

# 陆地棉MIPS蛋白生物信息学分析

## The Basic Bioinformatic Analysis of MIPS in *Gossypium hirsutum*

---



报告人：陈佩双

组 员：张轶伟 邢悦婷 常子焯

2015-06-27

# Content

---

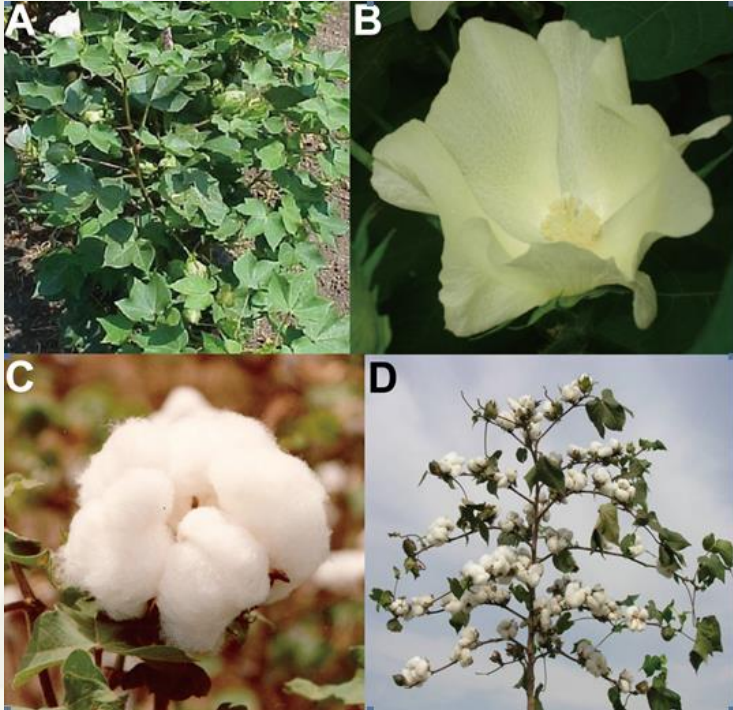
- \* **Background**
- \* **Sequence analysis and phylogeny construction**
- \* **Motif prediction**
- \* **3D Structure prediction**

---



# Background

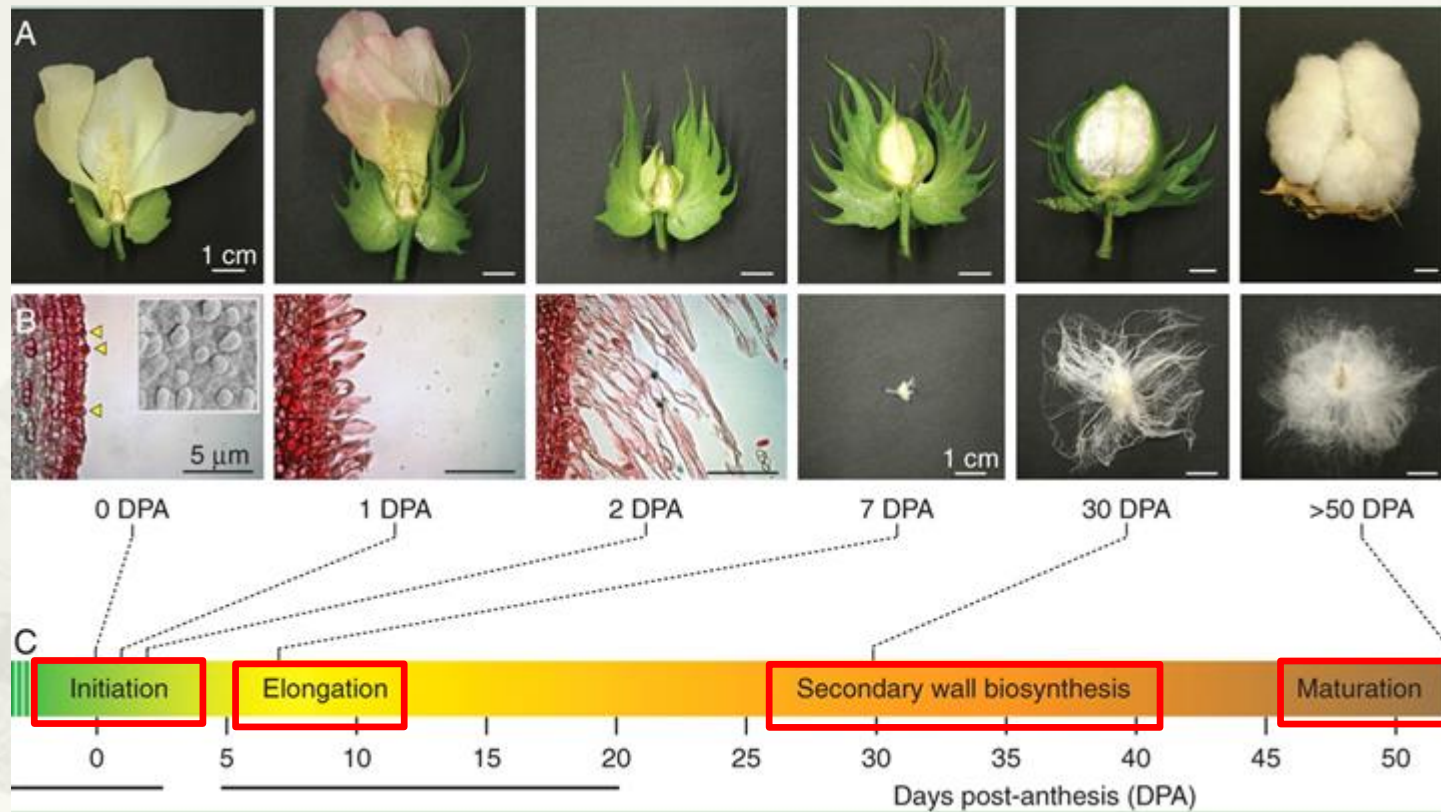
# Upland Cotton



<http://icgr.caas.net.cn>

- \* The cotton fiber is a single-celled trichome originating from seed coat and is the major natural material used in the modern textile industry.
- \* Belong to *Malvaceae*, *Gossypium*.
- \* N=13, diploid and tetraploid; A、 B、 C、 E、 F、 G、 K.
- \* Upland cotton (*Gossypium hirsutum*) is tetraploid combined by D-genome (*Gossypium raimondi*) and A-genome (*Gossypium arboreum*), and genome sequence of these three species were sequenced.

