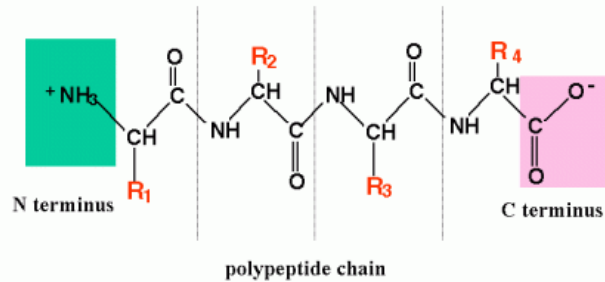


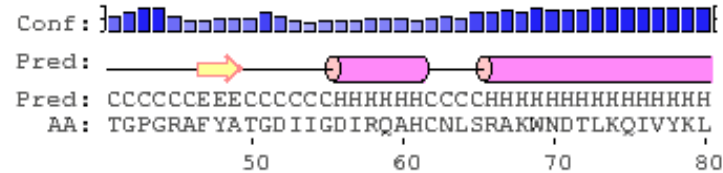
The structure levels in proteins

The **primary structure** is the "one-dimensional" sequence of amino acids in the protein or peptide.

Peptide = chain of amino acids



Secondary structure: Regular structural unit formed by sequence regions in a protein (from a sequence perspective: two-dimensional). Basically, these are of two types: α -helices and β -sheets.

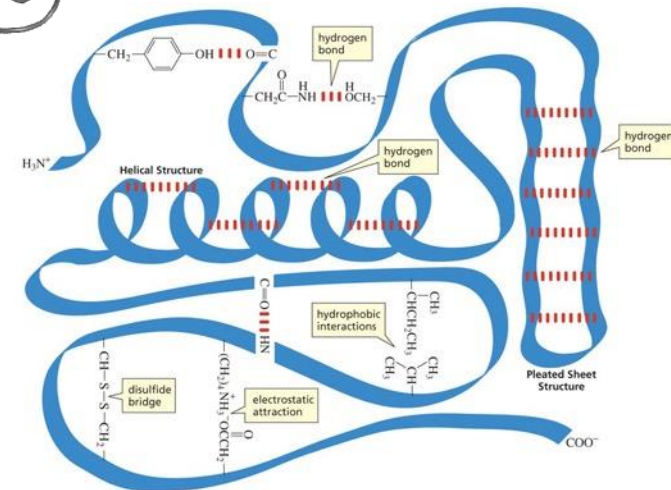
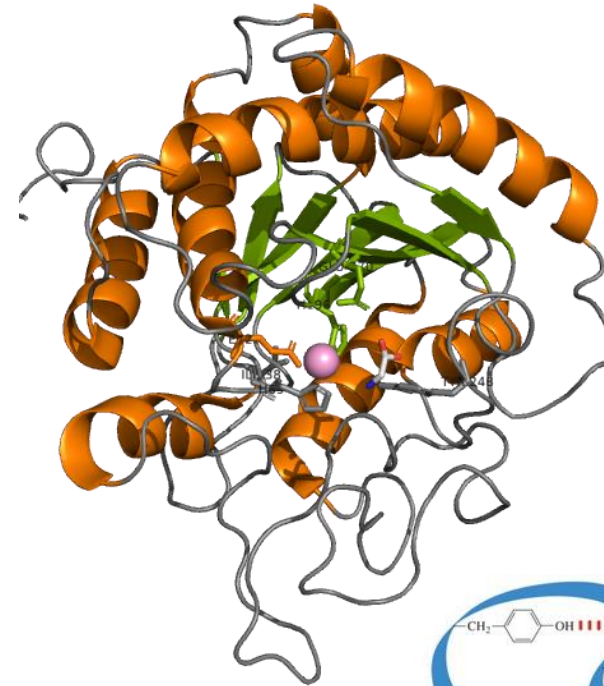


Legends:

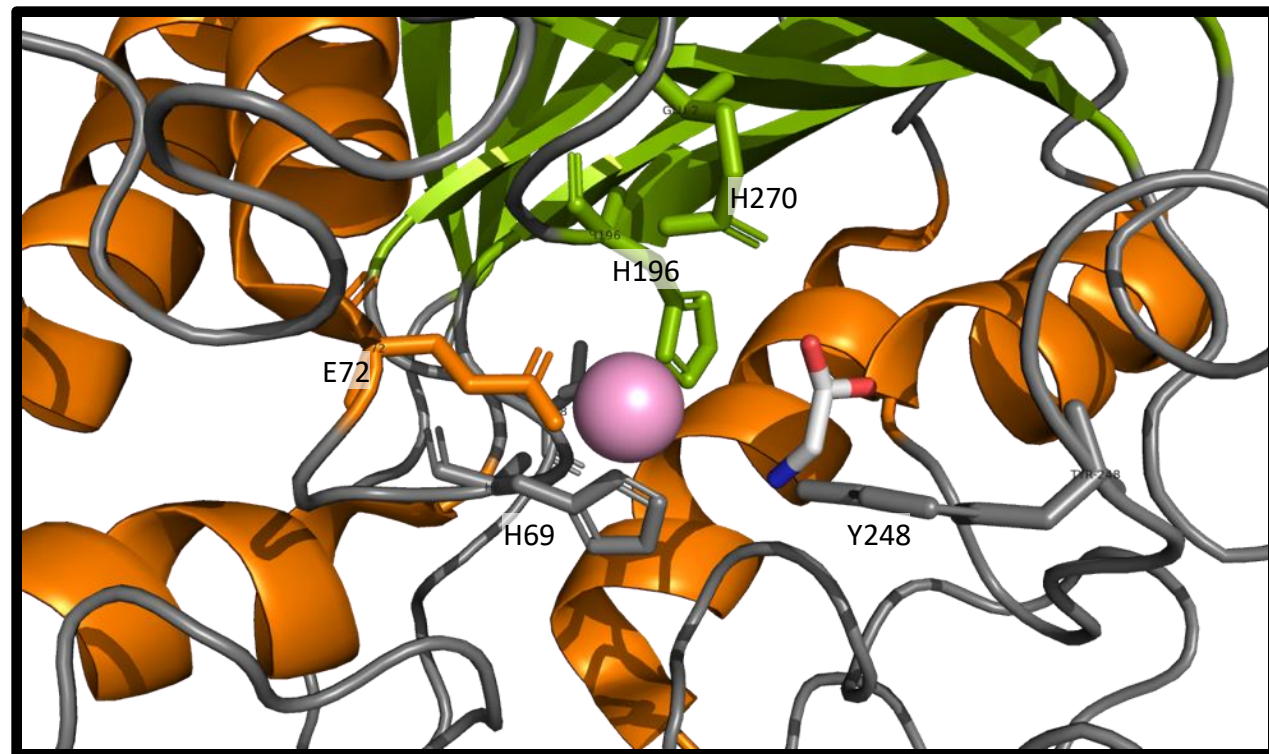
- helix
- strand
- coil
- Conf: - confidence of prediction
- Pred: - predicted secondary structure
- AA: target sequence

>NP_000508.1
 MVLSPADKTNVKAAWGKVGAHAGEYGAEALERM
 FLSFPTTKTYFPHFDLSHGSAQVKGHGKKVADA
 LTNVAHVDDMPNALSALSDLHAHKLRVDPVNF
 KLLSHCLLVTLAAHLPAEFTPAVHASLDKFLAS
 VSTVLTISKYR

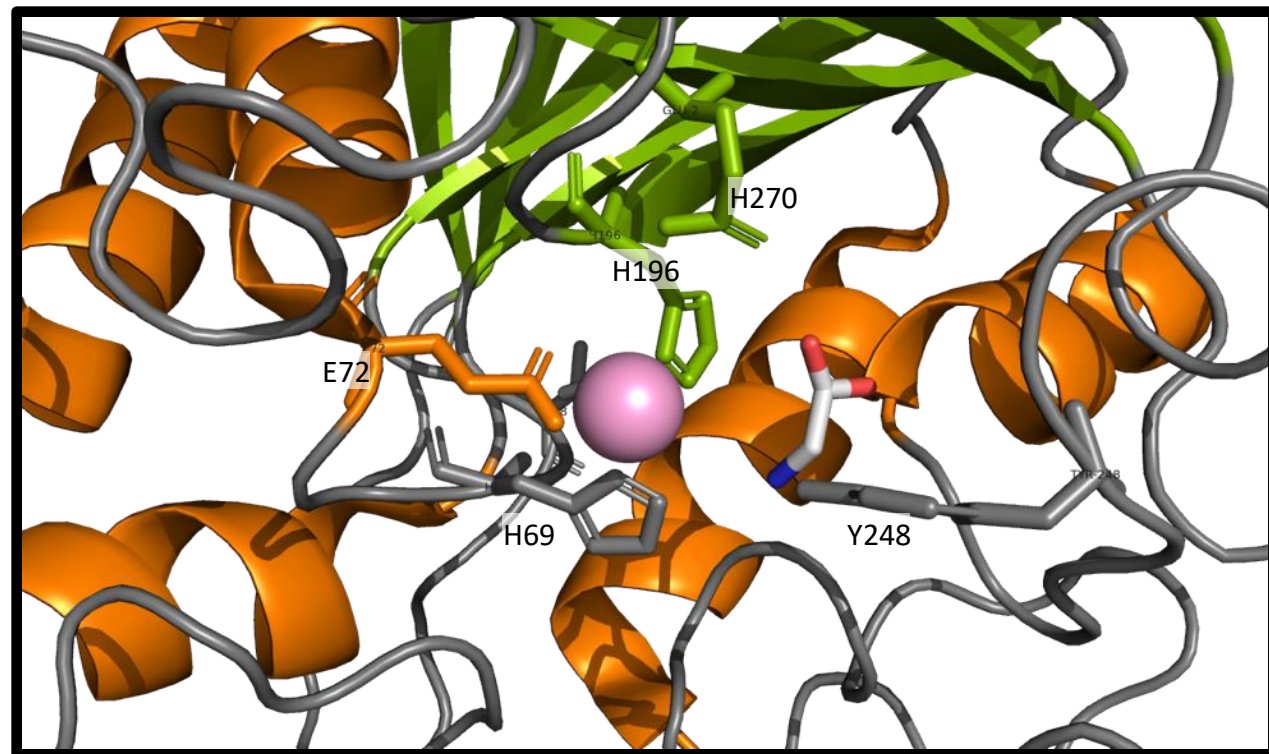
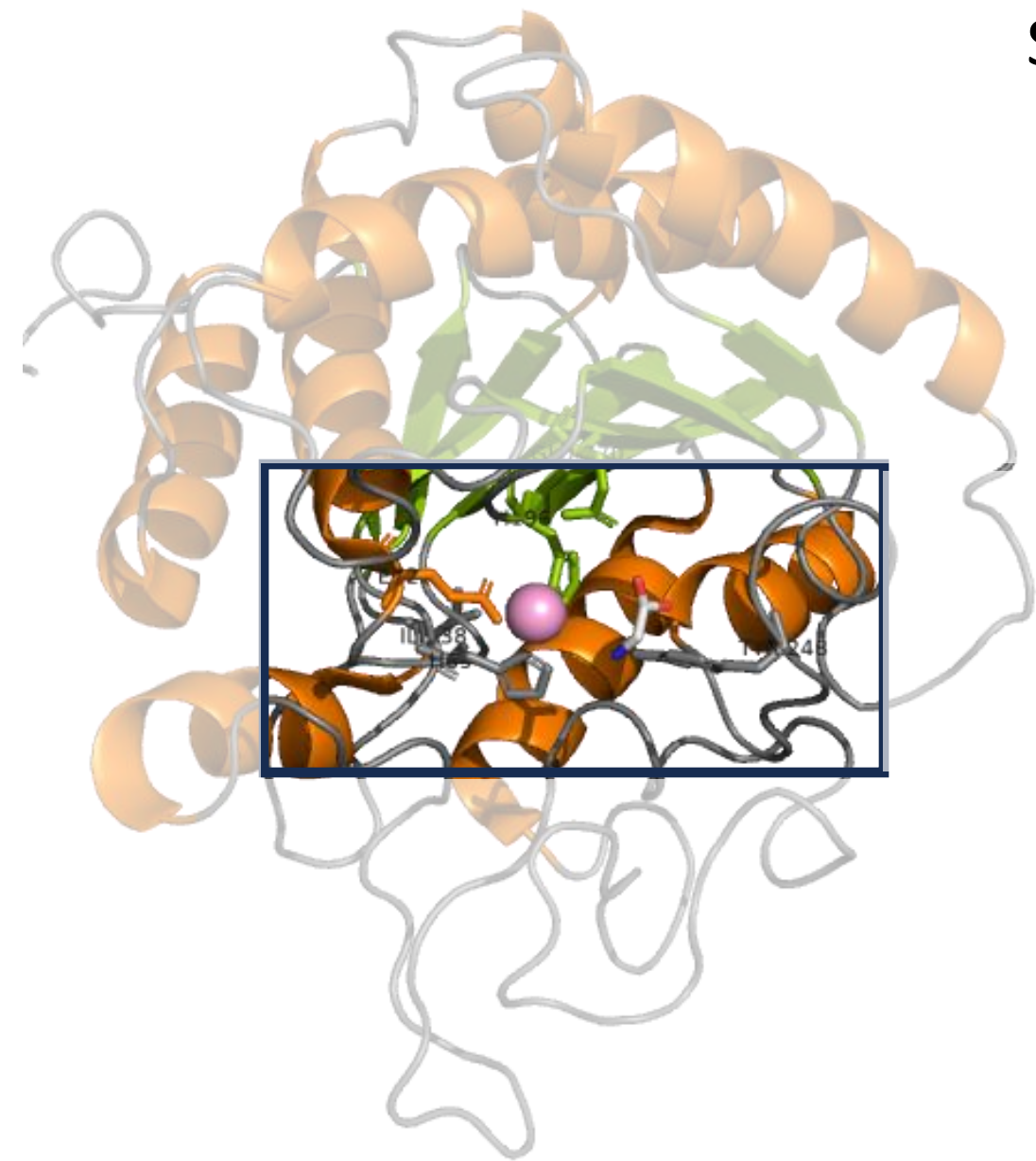
Tertiary structure is the three-dimensional fold of the protein, or how the secondary structure elements are packed against each other.

