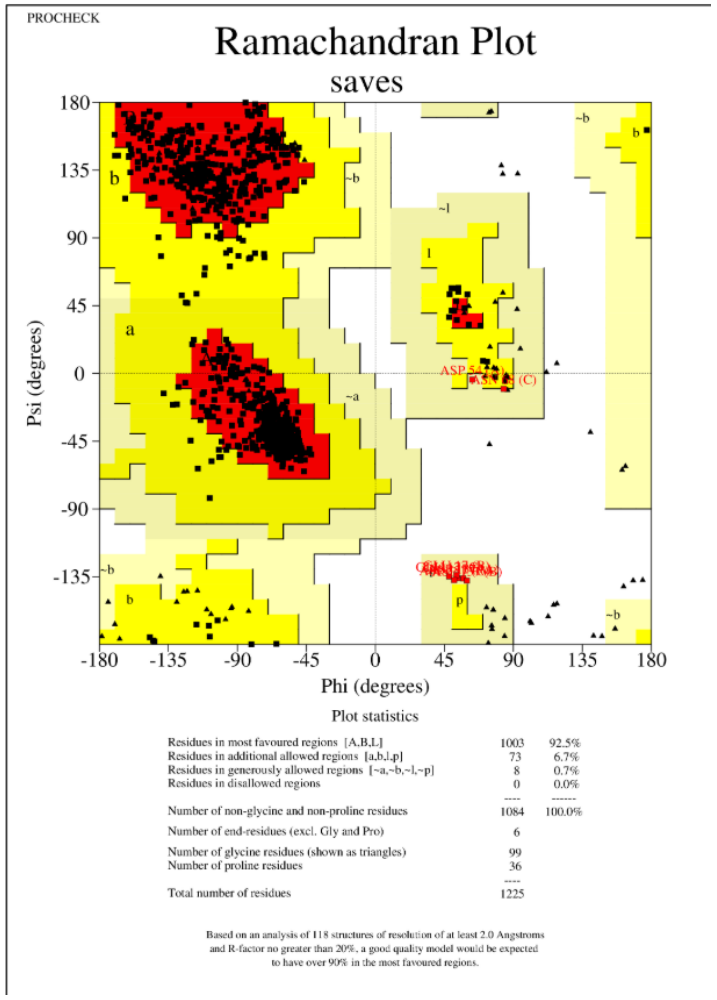
Fewer than 80% of the amino acids have scored >= 0.1 in the 3D/1D profile.



What check results



ProCheck Ramachandran plot results



Plot statistics

Residues in most favoured regions [A,B,L]	1003	92.5%
Residues in additional allowed regions [a,b,l,p]	73	6.7%
Residues in generously allowed regions [~a,~b,~l,~p]	8	0.7%
Residues in disallowed regions	0	0.0%
Number of non-glycine and non-proline residues	1084	100.0%
Number of end-residues (excl. Gly and Pro)	6	
Number of glycine residues (shown as triangles)	99	
Number of proline residues	36	
Total number of residues	1225	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

References

- “IDH1 isocitrate dehydrogenase (NADP(+)) 1 [Homo sapiens (human)] - Gene - NCBI,” *www.ncbi.nlm.nih.gov*. https://www.ncbi.nlm.nih.gov/gene/3417
- Lennard *et al.*, “Effects of the IDH1 R132H Mutation on the Energy Metabolism: A Comparison between Tissue and Corresponding Primary Glioma Cell Cultures,” *ACS omega*, vol. 7, no. 4, pp. 3568–3578, Jan. 2022, doi: https://doi.org/10.1021/acsomega.1c06121.
- S. Najafi, S. Esmaeili, Hossein Zhaleh, and Y. Rahmati, “The role of IDH1 mutation on gene expression in glioblastoma,” *Informatics in Medicine Unlocked*, vol. 28, pp. 100812–100812, Jan. 2022, doi: https://doi.org/10.1016/j.imu.2021.100812.
- Wikipedia Contributors, “IDH1,” *Wikipedia*, Aug. 08, 2024. https://en.wikipedia.org/wiki/IDH1

Thank you