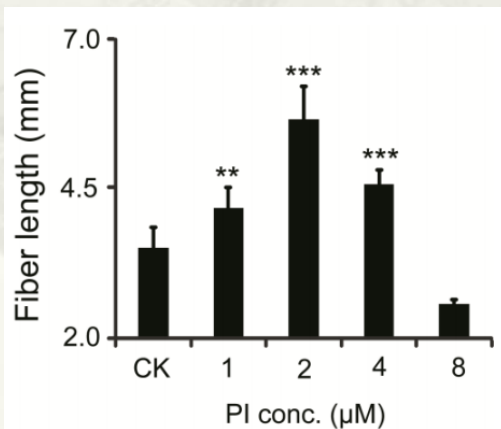
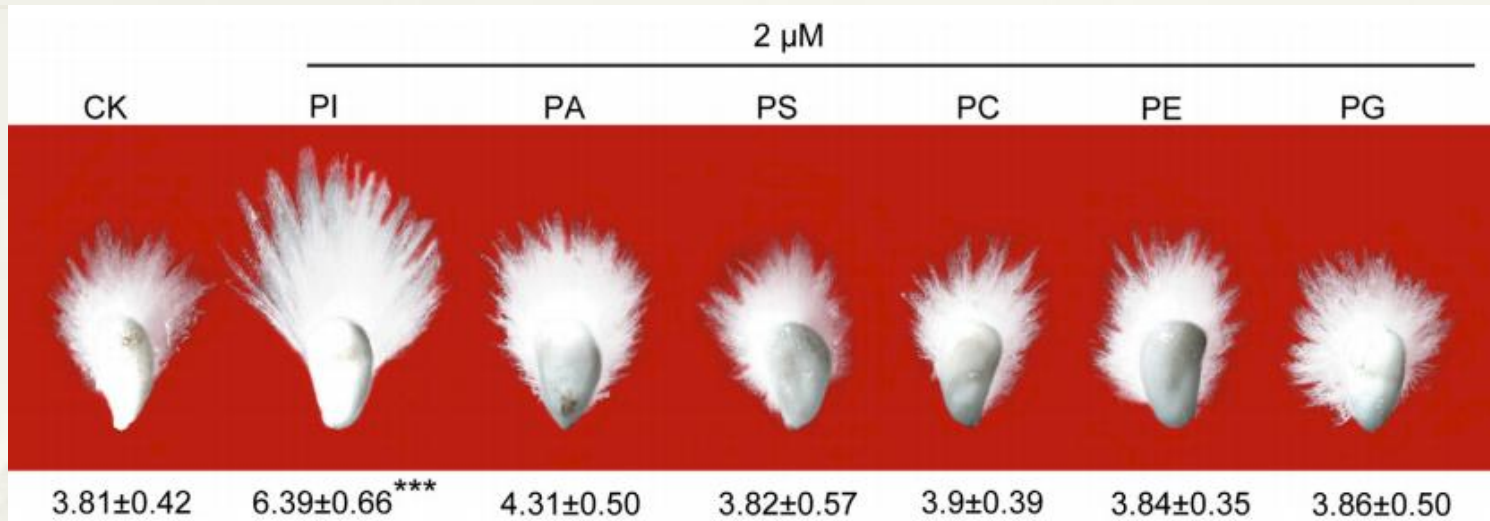
# Development of cotton fiber



(Lee *et al.*, 2007)

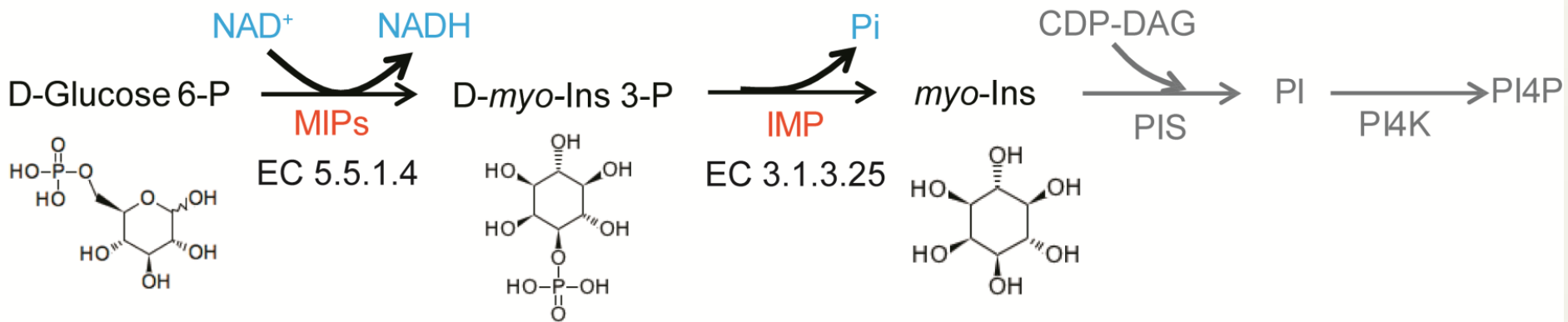


# Latest data from our lab



Phosphatidylinositol(PI) promote cotton fiber elongation in ovule culture.

# The *myo*-inositol biosynthesis pathway (Loewus pathway)



- \* EC 5.5.1.4 D-*myo*-inositol 3-phosphate (or L-*myo*-inositol 1-phosphate) synthase (**MIPS**)
- \* EC 3.1.3.25 *myo*-inositol monophosphatase (**IMP**)

# About MIPS

- \* The rate-limiting step of myo-inositol synthesis is catalyzed by MIPS.
- \* Yeast and animal genomes contain a single gene encoding MIPS, plants have multiple *MIPS* genes.
- \* In plant

**The *Arabidopsis thaliana* Myo-Inositol 1-Phosphate Synthase1 Gene Is Required for Myo-inositol Synthesis and Suppression of Cell Death** <sup>W</sup>

**D-*myo*-Inositol-3-Phosphate Affects Phosphatidylinositol-Mediated Endomembrane Function in *Arabidopsis* and Is Essential for Auxin-Regulated Embryogenesis** <sup>W|OA</sup>



---

**Sequence analysis  
&  
phylogeny construction**

The background features a large, light-colored fan shape. Inside the fan, there is a faint, grey-scale phylogenetic tree with many branches, representing evolutionary relationships. The text is centered over the fan.

# Basic annotation information of GhMIPS



## D2D2Z4 - D2D2Z4\_GOSHI

**Protein** | Submitted name: **Myo-inositol-1-phosphate synthase**

**Gene** | N/A

**Organism** | *Gossypium hirsutum* (Upland cotton) (*Gossypium mexicanum*)

**Status** | Unreviewed - Annotation score: ●○○○○○ - Experimental evidence at transcript level<sup>2</sup>

```
>tr|D2D2Z4|D2D2Z4_GOSHI Myo-inositol-1-phosphate synthase OS=Gossypium hirsutum PE=2 SV=1
MFIENFKVESPNVKYTENEIQSVYNYETTELWHENKNGTYQWVVKPKTVKYEFKTDIHVP
KLGVMVLVGWGGNNGSTLTGGVIANKEGISWATKDKVQQANYFGSLTQASTIRIGSYNGEE
IYAPFKSLLPMVNPNDIVFGGWDISDMNLADAMARAKVFDIDLQQLRPYMESMVPLPGI
YDPDFIAANQGERANNVIKGTKKEQVQQVIKDIKEFKEKNKVDKVVVLWTANTERYSNVI
VGLNDTVESLMASLEKNESEISPSSTLYAIACVLENVFFINGSPQNTFVPGIDLAIQRNC
LIGGDDFKSGQTKMKSVLVDFLVGAGIKPTSIVSYNHLGNNDGMNLSAPQTFRSKEISKS
NVVDDMVSSNGILYEPGEHPDHVVVIKYVPYVGD SKRAMDEYTSEIFMGGKNTIVLHNTC
EDSLLAAPIIILDLVLLAELSTRIQFKADGEGKFHSFHPVATILSYLTKAPLVPPGTPVVN
ALSKQRAMLENILRASIGLAPENNMILEYK
```