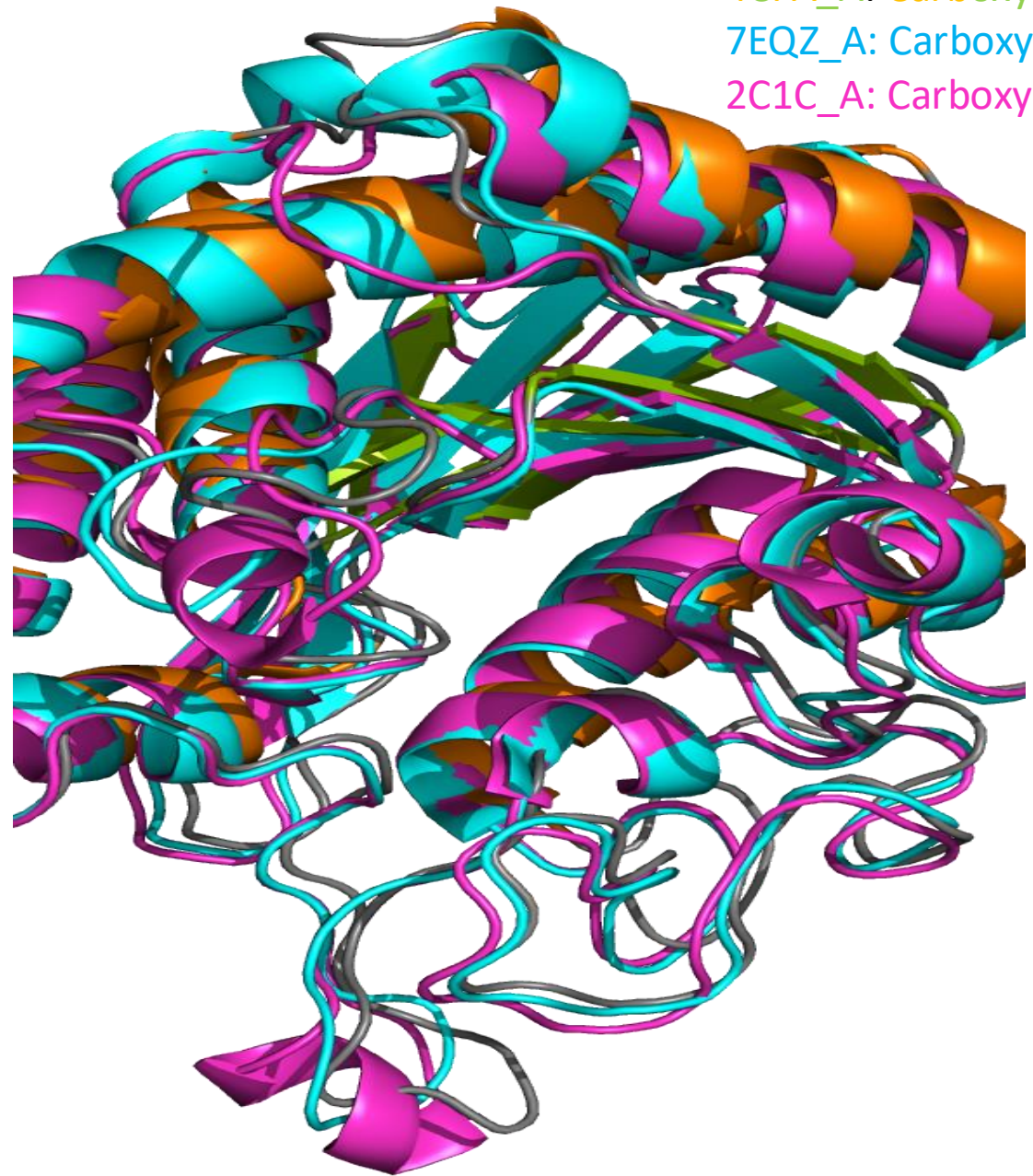


Structure \Rightarrow Function
Function \Leftarrow



Structure \Rightarrow Function
Function \Leftarrow





4CPA_A: Carboxypeptidase A (Bos taurus)

7EQZ_A: Carboxypeptidase B (Aedes aegypti)

2C1C_A: Carboxypeptidase B (Helicoverpa zea)

```

Query=4cpaA Sbjct=7eqzA
Chain  Z  rmsd lali nres %id Description
7eqz-A 43.9  1.6  296  299  33 MOLECULE: CARBOXYPEPTIDASE B;

Query  arstntFNyATyHTLDEIyDFMDLLVAQHPELVSKLQIGRSyEGRPIyVLKfSTGgSNRP  60
ident  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct  -----DVSTsYLrhNEINEyLQTLsQkYPSLVSVEEAGTSyEGRsIKTITINKK-PGNA  53

Query  AIWIDLGIHSREWITQATGvWfAKKFTENyGQnpsFTAILDSMDIFLEIvTNPNGfAFTH  120
ident  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct  vVfLDAGIHAREWIAPATALyAIEQLVEHSSE---NQEVLSNLTWIMpVvNPDGyEFsH  110

Query  SENRLWRKTRsVTSSsSLCVGDANRNWDAGfGKAGASSSPCSEyHGKYANSEVEVKsIV  180
ident  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct  EDRfWRKTRKPTG-KSCKGTdGNRNfDYHWGEVGASTQACADTFRGETAFSEpETRAVR  169

Query  DFvKNHG-NfKAFLSIHsYSQLLLYPyGYTTQsIPDKTELnQvAKsAVAAALkSLyGTsYK  239
ident  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct  DAVMKLkGScKFyLSLHsYGNyILyPWGwTskLPETWEAIDEVAQAGAEAIKQSTGSRyT  229

Query  YGSIIITTIyQASGGsIDWSyN-QGIKysFTfELRDtGRyGfLLPASQIIPtAQETwLgVL  298
ident  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct  VGSSTNVLYAAAGSDDWAFaVAEVPISITMELPGGNGGFNPPSSIEKIVNESWVGIK  289

Query  TIMEHTVNNg  308
ident
Sbjct  AMALKVAQMf  299

Query=4cpaA Sbjct=2c1cA
Chain  Z  rmsd lali nres %id Description
2c1c-A 43.3  1.6  297  312  33 MOLECULE: CARBOXYPEPTIDASE B;

Query  arstntFNyATyHTLDEIyDFMDLLVAQHPELVSKLQIGRSyEGRPIyVLKfST--GGSN  58
ident  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct  -----LPyDNYQELeVIDEYLDyIGEKyPDVATvVNAAESfEGRPIKykISTnFEDE  54

Query  -RPAIWIDLGIHSREWITQATGvWfAKKFTENyGQnpsFTAILDSMDIFLEIvTNPNGfA  117
ident  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct  nKPVIFIDGGIHAREWISPPSVTWAIHKLVEDVT---ENDLLEKfDWILLPVvNPDGyK  110

Query  FTHeNRLWRKTRsVTSS---SLCVGDANRNWDAGfGKAGASSSPCSEyHGKYANSEV  174
ident  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct  YTFtNERfWRKTRSTNNpIsQICRGADGNRNfDFVWNSIGTSNSPCSDIyAGTSAFSEV  170

Query  EVKsIVDFvKNH-GNfKAFLSIHsYSQLLLYPyGYTTQsIPDKTELnQvAKsAVAAALkSL  233
ident  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct  ETRWRDILHEHlARMALyLTMHsFGSMILyPWGHdGSLsQNALGLHTVgVAMASVIQSN  230

Query  YG--TSyKYGSIIITTI-YQASGGsIDWSyNQGIKysFTfELRDtGR--YGFLLPASQII  287
ident  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct  ALpnfPPyTVGNSALVIGyYIAGSSeDYAHsIGVPLsYTYELPGLSSgWdGFHLPPQyIE  290

Query  PTAQETwLgVLTIMEHTVNn-g  308
ident  | | | |
Sbjct  QVCRETWEGIVVGARRAGDLfr  312

```

Gene: Leghemoglobin-1

Organism: *Lupinus luteus* (Llobí groc)

Function: It facilitates the diffusion of oxygen to symbiotic bacteriods in order to promote nitrogen fixation.



Gene: Leghemoglobin-1

Organism: *Lupinus luteus* (Llobí groc)

Function: It facilitates the diffusion of oxygen to symbiotic bacteriods in order to promote nitrogen fixation.



Gene: Hemoglobin subunit beta

Organism: *Homo sapiens* (Humà)

Function: Involved in oxygen transport from the lung to the various peripheral tissues.



```

CAA68462.1 1 ATGGGTGTTTAACTGATGTG--CAAGT-----GGCTTTGGTG 36
|||...:|.||...|| | ||| ..|.| |||.|
AAA35597.1 1 ATGGTNCAYYTACNCNCNGTGGAGAAGTCYGCYGTNACNCNCNTG 50
GGG
CAA68462.1 37 AAGAGCTCA--TTTGAAGAATT-----TAATGCAAATATTCCTAAAAAC 78
.| |||.|. | .|.|||.||.: :|.||
AAA35597.1 51 YA-AGGTNAAYGTGGATGAAGYYGGYGGYGAGGC----- 83
CAA68462.1 79 ACC----CA--CCGTT-----TCTTCACCTTGGTACTAGAGATTGCACCA 117
|| || |.|. | |||. | |||||. .|. ||||
AAA35597.1 84 -CCTGGGCAGNCTGCTNGTGGTCTAC-CCTTGGACCCAGAG----- 122
CAA68462.1 118 GGAGCAAAGGATTTGTTCT--CATT--TTTGAAGGGATCTAGTGAAG--T 161
||| |. | ||: | ||||| | .
AAA35597.1 123 -----GTTCTTNGANTCNTTYG--GGGATCT-----GNNN 150
CAA68462.1 162 ACCCCAGA-----AT----AATCCTGATCTTCAGGCCCAT-GCTGGAA 199
||. ||. || || |. |||. |. . ||||. ||| ||. |||
AAA35597.1 151 ACNCCNGANGCAGTTATGGGCAACCTAAGGTGAAGGCTCATGGCAAGAA 200
CAA68462.1 200 A-----GGT---TTTAAAGT-----TGACTTAC--GAAGCAGCAAT 230
| ||| ||| || |. ||. || ||
AAA35597.1 201 AGTGCTCGGTGCCTTT--AGTGATGGCTGGCTCACCTGGA----- 239
CAA68462.1 231 TCAACTTCAAGTGAATGGAGCAGTGGCTTCAGATGCCACGTTGAAAAGTT 280
||||. |||||. |. | . ||| |||||. |||
AAA35597.1 240 -CAACCTCAAGGGCA-----CCTT----TGCCACACTGA----- 268
CAA68462.1 281 TGGGTTCTGTCCA-TGTCTCAAAAGGAGTCGTTGATGCCCATTTTCCGGT 329
|. |. . ||| || |||. ||| || |. |. . |||
AAA35597.1 269 -GTGAGCTG--CACTGTGACAA-----GC-----TNCAYGT 296
CAA68462.1 330 GGTGAAGGAAGCAATCCTGAAAACAATAAAGG----- 361
|| ||| |||||. ||| |. |||
AAA35597.1 297 GG-----ATCTGAGAAC--TTCAGGCTNCTNGGCAACGTGYTN 333
CAA68462.1 362 ----GAGTGGTGGGAGA-CAAATGGAGCGAGGAACTGAACACTGCTTGG 406
|. |||. |||. |. |||. |. |||. |. |. |. |
AAA35597.1 334 GTCTGYGTGCTGGCCCATCACTTTG-GCAAAGAATTACCCCC-----A 375
CAA68462.1 407 CCA-----TAGCCTATGACGAATTGG----- 427
||| .. |||||. |. |||. |||
AAA35597.1 376 CCAGTGCANGCNGCCTATCAGAAAGTGGTNGCTGGTNGCTAATGCCCT 425
CAA68462.1 428 -----CAATTATAATTAAGAAGGAGATGAAGGATGCTGCTTAA 465
|||. |||. |. |||
AAA35597.1 426 GGCCCAAGTATCACTAA----- 444