510 aa

- Function
- Names & Taxonomy
- Subcellular location
- Pathology & Biotech
- PTM / Processing
- Expression
- Interaction
- Structure
- Family & Domains
- Sequence
- Cross-references
- Publications
- Entry information
- Miscellaneous
- Similar proteins

# Sequence analysis of MIPS from different species



```
Q94C02 Q94C02_SOYBN 1 -----MFLENFKVCEPNVKYITETIHSQVYVYETTELWHEH-RNGTYQWVVKPKTKVKEF
P42801 INO1_ARATH 1 -----MFLESFKVESPVNVKITEHEIHSVDYVETTEWVHEKTVNGTYQWVVKPKTKVKEF
Q38862 INO2_ARATH 1 -----MFLESFKVESPVNVKITEHEIHSVDYVETTEWVHEH-RNGTYQWVVKPKTKVKEF
P11986 INO1_YEAST 1 MTEDNAIPITTSVKVWTRDCTVKNDELLTKYSYENAVVT---KTASGRFDVTPVQDVVF
D2D2Z4 D2D2Z4_GOSHI 1 -----MFLENFKVESPVNVKITEHEIHSQVYVYETTELWHEH-KNGTYQWVVKPKTKVKEF
064437 INO1_ORYSJ 1 -----MFLESFKVESPVNVKITEHEIHSVDYVETTEWVHEH-HDGASRWVVRHSVRYNF
Q9S7U0 INO1_WHEAT 1 -----MFLESFKVESPVNVKITEHEIHSVDYVETTEWVHEH-RNGTYQWVVKPKTKVKEF
Q9LX12 INO3_ARATH 1 -----MFLESFKVESPVNVKITEHEIHSVDYVETTELWHEH-KNGAFQHTVVKPKTKVKEF
* . . . * . * * : * . : . . . : . : . : * . * *
```

```
Q94C02 Q94C02_SOYBN 54 KTIHIVP-KLGVMLVGGGNGSILTGGVIANREGISWATKDKVQQANYFGSLTQASSIR
P42801 INO1_ARATH 55 KTDIRVP-KLGVMLVGLGGNGSILTAGVIANKEGISWATKDKVQQANYFGSLTQASSIR
Q38862 INO2_ARATH 54 KTDTRVP-KLGVMLVGGGNGSILTAGVIANKEGISWATKDKVQQANYFGSLTQASSIR
P11986 INO1_YEAST 57 KLDLKKPEKLGIMLILGLGGNGSILTAVSILANKHVEFQTKGVRQPNYFGSMTCSTLK
D2D2Z4 D2D2Z4_GOSHI 54 KTDIHPV-KLGVMLVGGGNGSILTGGVIANKEGISWATKDKVQQANYFGSLTQASTIR
064437 INO1_ORYSJ 54 RTTTTVP-KLGVMLVGGGNGSILTAGVIANREGISWATKDKVQQANYFGSLTQASTIR
Q9S7U0 INO1_WHEAT 54 KTDTRVP-KLGVMLVGGGNGSILTAGVIANKEGISWATKDKVQQANYFGSLTQASSIR
Q9LX12 INO3_ARATH 54 KTDTHVP-KLGVMLVGGGNGSILTAGVIANREGISWATKEKVVQANYFGSLTQASSIR
: * **:* * * * * * . . * * : . : * * : * * * * * * . . . :
```

```
Q94C02 Q94C02_SOYBN 113 VGS-FNGEEIYAPFKSLLPMVNPDDVFGGWDISDMNLADAMARARVLDIDLQKRLRPYM
P42801 INO1_ARATH 114 VGS-FNGEEIYAPFKSLLPMVNPDDVFGGWDISDMNLADAMARARVLDIDLQKRLRPYM
Q38862 INO2_ARATH 113 VGS-YNGEEIYAPFKSLLPMVNPDDVFGGWDISDMNLADAMARARVLDIDLQKRLRPYM
P11986 INO1_YEAST 117 LGIDAEAGNDVYAPFNSLLPMVNPDDVFGGWDISDMNLADAMARARVLDIDLQKRLRPYM
D2D2Z4 D2D2Z4_GOSHI 113 VGS-YNGEEIYAPFKSLLPMVNPDDVFGGWDISDMNLADAMARARVLDIDLQKRLRPYM
064437 INO1_ORYSJ 113 VGS-YNGEEIYAPFKSLLPMVNPDDVFGGWDISDMNLADAMARARVLDIDLQKRLRPYM
Q9S7U0 INO1_WHEAT 113 VGS-FNGEEIYAPFKSLLPMVNPDDVFGGWDISDMNLADAMARARVLDIDLQKRLRPYM
Q9LX12 INO3_ARATH 113 VGS-FNGEEIYAPFKSLLPMVNPDDVFGGWDISDMNLADAMARARVLDIDLQKRLRPYM
: * * : : : * * * * * * . . : * * * : : * * * : : * * * * * * . . . :
```

```
Q94C02 Q94C02_SOYBN 172 ESMVPLPGLYDPDFIAANQGRANNVKIGTKK-----EQVQVHIDIREFKKPKNKV
P42801 INO1_ARATH 173 ENIVPLPGLYDPDFIAANQGRANNVKIGTKK-----EQVDHIIKDMREFKPKNKV
Q38862 INO2_ARATH 172 ENMIVPLPGLYDPDFIAANQGRANNVKIGTKK-----EQVDHIIKDMREFKPKNKV
P11986 INO1_YEAST 177 SLVKPLPSIYYPDFIAANQGRANNCINLDEKGNVTRGKWHLQRIRRDIDQNFKEENAL
D2D2Z4 D2D2Z4_GOSHI 172 ESMVPLPGLYDPDFIAANQGRANNVKIGTKK-----EQVQVHIDIREFKKPKNKV
064437 INO1_ORYSJ 172 ESMVPLPGLYDPDFIAANQGRANNVKIGTKK-----EQMEQIHKDIREFKKPKNKV
Q9S7U0 INO1_WHEAT 172 EHMVPLPGLYDPDFIAANQGRANNVKIGTKK-----EQVDHIIKDMREFKPKNKV
Q9LX12 INO3_ARATH 172 EHMVPLPGLYDPDFIAANQGRANNVKIGTKK-----QQLGQVHKDIREFKKPKNKV
: . : * * * * * : * * * * * * . . : . . : : : : : * * * * * . . . :
```

```
Q94C02 Q94C02_SOYBN 223 DKVVLWTANTERYSNLVVGLNDTMENTLMSLDKDEAEISFSTLYAIACVLENVFPINGS
P42801 INO1_ARATH 224 DKVVLWTANTERYSNLVVGLNDTMENTLMSVDRDEAEISFSTLYAIACVLEGPINGS
Q38862 INO2_ARATH 223 DKLVVLWTANTERYSNLVVGLNDTMENTLMSVDRDEAEISFSTLYAIACVLEGPINGS
P11986 INO1_YEAST 237 DKVVLWTANTERYSNLVVGLNDTMENTLMSLDKDEAEISFSTLYAIACVLENVFPINGS
D2D2Z4 D2D2Z4_GOSHI 223 DKVVLWTANTERYSNLVVGLNDTMENTLMSLDKDEAEISFSTLYAIACVLENVFPINGS
064437 INO1_ORYSJ 223 DKVVLWTANTERYSNLVVGLNDTMENTLMSLDKDEAEISFSTLYAIACVLEGPINGS
Q9S7U0 INO1_WHEAT 223 DKLVVLWTANTERYSNLVVGLNDTMENTLMSLDKDEAEISFSTLYAIACVLEGPINGS
Q9LX12 INO3_ARATH 223 DKVVLWTANTERYSNLVVGLNDTMENTLMSLDKDEAEISFSTLYAIACVLENVFPINGS
* * : * * * * * * * : * * * * * * . . : * * * * * * * * : * * * * * *
```