```

```

#####
# Program: needle
#
# Aligned_sequences: 2
# 1: CAA68462.1
# 2: AAA35597.1
# Matrix: EDNAFULL
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 594
# Identity:      225/594 (37.9%)
# Similarity:   230/594 (38.7%)
# Gaps:         279/594 (47.0%)
# Score: 309.0
#
#
#=====

```

```

CAA68462.1      1  ATGGGTGTTTAACTGATGTG--CAAGT-----GGCTTTGGTG      36
   |||...:|.||...||| .||| ..||.|||.|
AAA35597.1      1  ATGGTNCAYYTACNCCNGTGGAGAAGTCYGCYGTACNCGCNCNTNTGGGG      50

CAA68462.1     37  AAGAGCTCA--TTTGAAGAATT-----TAATGCAAAATATTCCTAAAAAC      78
   .| |.|.| .|.||.||||.: :|.||
AAA35597.1     51  YA-AGGTNAAYGTGGATGAAGYGGYGGYGAGGC-----            83

CAA68462.1     79  ACC----CA--CCGTT----TCTTCACCTTGGTACTAGAGATTGCACCA      117
   ||  || |.|.| |||.| |||||.|.|||
AAA35597.1     84  -CCTGGGCAGNCTGCTNGTGGTCTAC-CCTTGGACCCAGAG-----      122

CAA68462.1    118  GGAGCAAAGGATTTGTTCT--CATT--TTTGAAGGATCTAGTGAAG--T      161
   ||||| .|.| ||:| ||||| | .
AAA35597.1    123  -----GTTCTTNGANTCNTTYG--GGGATCT-----GNNN      150

CAA68462.1    162  ACCCCAGA-----AT----AATCCTGATCTTCAGGCCCAT-GCTGGAA      199
   ||.|||.||  ||  ||.|||.|.|.|.|||.||| ||..|||
AAA35597.1    151  ACNCCNGANGCAGTTATGGGCAACCCTAAGGTGAAGGCTCATGGCAAGAA      200

CAA68462.1    200  A-----GGT---TTTTAAGT-----TGACTTACGAAGCAGCAATTCA      233
   |  ||  ||| |||  ||.||.|| |.| | ||
AAA35597.1    201  AGTGCTCGGTGCCTTT--AGTGATGGCCTGGCTCAC---CTGGA---CA      241

CAA68462.1    234  ACTTCAAGTGAATGGAGCAGTGGCTTCAGATGCCACGTTGAAAAGTTTGG      283
   ||.|||||.|.| .|||  |||||.||| |.
AAA35597.1    242  ACCTCAAGGGCA-----CCTT----TGCCACACTGA-----GT      270

CAA68462.1    284  GTTCTGTCCA-TGTCTCAAAGGAGTCGTTGATGCCCATTTTCCGGTGGT      332
   |..|||  || |||.|||  ||  |.|..|||
AAA35597.1    271  GAGCTG--CACTGTGACAA-----GC-----TNCAYGTGG-      298

CAA68462.1    333  GAAGGAAGCAATCCTGAAAACAATAAAGG-----            361
   |||||.||| |..|||
AAA35597.1    299  -----ATCCTGAGAAC--TTCAGGCTNCTNGGCAACGTGYTNGTC      336

CAA68462.1    362  -GAGTGGTGGGAGA-CAAATGGAGCGAGGAACTGAACACTGCTTGGACCA      409
   |.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||
AAA35597.1    337  TGYGTGCTGGCCCATCACTTTG-GCAAAGAATTACCCCC-----ACCA      378

CAA68462.1    410  -----TAGCCTATGACGAATTGG-----            427
   ..|||||.|||.|||
AAA35597.1    379  GTGCANGCNGCCTATCAGAAAGTGGTNGCTGGTNGCTAATGCCCTGGC      428

CAA68462.1    428  ---CAATTATAATTAA      440
   |||.|||.|||
AAA35597.1    429  CCACAAGTATCACTAA      444

```

```

#####
# Program: water
#
# Aligned_sequences: 2
# 1: CAA68462.1
# 2: AAA35597.1
# Matrix: EDNAFULL
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 566
# Identity:      226/566 (39.9%)
# Similarity:   231/566 (40.8%)
# Gaps:         248/566 (43.8%)
# Score: 309.0
#
#
#=====

```


CAA68462.1 1 ATGGGTGTTTAACTGATGTG--CAAGT-----GGCTTTGGTG 36
|||...:|.||...|| |.|| | | | .|.|||. |

AAA35597.1 1 ATGGTNCAYYTACNCNGTGGAGAAGTCYGCYGTACNCNCNTG 50

CAA68462.1 37 AAGAGCTCA--TTTGAAGAATT-----TAATGCAAATATTCC 78
|.|||. | |.|||.|||.: :|.||

AAA35597.1 51 YA-AGGTNAAYGTGGATGAAGYYGGYGGYGAGGC----- 83

CAA68462.1 79 ACC---CA--CCGTT-----TCTTACCTTGGTACTAGAGATT 117
|| || |.|. | |||. | |||||. .|. ||||

AAA35597.1 84 -CCTGGGCAGNCTGCTNGTGGTCTAC-CCTTGGACCCAGAG- 122

CAA68462.1 118 GGAGCAAAGGATTTGTTCT--CATT--TTTGAAGGGATCTAGT 161
|||| |. | ||:| ||||| | .

AAA35597.1 123 -----GTTCTTNGANTCNTTYG--GGGATCT-----GN 150

CAA68462.1 162 ACCCCAGA-----AT----AATCCTGATCTTCAGGCCCAT - 199
||. ||. || | ||. |||. |. |. |. |. |. |. |. |. |. |. |. |. |. |. |. |

AAA35597.1 151 ACNCCNGANGCAGTTATGGGCAACCTAAGGTGAAGGCTCATGG 200

CAA68462.1 200 A-----GGT---TTTTAAGT-----TGACTTAC---GA 230
| ||| ||| || | |. |. | | |

AAA35597.1 201 AGTGCTCGGTGCC TTT--AGTGATGGCTGGCTCACCTGGA- 239

CAA68462.1 231 TCAACTTCAAGTGAATGGAGCAGTGGCTTTCAGATGCCACG 280
||||. | |||||. |. | . || | | |||||. | | |

AAA35597.1 240 -CAACCTCAAGGGCA-----CCTT----TGCCACACTGA- 268

CAA68462.1 281 TGGGTTCTGTCCA-TGTCTCAAAGGAGTCGTTGATGCCCAT 329
|. |. |. || | ||. |. || | ||. | | | |. |. |. |

AAA35597.1 269 -GTGAGCTG--CACTGTGACAA-----GC-----TN 296

CAA68462.1 330 GGTGAAGGAAGCAATCCTGAAAACAATAAAGG----- 361
|| | | | | | | | | | |. |. | | |

AAA35597.1 297 GG-----ATCTGAGAAC--TTCAGGCTNCTNGGCAACGTG 333

CAA68462.1 362 ---GAGTGGTGGGAGA-CAAATGGAGCGAGGAACTGAAC 406
|. |||. |||. |. | | |. |. | | |. |. |. |. |. |. | |

AAA35597.1 334 GTCTGYGTGCTGGCCATCACTTTG-GCAAAGAATTCACCCC 375

CAA68462.1 407 CCA-----TAGCCTATGACGAATTGG----- 427
|| | . | | | | | |. |. | | | | |

AAA35597.1 376 CCAGTGCANGCNGCCTATCAGAAAGTGGTNGCTGGTNGCT 425

CAA68462.1 428 -----CAATTATAATTAAGAAGGAGATGAAGGATGC 465
|| |. | | |. |. | | |

AAA35597.1 426 GGCCACCAAGTATCACTAA----- 444

CAA68462.1 1 ATGGGTGTTTAACTGATGTG--CAAGT-----GGCTTTGGTG 36
|||...:|.||...|| |.|| | | | .|.|||. |

AAA35597.1 1 ATGGTNCAYYTACNCNCNGTGGAGAAGTCYGCYGTACNCNCNTG 50

CAA68462.1 37 AAGAGCTCA--TTTGAAGAATT-----TAATGCAAATATTCC 78
|.|||. | |.|||.|||.: :|.||

AAA35597.1 51 YA-AGGTNAAYGTGGATGAAGYYGGYGGYGAGGC----- 83

CAA68462.1 79 ACC---CA--CCGTT-----TCTTACCTTGGTACTAGAGATT 117
|| || |.|. | |||. | |||||. .|. ||||

AAA35597.1 84 -CCTGGGCAGNCTGCTNGTGGTCTAC-CCTTGGACCCAGAG- 122

CAA68462.1 118 GGAGCAAAGGATTTGTTCT--CATT--TTTGAAGGGATCTAGT 161
|||| |. | ||:| ||||| | .

AAA35597.1 123 -----GTTCTTNGANTCNTTYG--GGGATCT-----GN 150

CAA68462.1 162 ACCCCAGA-----AT----AATCCTGATCTTCAGGCCCAT - 199
||. ||. || | ||. |||. |. |. |. |. |. |. |. |. |. |. |. |. |. |. |

AAA35597.1 151 ACNCCNGANGCAGTTATGGGCAACCTAAGGTGAAGGCTCATGG 200

CAA68462.1 200 A-----GGT---TTTTAAGT-----TGACTTACGAAGC 233
| ||| ||| || | |. |. | | | |. |. |. | |

AAA35597.1 201 AGTGCTCGGTGCC TTT--AGTGATGGCTGGCTCAC--- 241
CTGGA---CA

CAA68462.1 234 ACTTCAAGTGAATGGAGCAGTGGCTTCAGATGCCACGTT 283
||. | | | | |. |. | . || | | | | |. | | | |. | .