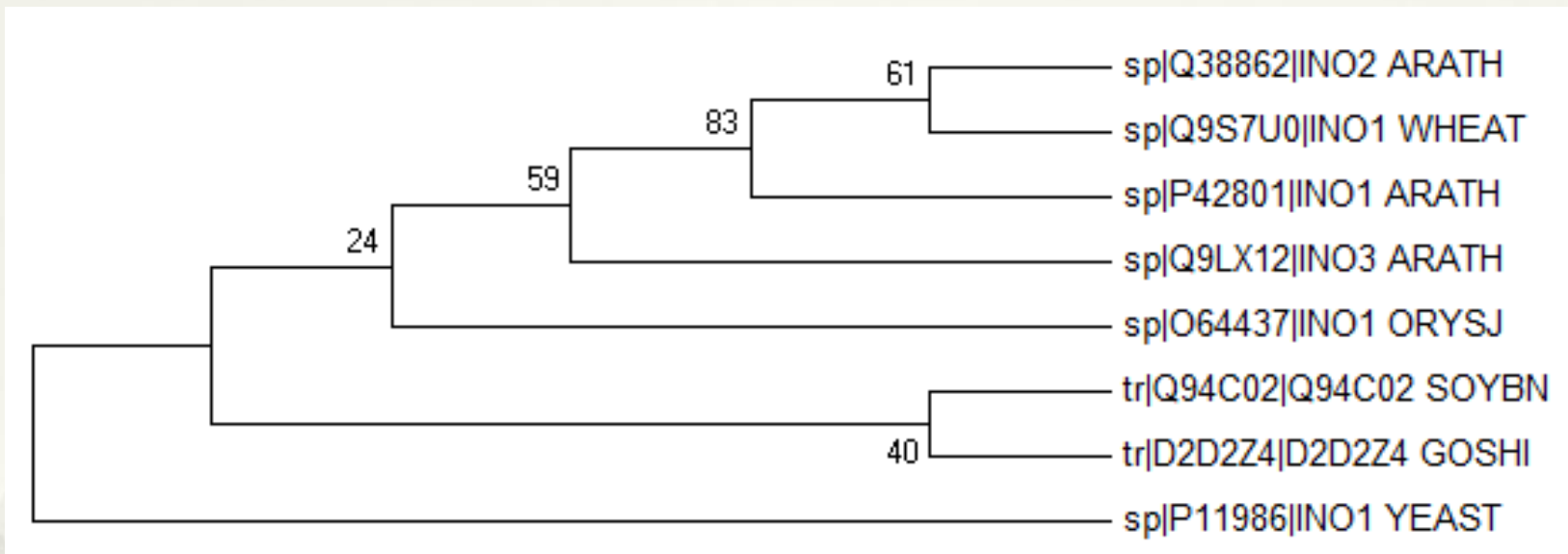
```
Q94C02 Q94C02_SOYBN 283 PQNTFVPLGLIDLAIAKRNLIIGDDFKSGQTKMKSVLVDFLVGAGIKPTSTVSYNHLGND
P42801 INO1_ARATH 284 PQNTFVPLGLIDLAIAKRNLIIGDDFKSGQTKMKSVLVDFLVGAGIKPTSTVSYNHLGND
Q38862 INO2_ARATH 283 PQNTFVPLGLIELAISKNCILIGDDFKSGQTKMKSVLVDFLVGAGIKPTSTVSYNHLGND
P11986 INO1_YEAST 297 PQNTFVPLGLIAEHGEGTFIAGDDLKSGQTKLKSVLVDFLVGAGIKPTSTVSYNHLGND
D2D2Z4 D2D2Z4_GOSHI 283 PQNTFVPLGLIDLAIAKRNLIIGDDFKSGQTKMKSVLVDFLVGAGIKPTSTVSYNHLGND
064437 INO1_ORYSJ 283 PQNTFVPLGLIDLAIAKRNLIIGDDFKSGQTKMKSVLVDFLVGAGIKPTSTVSYNHLGND
Q9S7U0 INO1_WHEAT 283 PQNTFVPLGLIELAISKNCILIGDDFKSGQTKMKSVLVDFLVGAGIKPTSTVSYNHLGND
Q9LX12 INO3_ARATH 283 PQNTFVPLGLIELAISKNCILIGDDFKSGQTKMKSVLVDFLVGAGIKPTSTVSYNHLGND
* * * * * * * * : * * * * * * * * : * * * * * * * * : * * * * * *
```

```
Q94C02 Q94C02_SOYBN 343 GMNLSAPQTFRSKEISKSNVDDMVSNAILIY--PGEHPDHVVVVKIYVYVWDSKRAMD
P42801 INO1_ARATH 344 GMNLSAPQTFRSKEISKSNVDDMVSNAILIY--PGEHPDHVVVVKIYVYVWDSKRAMD
Q38862 INO2_ARATH 343 GMNLSAPQTFRSKEISKSNVDDMVSNAILIY--PGEHPDHVVVVKIYVYVWDSKRAMD
P11986 INO1_YEAST 357 GYNLSAPQTFRSKEISKSNVDDMVSNAILIY--PGEHPDHVVVVKIYVYVWDSKRAMD
D2D2Z4 D2D2Z4_GOSHI 343 GMNLSAPQTFRSKEISKSNVDDMVSNAILIY--PGEHPDHVVVVKIYVYVWDSKRAMD
064437 INO1_ORYSJ 343 GMNLSAPQTFRSKEISKSNVDDMVSNAILIY--LGEHPDHVVVVKIYVYVWDSKRAMD
Q9S7U0 INO1_WHEAT 343 GMNLSAPQTFRSKEISKSNVDDMVSNAILIY--PGEHPDHVVVVKIYVYVWDSKRAMD
Q9LX12 INO3_ARATH 343 GMNLSAPQTFRSKEISKSNVDDMVSNAILIY--PGEHPDHVVVVKIYVYVWDSKRAMD
* * * * * * * * : * * * * * * * * : * * * * * * * * : * * * * * *
```

```
Q94C02 Q94C02_SOYBN 401 EYTSIFMGGKNTIVLHNTCEDSLAAPIILDLVLLAELSTRIQFKA----ENEKGFHS
P42801 INO1_ARATH 402 EYTSIFMGGKNTIVLHNTCEDSLAAPIILDLVLLAELSTRIQFKA----EGEGKGFHS
Q38862 INO2_ARATH 401 EYTSIFMGGKNTIVLHNTCEDSLAAPIILDLVLLAELSTRIQFKA----EGEGKGFHS
P11986 INO1_YEAST 417 EYSELMLGGHNRISLHNVCEDSLATPLIDLLVMTLLEFCTRVSYKVVDPKEDAGKFN
D2D2Z4 D2D2Z4_GOSHI 401 EYTSIFMGGKNTIVLHNTCEDSLAAPIILDLVLLAELSTRIQFKA----DGEKGFHS
064437 INO1_ORYSJ 401 EYTSIFMGGKNTIVLHNTCEDSLAAPIILDLVLLAELSTRIQFKA----EGEGKGFHS
Q9S7U0 INO1_WHEAT 401 EYTSIFMGGKNTIVLHNTCEDSLAAPIILDLVLLAELSTRIQFKA----EGEGKGFHS
Q9LX12 INO3_ARATH 401 EYTSIFMGGKNTIVLHNTCEDSLAAPIILDLVLLAELSTRIQFMS----ENEKGFHS
* * * * * * * * : * * * * * * * * : * * * * * * * * : * * * * * *
```

```
Q94C02 Q94C02_SOYBN 456 EHPVATILSYLTKAPLVPFGTPVNVALSQRAMLENIRACVGLAPENNMILEYK--
P42801 INO1_ARATH 457 EHPVATILSYLTKAPLVPFGTPVNVALSQRAMLENIRACVGLAPENNMILEYK--
Q38862 INO2_ARATH 456 EHPVATILSYLTKAPLVPFGTPVNVALSQRAMLENIRACVGLAPENNMILEYK--
P11986 INO1_YEAST 477 FYPVLTPLSYLWIKAPLIRFPFPVNLQKQRTALENPLRLILGLPSQNLRFPERLL
D2D2Z4 D2D2Z4_GOSHI 456 EHPVATILSYLTKAPLVPFGTPVNVALSQRAMLENIRACVGLAPENNMILEYK--
064437 INO1_ORYSJ 456 EHPVATILSYLTKAPLVPFGTPVNVALSQRAMLENIRACVGLAPENNMILEYK--
Q9S7U0 INO1_WHEAT 456 EHPVATILSYLTKAPLVPAGTPVNVALSQRAMLENIRACVGLAPENNMILEYK--
Q9LX12 INO3_ARATH 456 EHPVATILSYLTKAPLVPFGTPVNVALSQRAMLENIRACVGLAPENNMILEYK--
* * * * * * * * : * * * * * * * * : * * * * * * * * : * * * * * *
```

# phylogeny construction



Phylogenetic tree constructed by MEGA 6.0



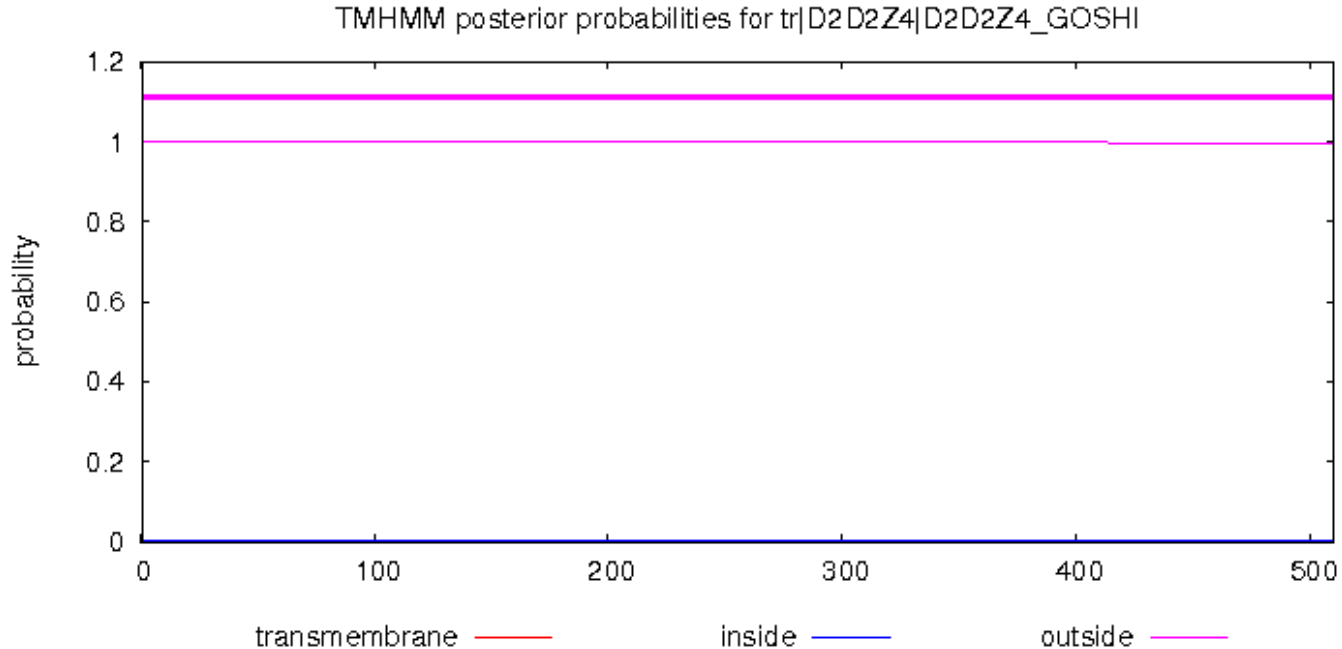
---



# Motif prediction

# Transmembrane structure prediction

```
# tr|D2D2Z4|D2D2Z4_GOSHI Length: 510
# tr|D2D2Z4|D2D2Z4_GOSHI Number of predicted TMHs: 0
# tr|D2D2Z4|D2D2Z4_GOSHI Exp number of AAs in TMHs: 0.0901
# tr|D2D2Z4|D2D2Z4_GOSHI Exp number, first 60 AAs: 0.00054
# tr|D2D2Z4|D2D2Z4_GOSHI Total prob of N-in: 0.00196
tr|D2D2Z4|D2D2Z4_GOSHI TMHMM2.0      outside 1 510
```