AAA35597.1 242 ACCTCAAGGGCA-----CCTT----TGCCACACTGA- 270
-----GT

CAA68462.1 284 GTTCTGTCCA-TGTCTCAAAGGAGTCGTTGATGCCCAT 332
|. |. | | | | | | | |. | | | |. |. |. |. | | | |

AAA35597.1 271 GAGCTG--CACTGTGACAA-----GC-----TN 298
CAYGTGG-

CAA68462.1 333 GAAGGAAGCAATCCTGAAAACAATAAAGG----- 361
| | | | | | | | | | |. |. | | |

AAA35597.1 299 -----ATCTGAGAAC--TTCAGGCTNCTNGGCAACGT 336
GTYTNGTC

CAA68462.1 362 -GAGTGGTGGGAGA-CAAATGGAGCGAGGAACTGAAC 409
|. |||. |||. |. | | |. |. | | |. |. |. |. |. |. | |

AAA35597.1 337 TGYGTGCTGGCCATCACTTTG-GCAAAGAATTCACCCC 378
-----ACCA

CAA68462.1 410 -----TAGCCTATGACGAATTGG----- 427
|. | | | | | |. |. | | | | |

AAA35597.1 379 GTGCANGCNGCCTATCAGAAAGTGGTNGCTGGTNGCT 428
AATGCCCTGGC

CAA68462.1 428 ---CAATTATAATTAA 440
|| |. | | |. |. | | |

AAA35597.1 429 CCACAAGTATCACTAA 444

		Second Letter				
		U	C	A	G	
1st letter	U	UUU Phe UUC UUA Leu UUG	UCU Ser UCC UCA UCG	UAU Tyr UAC UAA Stop UAG Stop	UGU Cys UGC UGA Stop UGG Trp	U C A G
	C	CUU Leu CUC CUA CUG	CCU Pro CCC CCA CCG	CAU His CAC CAA Gln CAG	CGU Arg CGC CGA CGG	U C A G
	A	AUU Ile AUC AUA AUG Met	ACU Thr ACC ACA ACG	AAU Asn AAC AAA Lys AAG	AGU Ser AGC AGA Arg AGG	U C A G
	G	GUU Val GUC GUA GUG	GCU Ala GCC GCA GCG	GAU Asp GAC GAA Glu GAG	GGU Gly GGC GGA GGG	U C A G

		C	S	T	P	A	G	N	D	E	Q	H	R	K	M	I	L	V	F	Y	W		
3rd letter	C	9																				C	
	S	-1	4																				S
	T	-1	1	5																			T
	P	-3	-1	-1	7																		P
	A	0	1	0	-1	4																	A
	G	-3	0	-2	-2	0	6																G
	N	-3	1	0	-2	-2	0	6															N
	D	-3	0	-1	-1	-2	-1	1	6														D
	E	-4	0	-1	-1	-1	-2	0	2	5													E
	Q	-3	0	-1	-1	-1	-2	0	0	2	5												Q
	H	-3	-1	-2	-2	-2	-2	1	-1	0	0	8											H
	R	-3	-1	-1	-2	-1	-2	0	-2	0	1	0	5										R
	K	-3	0	-1	-1	-1	-2	0	-1	1	1	-1	2	5									K
	M	-1	-1	-1	-2	-1	-3	-2	-3	-2	0	-2	-1	-1	5								M
	I	-1	-2	-1	-3	-1	-4	-3	-3	-3	-3	-3	-3	-3	1	4							I
	L	-1	-2	-1	-3	-1	-4	-3	-4	-3	-2	-3	-2	-2	2	2	4						L
	V	-1	-2	0	-2	0	-3	-3	-3	-2	-2	-3	-3	-2	1	3	1	4					V
	F	-2	-2	-2	-4	-2	-3	-3	-3	-3	-3	-1	-3	-3	0	0	0	-1	6				F
	Y	-2	-2	-2	-3	-2	-3	-2	-3	-2	-1	2	-2	-2	-1	-1	-1	-1	3	7			Y
	W	-2	-3	-2	-4	-3	-2	-4	-4	-3	-2	-2	-3	-3	-1	-3	-2	-3	1	2	11		W
		C	S	T	P	A	G	N	D	E	Q	H	R	K	M	I	L	V	F	Y	W		

```
#####
# Program: needle
# Aligned_sequences: 2
# 1: LGB1_LUPLU
# 2: HBB_HUMAN
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 187
# Identity:      34/187 (18.2%)
# Similarity:   54/187 (28.9%)
# Gaps:         73/187 (39.0%)
# Score: 39.5
#
#
#=====
```

```
LGB1_LUPLU      1  MGVLTDVQVALVKSSFEEFNA-----NIPKNTHRFFTLVLEIAPGAK      42
                        :|...:| |:|      |::.....:|:..|...:
HBB_HUMAN       1  -----MVHLTPEEKSAVTALWGKVNVDVGGGALGRLLVVYPWTQ      40

LGB1_LUPLU     43  DLF-SFLKGSSEVP---QNNPDLQAHAGKV-----FKLTYEA      75
                        ..| || |...| ..||:..||..||      .|.|:...
HBB_HUMAN      41  RFFESF--GDLSTPDAVMGNPKVKAHGKKVLGAFSDGLAHL DNLKGT FAT      88

LGB1_LUPLU     76  AIQLQVNGAVASDATLKSLGSHVSKGVVDAHF-----PVVKEAILKTIK    120
                        ..:|...:.....:|:|.|. |:..||      |.:|.|.|.:.
HBB_HUMAN      89  LSELHCDKLHVDPENFRLLGNVLVC--VLAH HFGKEFTPPVQAAYQKVVA    136

LGB1_LUPLU    121  EV---VGDKWSEELNTAWTIAYDELAIIKKEMKDAA      154
                        .|  :..|:.
HBB_HUMAN    137  GVANALAHKYH-----      147
```

```
#####
# Program: water
# Aligned_sequences: 2
# 1: LGB1_LUPLU
# 2: HBB_HUMAN
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 117
# Identity:      29/117 (24.8%)
# Similarity:   39/117 (33.3%)
# Gaps:         38/117 (32.5%)
# Score: 51.0
#
#
#=====
```

```
LGB1_LUPLU     27  THRFFTLVLEIAPGAKDLFSFLKGSSEVP---QNNPDLQAHAGKV-----
68
                        |.|||.      || |...| ..||:..||..||
HBB_HUMAN      39  TQRFFE-----SF--GDLSTPDAVMGNPKVKAHGKKVLGAFS
73

LGB1_LUPLU     69  -----FKLTYEAAIQLQVNGAVASDATLKSLGSHVSKGVVDAHF--
108
                        .|.|:.....:|...:.....:|:|.|. |:..||
HBB_HUMAN      74  DGLAHL DNLKGT FATLSELHCDKLHVDPENFRLLGNVLVC--VLAH HFGK
121

LGB1_LUPLU    109  ---PVVKEAILKTIKEV      122
                        |.:|.|.|.|.|.|.
HBB_HUMAN    122  EFTPPVQAAYQKV VAGV      138
```


Gene: Leghemoglobin-1

Organism: *Lupinus luteus* (Llobí groc)

Function: It facilitates the diffusion of oxygen to symbiotic bacteroids in order to promote nitrogen fixation.

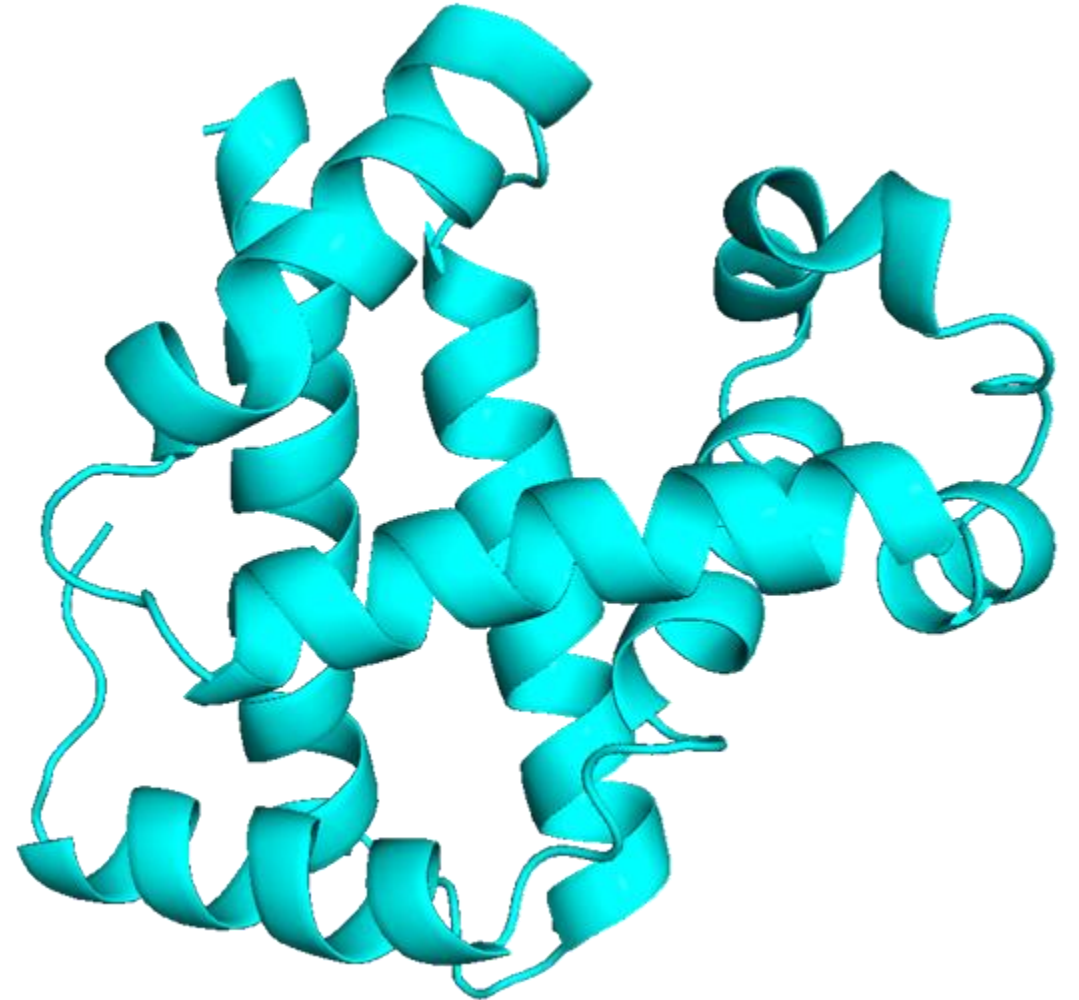
Structure: NO pdb code (No experimentally solved structure)

Gene: Hemoglobin subunit beta

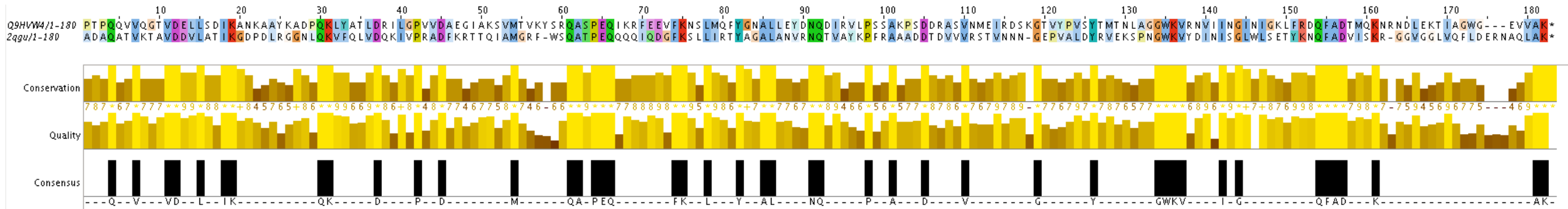
Organism: *Homo sapiens* (Humà)

Function: Involved in oxygen transport from the lung to the various peripheral tissues.