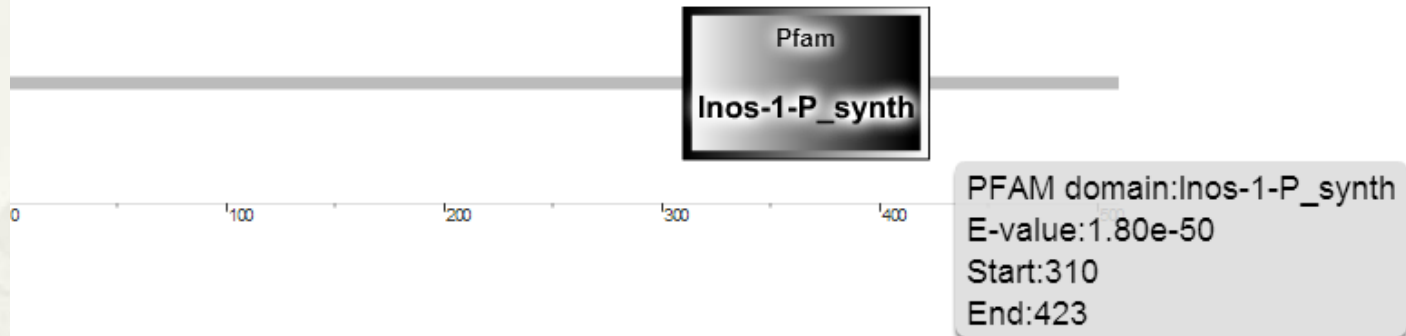


# Domain finding

## Domains within *Gossypium hirsutum* protein D2D2Z4\_GOSHI (D2D2Z4)

Myo-inositol-1-phosphate synthase

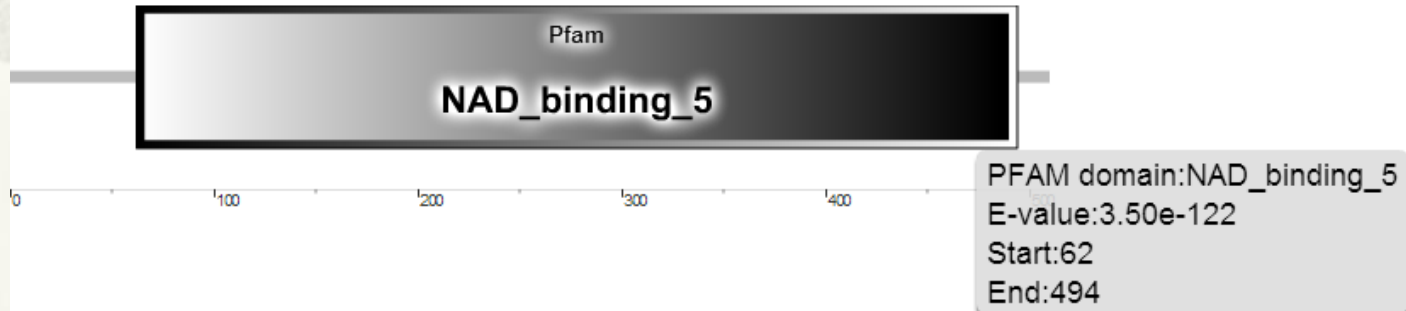
Alternative representations: 1 / 2



## Domains within *Gossypium hirsutum* protein D2D2Z4\_GOSHI (D2D2Z4)

Myo-inositol-1-phosphate synthase

Alternative representations: 2 / 2



# 3D structure prediction



# 1. Search for templates









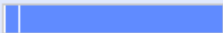










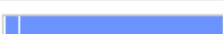


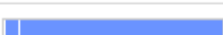


















## Template Results

Templates

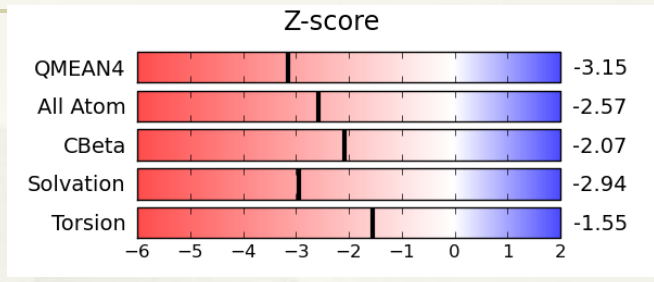
Sequence Similarity

Alignment of Selected Templates

More 

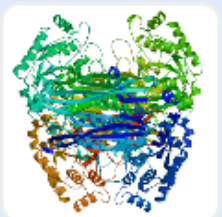
			Title					Ligands	
<input type="checkbox"/>	<input checked="" type="checkbox"/>	<a href="#">1p1f.1.B</a>	Inositol-3-phosphate synthase		55.64	X-ray, 2.6Å	homo-tetramer	None	
<input type="checkbox"/>	<input checked="" type="checkbox"/>	<a href="#">1p1h.1.C</a>	Inositol-3-phosphate synthase		55.64	X-ray, 2.0Å	homo-tetramer	4 x NAD 	
<input type="checkbox"/>	<input checked="" type="checkbox"/>	<a href="#">1p1h.1.B</a>	Inositol-3-phosphate synthase		55.64	X-ray, 2.0Å	homo-tetramer	4 x NAD 	
<input type="checkbox"/>	<input checked="" type="checkbox"/>	<a href="#">1p1h.1.A</a>	Inositol-3-phosphate synthase		55.64	X-ray, 2.0Å	homo-tetramer	4 x NAD 	
<input type="checkbox"/>	<input checked="" type="checkbox"/>	<a href="#">1p1i.1.B</a>	Inositol-3-phosphate synthase		55.64	X-ray, 2.4Å	homo-tetramer	4 x NAD 	
<input type="checkbox"/>	<input checked="" type="checkbox"/>	<a href="#">1la2.1.A</a>	Myo-inositol-1-phosphate synthase		55.64	X-ray, 2.6Å	homo-tetramer	4 x NAD 	
<input type="checkbox"/>	<input checked="" type="checkbox"/>	<a href="#">1jkf.1.B</a>	myo-inositol-1-phosphate synthase		55.64	X-ray, 2.4Å	homo-tetramer	4 x NAD 	
<input type="checkbox"/>	<input checked="" type="checkbox"/>	<a href="#">1jkf.1.A</a>	myo-inositol-1-phosphate synthase		55.64	X-ray, 2.4Å	homo-tetramer	4 x NAD 	
<input type="checkbox"/>	<input checked="" type="checkbox"/>	<a href="#">1vko.1.A</a>	inositol-3-phosphate synthase		54.49	X-ray, 2.3Å	homo-tetramer	4 x K  , 4 x POP  , 4 x NAD 	
<input type="checkbox"/>	<input checked="" type="checkbox"/>	<a href="#">1vko.1.A</a>	inositol-3-phosphate synthase		55.85	X-ray, 2.3Å	homo-tetramer	4 x K  , 4 x POP  , 4 x NAD 	

## 2. Build model



### Model Results

Order by: QMEAN4



Model 06

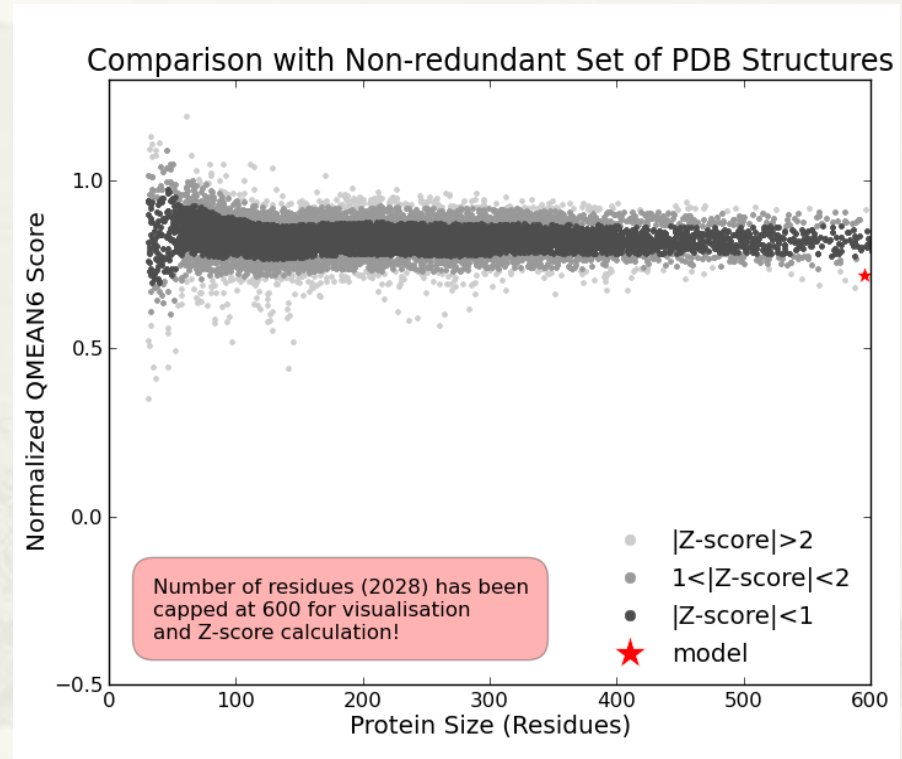
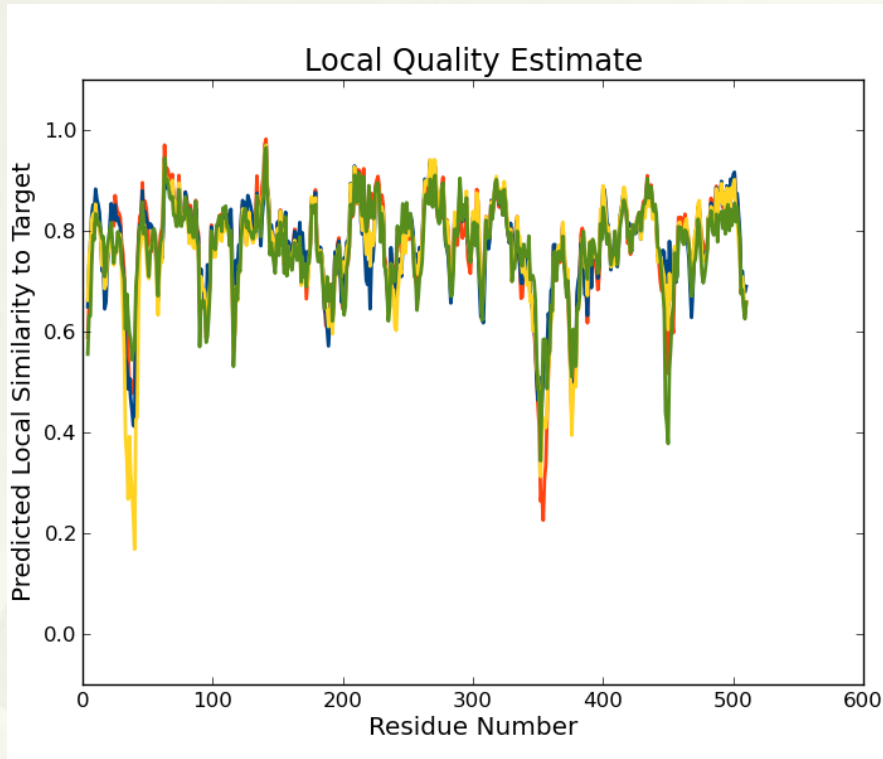
Oligo-State	Ligands	GMQE	QMEAN4
Homo-tetramer (matching prediction)	None	0.80	-3.15

Template	Seq Identity	Coverage	Description
1la2.1.A	55.64%	<div style="width: 55.64%;"></div>	Myo-inositol-1-phosphate synthase

Oligo-state	Method	Seq Similarity	Range	Coverage
homo-tetramer	X-ray, 2.65 Å	0.46	4 - 510	0.99

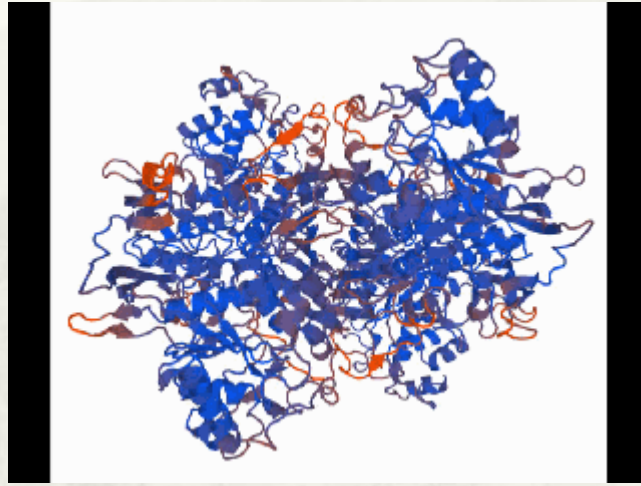
Ligand	Added to Model	Description
NAD	X - Binding site not conserved.	NICOTINAMIDE-ADENINE-DINUCLEOT...
NAD	X - Binding site not conserved.	NICOTINAMIDE-ADENINE-DINUCLEOT...
NAD	X - Binding site not conserved.	NICOTINAMIDE-ADENINE-DINUCLEOT...
NAD	X - Binding site not conserved.	NICOTINAMIDE-ADENINE-DINUCLEOT...