Structure: pdb code: 1IRD chain B



Homology modelling:



```

Q9HVM4      mtlrlrrgllvflaafpllsmaaptqqvVQGTVDLELLSDIKANKAAYKA
#=GR Q9HVM4 PP *****7999*****
2QGU       .....adaqatVKTAVDDVLATIKGDPDLRGG
#=GR 2QGU   PP .....*****899*****
#=GC PP_cons .....799*****
#=GC RF     .....XXXXXXXXXXXXXXXXXXXXXXXXXXXX
    
```

```

Q9HVM4      DPQKLYATLDRILGPPVDAEGIAKSVMTVKYSRQASPEQIKRFEEVFKNS
#=GR Q9HVM4 PP *****9999*****
2QGU       NLQKVFQLVDQKIVPRADFKRTTQIA-GRF.WSQATPEQQQIQDGFKSL
#=GR 2QGU   PP *****987.846*****
#=GC PP_cons *****98*977*****
#=GC RF     XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX.XXXXXXXXXXXXXXXXXXXXX
    
```

```

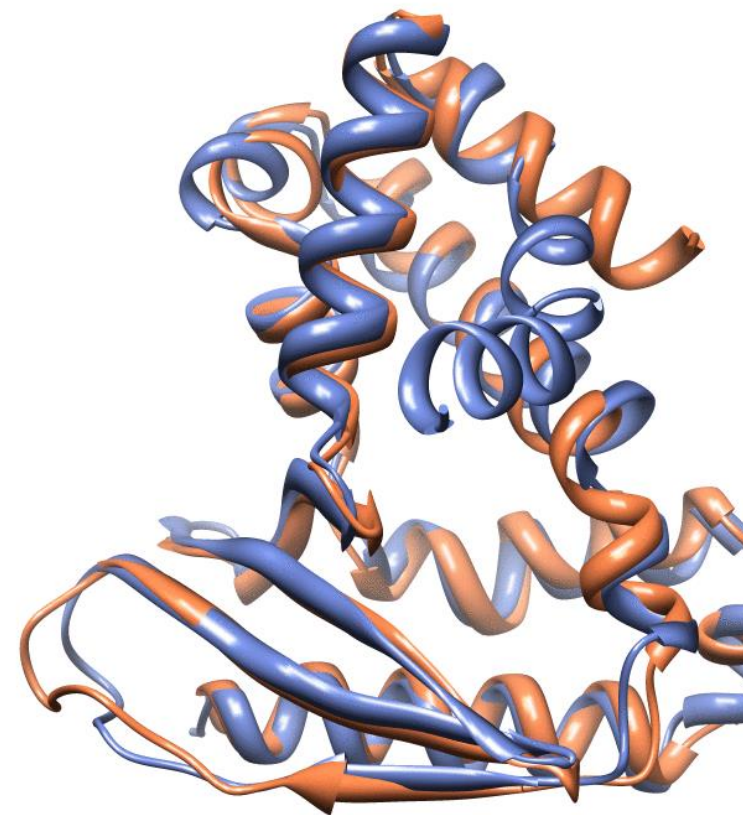
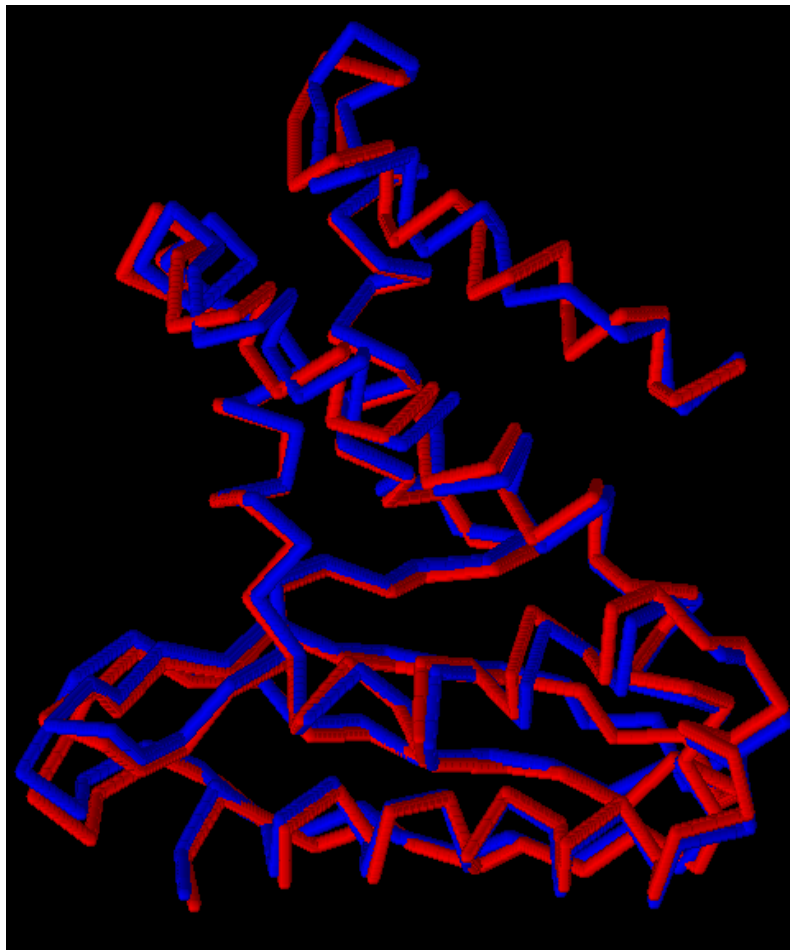
Q9HVM4      LMQFYGNALLEYDNQDIRVLPSSAKPSDDRASVNMETRDSKGTVPVSYT
#=GR Q9HVM4 PP *****6666679*****
2QGU       LIRTYAGALANVRNQTVA YKPFRAAADTDVVVRSTVNNN-GEPVALDYR
#=GR 2QGU   PP *****7777779*****877.5*****
#=GC PP_cons *****6666679*****999*8*****
#=GC RF     XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX.XXXXXXXXXXXXXXXXXXXXX
    
```

```

Q9HVM4      MTNLAGGWKVRNVIINGINIGKLFRDQFADTMQKNrNdLEKT IAGWG---
#=GR Q9HVM4 PP ***8889*****989*****9885...
2QGU       VEKSPNGWKVYDINISGLWLS ETYKNQFADVISKR.GgVGGVLVQFLDERN
#=GR 2QGU   PP ***6669*****.55*****9987
#=GC PP_cons ***7779*****.7.***9997987
#=GC RF     XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX.XXXXXXXXXXXXX
    
```

```

Q9HVM4      evvakaketakaeag
#=GR Q9HVM4 PP 689*****
2QGU       aqlak.....
#=GR 2QGU   PP *****
#=GC PP_cons .....
#=GC RF     .....
    
```

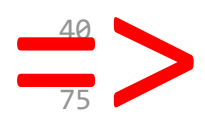


```
#####
# Program: needle
# Aligned_sequences: 2
# 1: LGB1_LUPLU
# 2: HBB_HUMAN
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 187
# Identity: 34/187 (18.2%)
# Similarity: 54/187 (28.9%)
# Gaps: 73/187 (39.0%)
# Score: 39.5
#
#
#=====
```

```
#####
# Program: water
# Aligned_sequences: 2
# 1: LGB1_LUPLU
# 2: HBB_HUMAN
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 117
# Identity: 29/117 (24.8%)
# Similarity: 39/117 (33.3%)
# Gaps: 38/117 (32.5%)
# Score: 50.0
#
#
#=====
```

VERY BAD ALIGNMENTS!!!
NO ALIGNMENT

```
LGB1_LUPLU 1 MGVLTDVQVALVKSSFEEFNA-----NIPKNTHRFFTLVLEIAPGAK 42
                :|...:|:|:| |::.....:|:~|...:
HBB_HUMAN 1 -----MVHLTPEEKSAVTALWGKVNVDDEVGGEALGRLLVVYPWTQ 40
LGB1_LUPLU 43 DLF-SFLKGSSEVP---QNNPDLQAHAGKV-----FKLTYEA 75
                ..| || |...| ..||:~||..|| |.~|...
HBB_HUMAN 41 RFFESF--GDLSTPDAVMGNPKVKAHGKKVLGAFSDGLAHL DNLKGT FAT 88
LGB1_LUPLU 76 AIQLQVNGAVASDATLKSLSGSLVSHGWEAHI----P~VKEAILTKI 120
                ...|:~.....:~||:~|...|| |:~|...:~:~
HBB_HUMAN 89 LSELHCDKLHVDPENFRLLGN~LV~V~AHHFCKEFTPPVQ~AYQ~VVA 136
LGB1_LUPLU 121 EV---VGDKWSEELNTAWTIAYDELAIIIKKEMKDA 154
                .| :~|:~
HBB_HUMAN 137 GVANALAHKYH----- 147
```



```
LGB1_LUPLU 27 THRFFTLVLEIAPGAKDLFSFLKGSSEVP---QNNPDLQAHAGKV----- 68
                |.~|. |~| |...| ..||:~||..||
HBB_HUMAN 39 TQRFFE-----SF--GDLSTPDAVMGNPKVKAHGKKVLGAFS 73
LGB1_LUPLU 69 -----FKLTYEAAIQLQVNGAVASDATLKSLSGSLVSHGWEAHI----P~VKEAILTKI 108
                ...|:~.....:~||:~|...|| |:~|...:~:~
HBB_HUMAN 74 DGLAHL DNLKGT FATLSELHCDKLHVDPENFRLLGN~LV~V~AHHFCKEFTPPVQ~AYQ~VVA 21
LGB1_LUPLU 109 ---P~VKEAILTKI KEV 122
                |.~|:~|...|...|
HBB_HUMAN 122 EFTPPVQAAYQKV VAGV 138
```

NO STRUCTURE

Gene: Leghemoglobin-1

Organism: *Lupinus luteus* (Llobí groc)

Function: It facilitates the diffusion of oxygen to symbiotic bacteroids in order to promote nitrogen fixation.

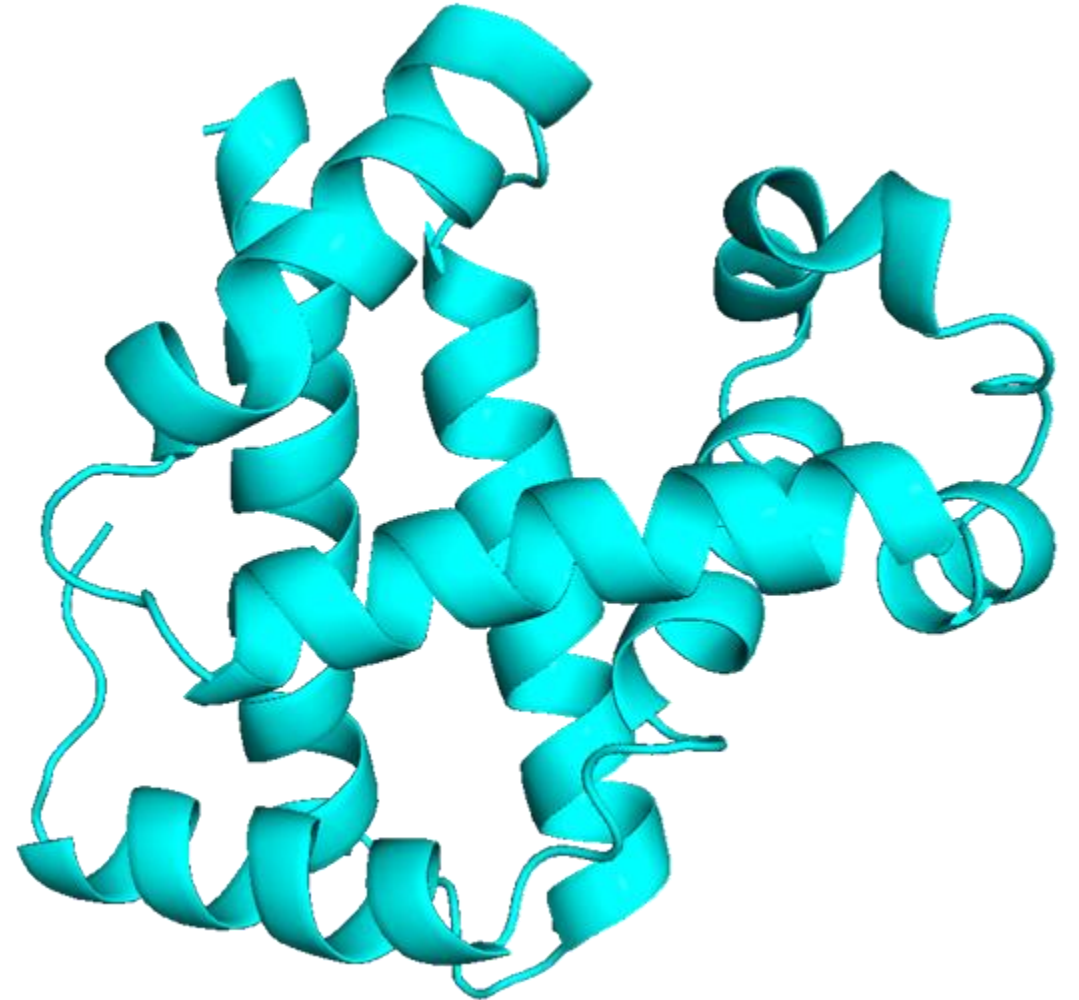
Structure: NO pdb code (No experimentally solved structure)

Gene: Hemoglobin subunit beta

Organism: *Homo sapiens* (Humà)

Function: Involved in oxygen transport from the lung to the various peripheral tissues.