# 3. Model evaluation



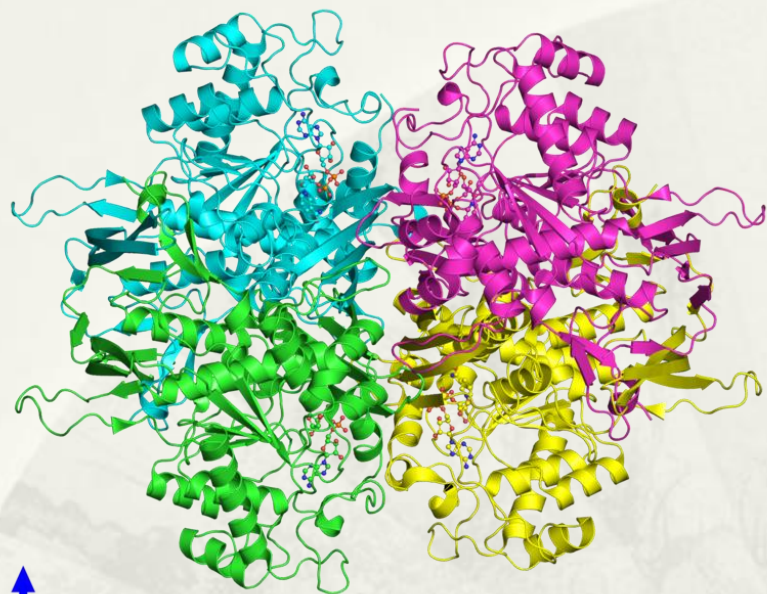
## 4. Modeling result

---

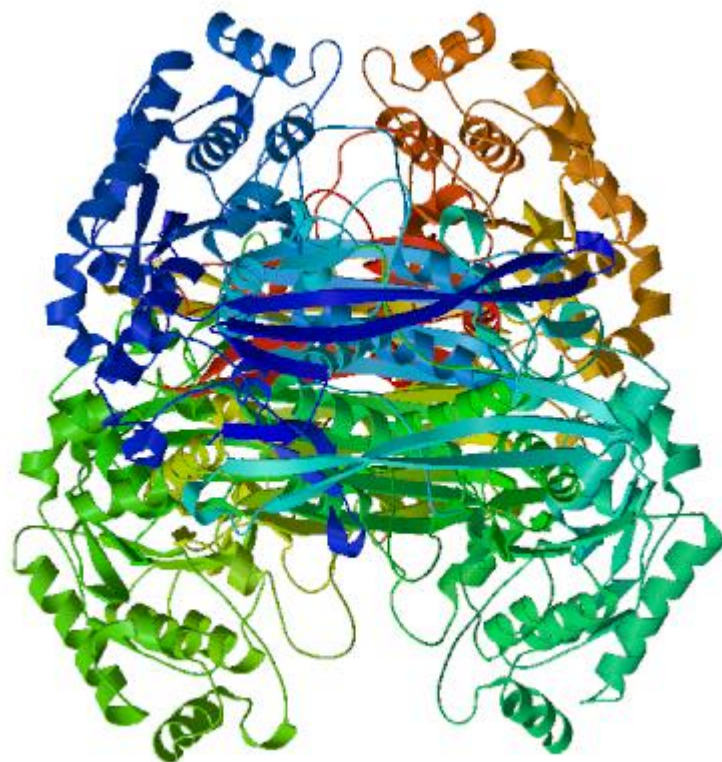




## 4. Modeling result

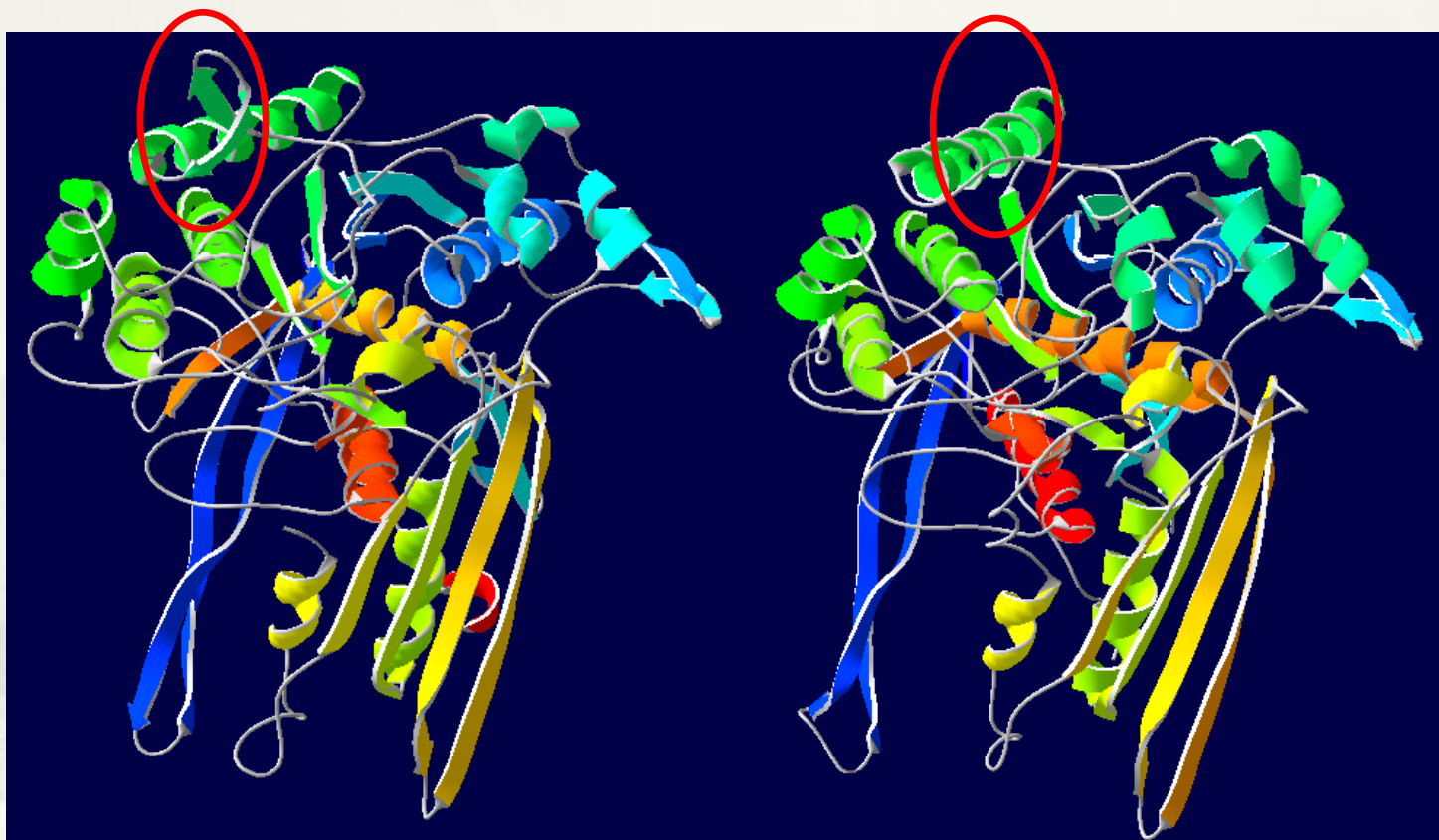


Template: 1la2



Model: D2D2Z4

## 4. Modeling result



1la2

D2D2Z4

# Acknowledgement

---

- \* Prof. Luo;
- \* Members of this class, especially G03 group.