Structure: pdb code: 1IRD chain B

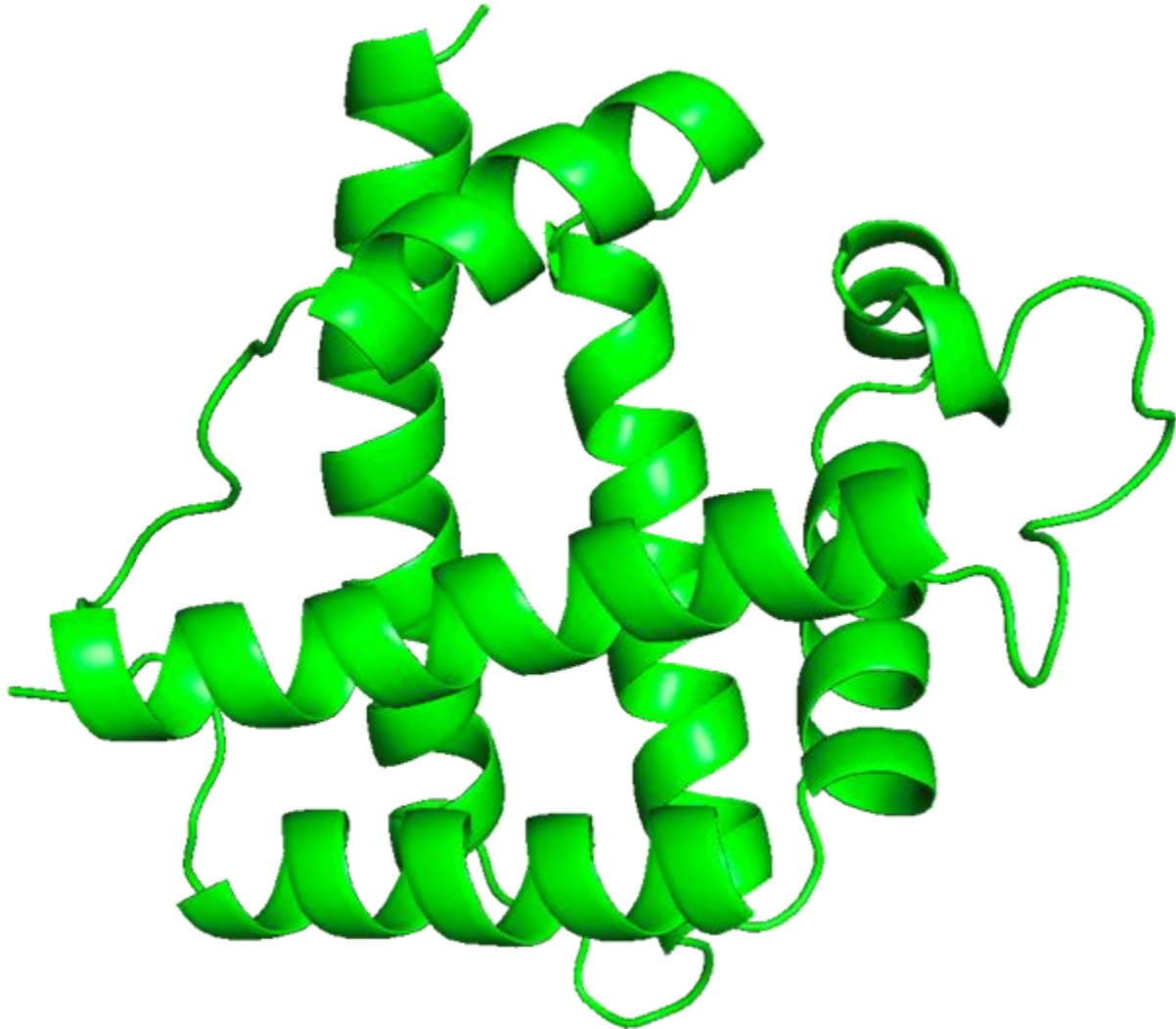


Gene: Leghemoglobin-1

Organism: *Lupinus luteus* (Llobí groc)

Function: It facilitates the diffusion of oxygen to symbiotic bacteriods in order to promote nitrogen fixation.

Structure: AlphaFoldDB: AF-P02239

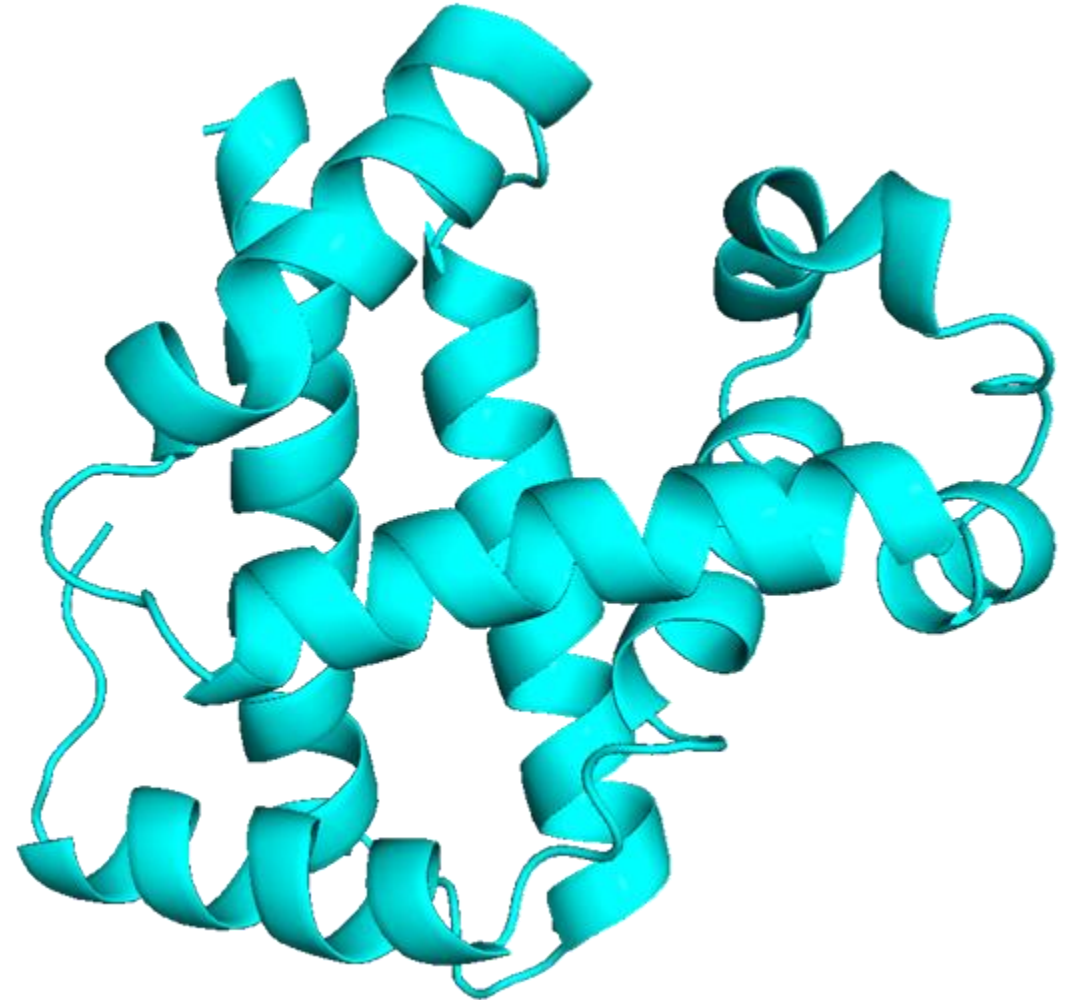


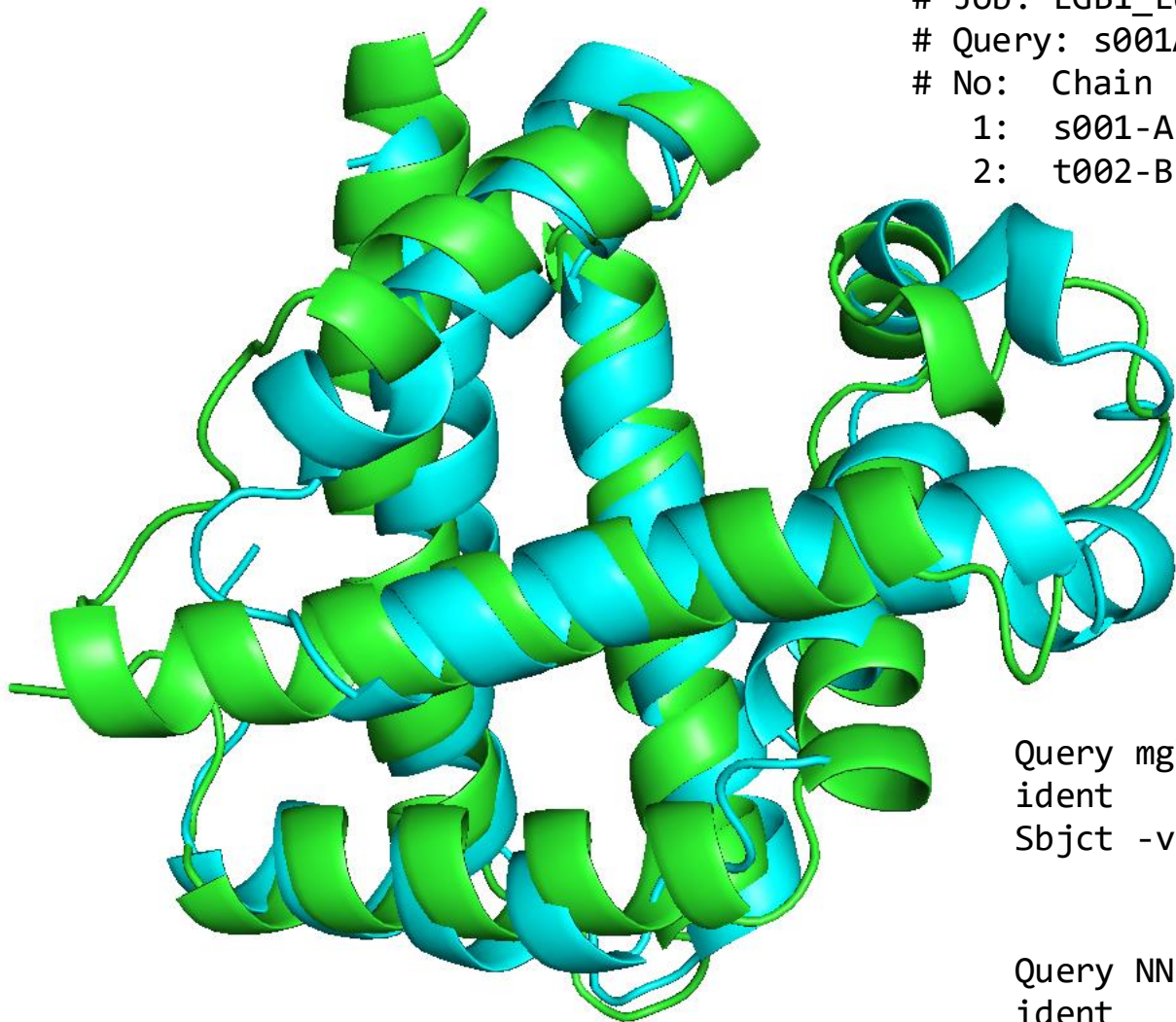
Gene: Hemoglobin subunit beta

Organism: *Homo sapiens* (Humà)

Function: Involved in oxygen transport from the lung to the various peripheral tissues.

Structure: pdb code: 1IRD chain B





Job: LGB1_LUPLU 1IRD

Query: s001A

# No:	Chain	Z	rmsd	lali	nres	%id	PDB	Description
1:	s001-A	32.0	0.0	154	154	100	MOLECULE: LEGHEMOGLOBIN-1;	
2:	t002-B	14.1	2.5	138	146	12	MOLECULE: HEMOGLOBIN ALPHA CHAIN;	

Query	mgVLT	DVQVAL	VKSS	FEEF	naNIP	KNTH	RFFTL	LVLEI	APGAK	DLFS	LKGSS	----	EVPQ	56	
ident															
Sbjct	-vHLT	PEEK	SAVT	ALWG	KV--	NVDE	VGGE	ALGR	LLWY	PWTQ	RFFES	FGDL	Stpda	VMGN	57

Query	NN-PDL	QAHAG	KVFKL	TYEAA	iq	lqvn	GAVAS	datL	KSLG	SVHV-	SKGV	VDAH	FPVV	KEA	114
ident															
Sbjct	PKvKA	HGKK	VLGA	FSDG	LAHL	-----	DNLKG	--t	FATL	SELH	CdKL	HVDP	ENFR	LLGNV	109

Query	ILKTI	KEVV	GDKW	SEEL	NTAW	TIAY	DELAI	IIK	KEMK	daa	154
ident											
Sbjct	LVCV	LAHH	FGKE	FTPP	VQAA	YQK	VVAG	VANAL	AHKY	H---	146



LGB1_LUPLU	1	MGVLTDVQVALVKSSFEEFNA-----NIPKNTHRFFTLVLEIAPGAK	42
		: ...: .: :..:.....: .:. ...:	
HBB_HUMAN	1	-----MVHLTPEEKSAVTALWGKVVNDEVGGGEALGRLLVVYPWTQ	40
LGB1_LUPLU	43	DLF-SFLKGSSEVP---QNNPDLQAHAGKV-----FKLTYEA	75
	: :..	
HBB_HUMAN	41	RF FESF--GDLSTPDAVMGNPKVKAHGKKVLGAFSDGLAHL DNLKGT FAT	88
LGB1_LUPLU	76	AIQLQVN GAVASDATLKSLGSHVSKGVDAHF-----PVVKEAILKTIK	120
		..: ...:.....: : . . :.. . :.. .. :..	
HBB_HUMAN	89	LSELHCDKLHVDPENFRLLGNVLC--VLAHHFGKEFTPPVQAAYQKVVA	136
LGB1_LUPLU	121	EV---VGDKWSEELNTAWTIAYDELAI I I I K K E M K D A A	154
		. :.. :.	
HBB_HUMAN	137	GVANALAHKYH-----	147

Needle: sequence based
18.2% ident.

Query	mgVLTDVQVALVKSSFEEFnaNIPKNTHRFFTLVLEIAPGAKDLFSFLKGSSEVPQ	56
ident		
Sbjct	-vHLTPEEKSAVTALWGKV--NVDEVGGGEALGRLLVVYPWTQRFFESFGDLStpdaVMGN	57

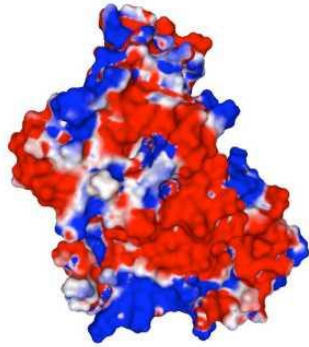
Query	NN-PDLQAHAGKVFKLTYEAA iqqlqvn GAVASdatLKSLGSHVH-SKGVDAHFVVKEA	114
ident		
Sbjct	PKvKAHGKKVLGAFSDGLAHL-----DNLKG--tFATLSELHCdKLHVDPENFRLLGNV	109

Dali: structure based
12% ident.

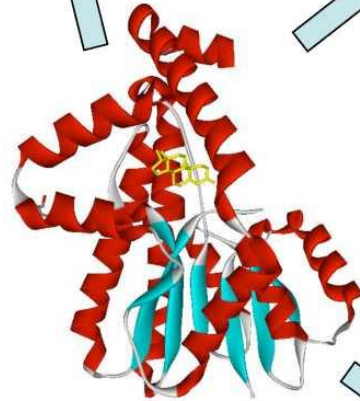
Query	ILKTIKEVVGDKWSEELNTAWTIAYDELAI I I I K K E M K d a a	154
ident		
Sbjct	LVCVLAHHFGKEFTPPVQAAYQKV VAGVANALAHKYH---	146

What can we do with the structural knowledge?

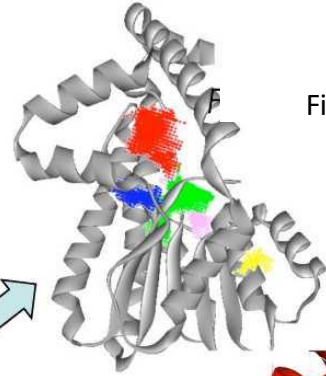
Electrostatics Surface Properties
(Predict interactions)



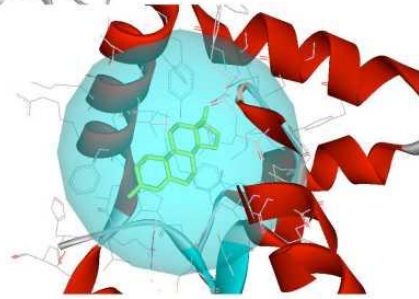
Compare structure of proteins
(molecular evolution)



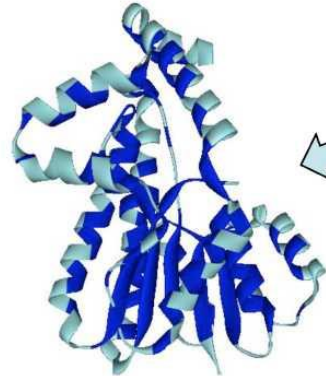
Find cavities within the protein



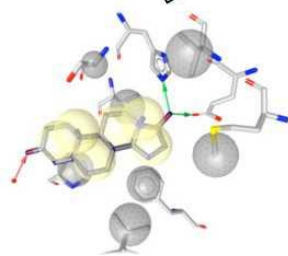
Define ligand binding site regions



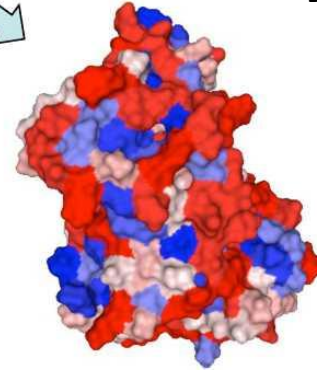
Find exposed-buried residues (design of mutations)



Design drugs



Map hydrophobicity





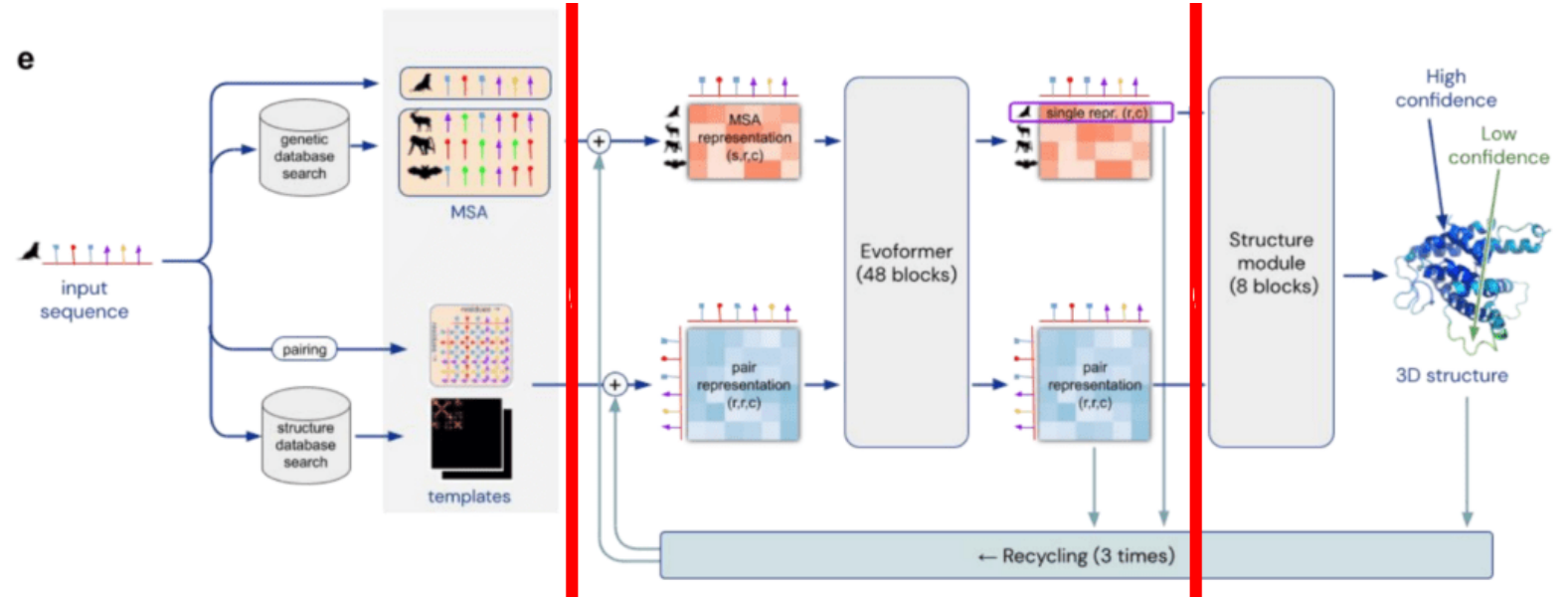
https://www.utep.edu/science/chemistry/_files/images/facilities/cryo%20em%20microscope.png



https://structbio.vanderbilt.edu/nmr/facility/Renovation_2010_11/900_2012_2_small.jpg



AlphaFold: High Accuracy protein structure predictions



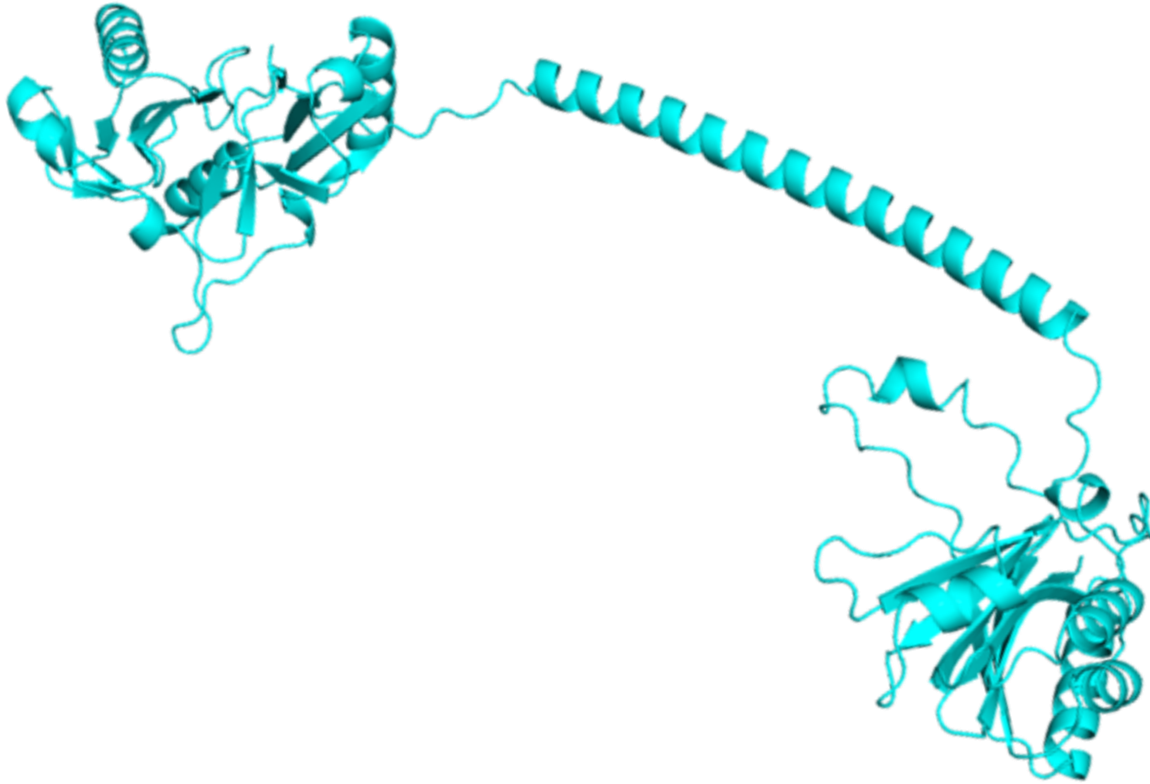
Gene: pilO

Organism: *Burkholderia pseudomallei* (strain K96243)

Function: Unknown function in Type IV pilus type B, variant BFP.

Structure: AlphaFoldDB: AF-Q63JW5

PDB: 4BYZ, 4BZ0 (only N-ter domain)



<https://alphafold.ebi.ac.uk/entry/Q63JW5>

Gene: bfpC

Organism: *Escherichia coli* O127:H6

Function: Unknown function in Type IV pilus type B, variant R64.

Structure: AlphaFoldDB: AF-B7UTD4

PDB: 3VHJ (only N-ter domain)




<https://alphafold.ebi.ac.uk/entry/B7UTD4>

Sequence of AF-Q63JW5-F1 Chain 1: Putative type A

```

1 11 21 31 41 51 61
MSAQVIQIGRQRFVGGLEWQSLSRRLRAEAVELAKKLFDLMLVLRIDRGVAAAGYANTRDGF
71 81 91 101 111 121
APGHLSLGAMVSRRAIALEGAFYNGRRQPAPNWLGAFAALPDGRWAYFAVRDHAFMPNGDWVGSRE
131 141 151 161 171 181 191
EALERLHTDYAWGGWNVVIGEPELERQGEONFQPKRLDDLPRRGGRPRTERWWALRPVERRLS
201 211 221 231 241 251

```



Structure Tools

Structure

AF-Q63JW5-F1

Type Model

Nothing Focused

Quick Styles

Default Illustrative Stylize Current

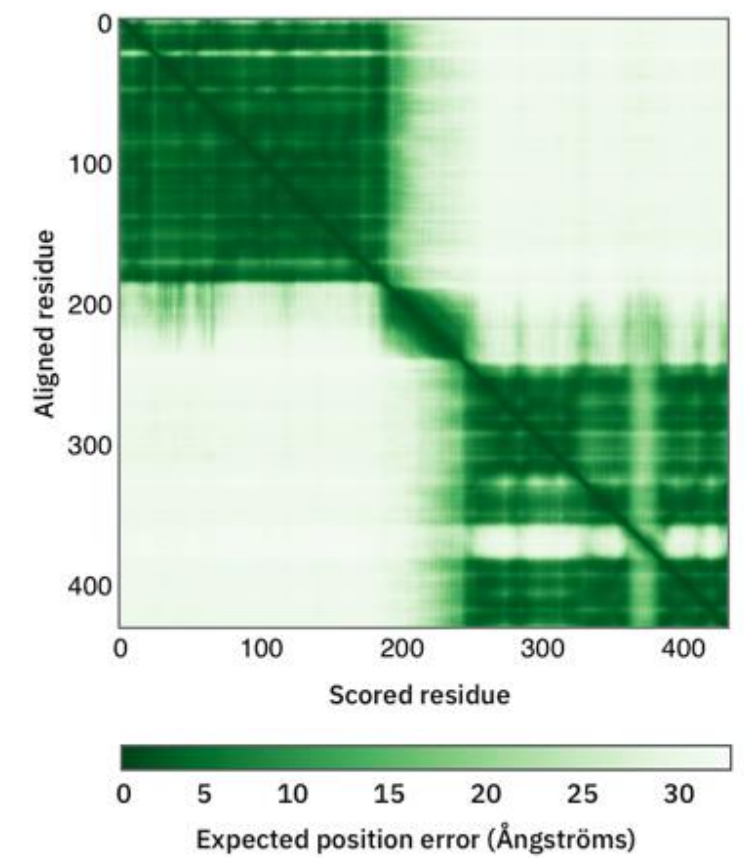
Components AF-Q63JW5-F1

Preset + Add

Polymer Cartoon

Measurements

+ Add



Model Confidence

■ Very high (pLDDT > 90)
 ■ High (90 > pLDDT > 70)
 ■ Low (70 > pLDDT > 50)
 ■ Very low (pLDDT < 50)

AlphaFold produces a per-residue model confidence score (pLDDT) between 0 and 100. Some regions below 50 pLDDT may be unstructured in isolation.

Predicted aligned error (PAE)

Click and drag a box on the PAE viewer to select regions of the structure and highlight them on the 3D viewer.

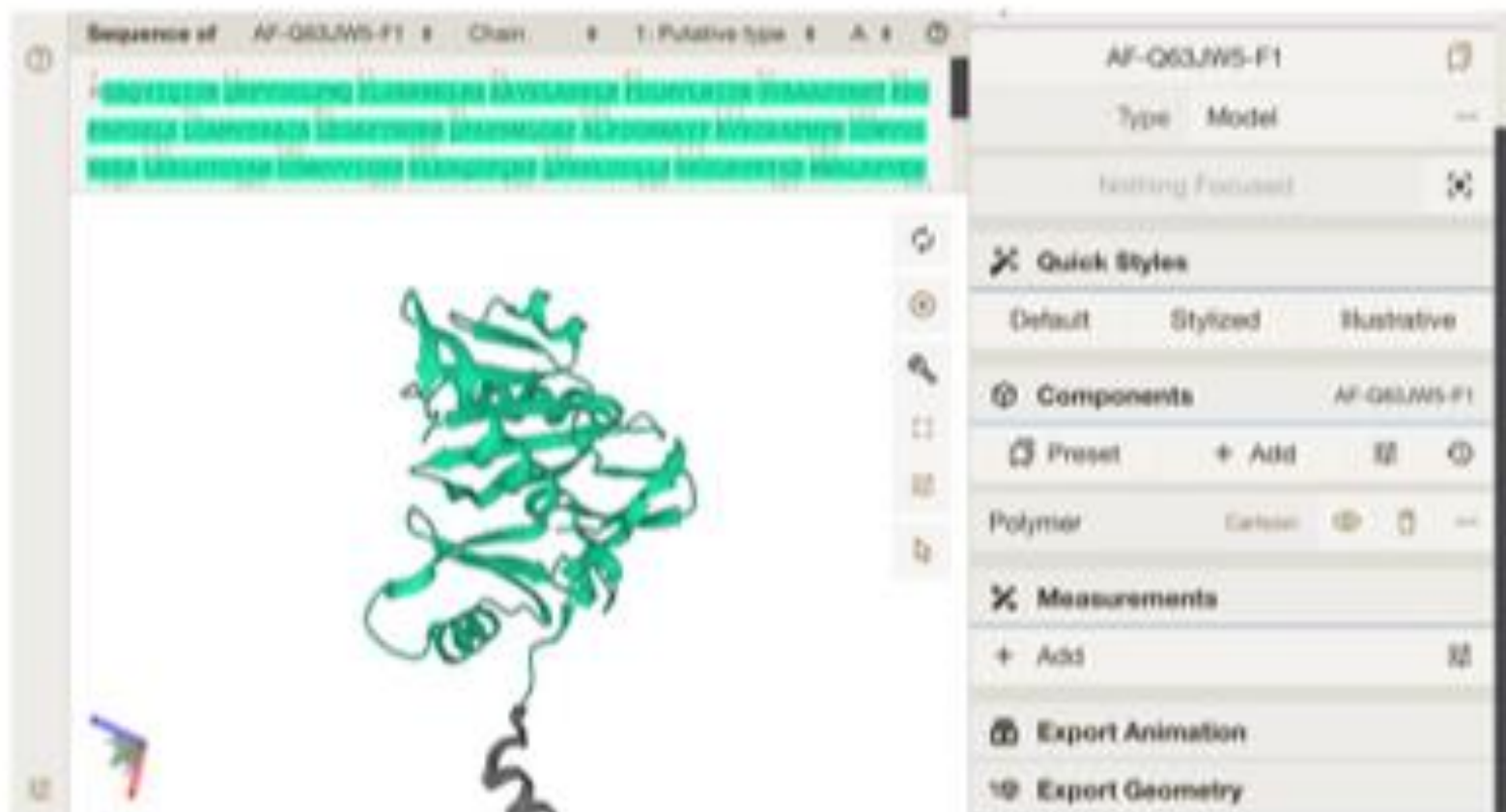
PAE data is useful for assessing inter-domain accuracy – [go to Help section below](#) for more information.

[Visit our online training course](#)

How to interpret the Predicted Aligned Error

Click and follow the tutorial "How to interpret the Predicted Aligned Error" to better understand this graphic / score

3D viewer

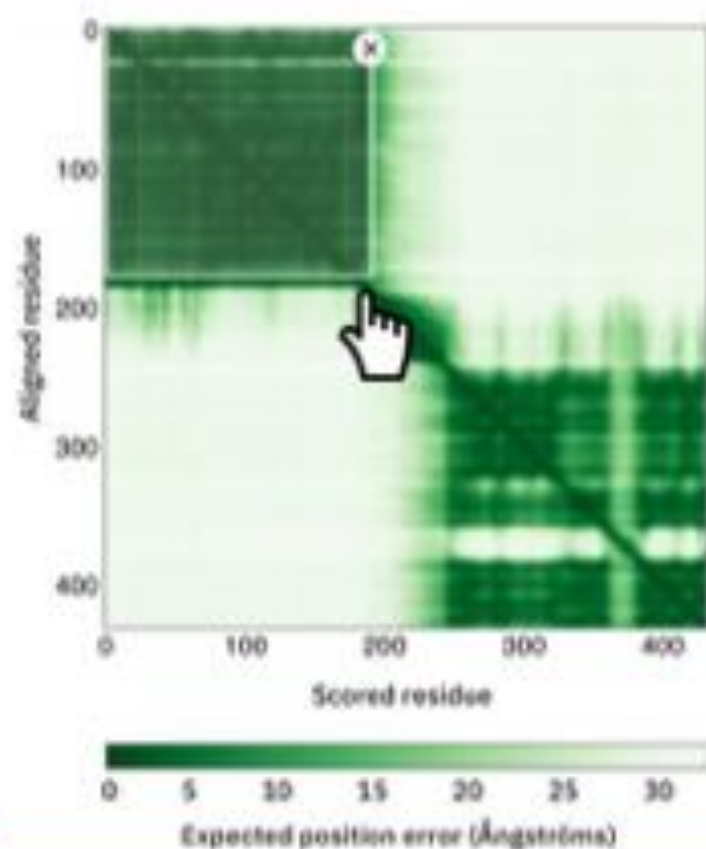


Model Confidence

- Very high (pLDDT > 90)
- High (90 > pLDDT > 70)
- Low (70 > pLDDT > 50)
- Very low (pLDDT < 50)

AlphaFold produces a per-residue model confidence score (pLDDT) between 0 and 100. Some regions below 50 pLDDT may be unstructured in isolation.

Predicted aligned error (PAE)



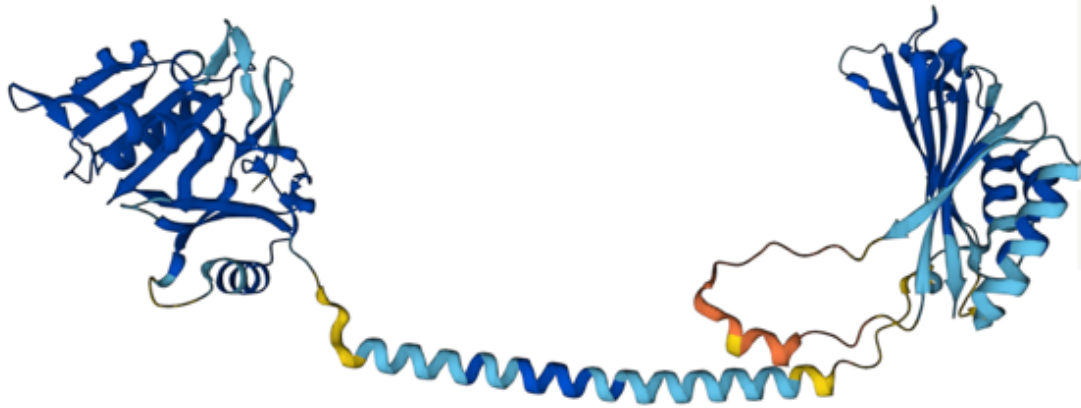
Click and drag a box on the PAE viewer to select regions of the structure and highlight them on the 3D viewer.

PAE data is useful for assessing inter-domain accuracy – [go to Help section below](#) for more information.

B7UTD4



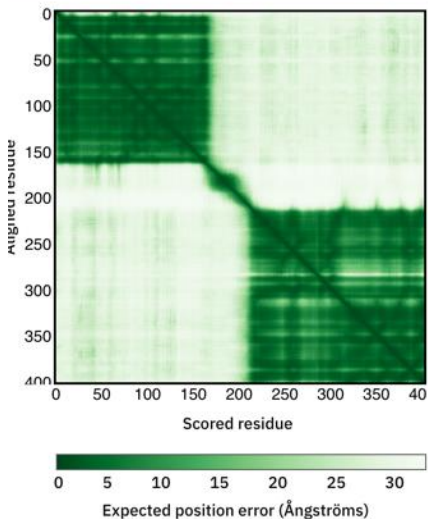
Q63JW5



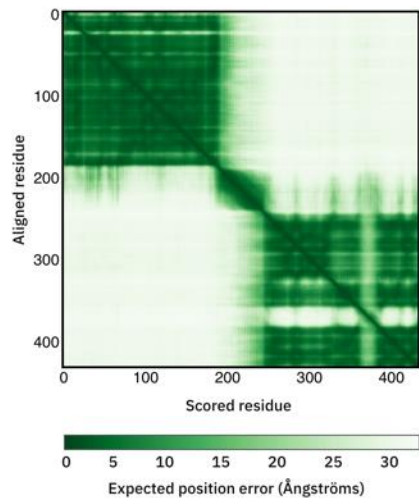
Model Confidence

- Very high (pLDDT > 90)
- High (90 > pLDDT > 70)
- Low (70 > pLDDT > 50)
- Very low (pLDDT < 50)

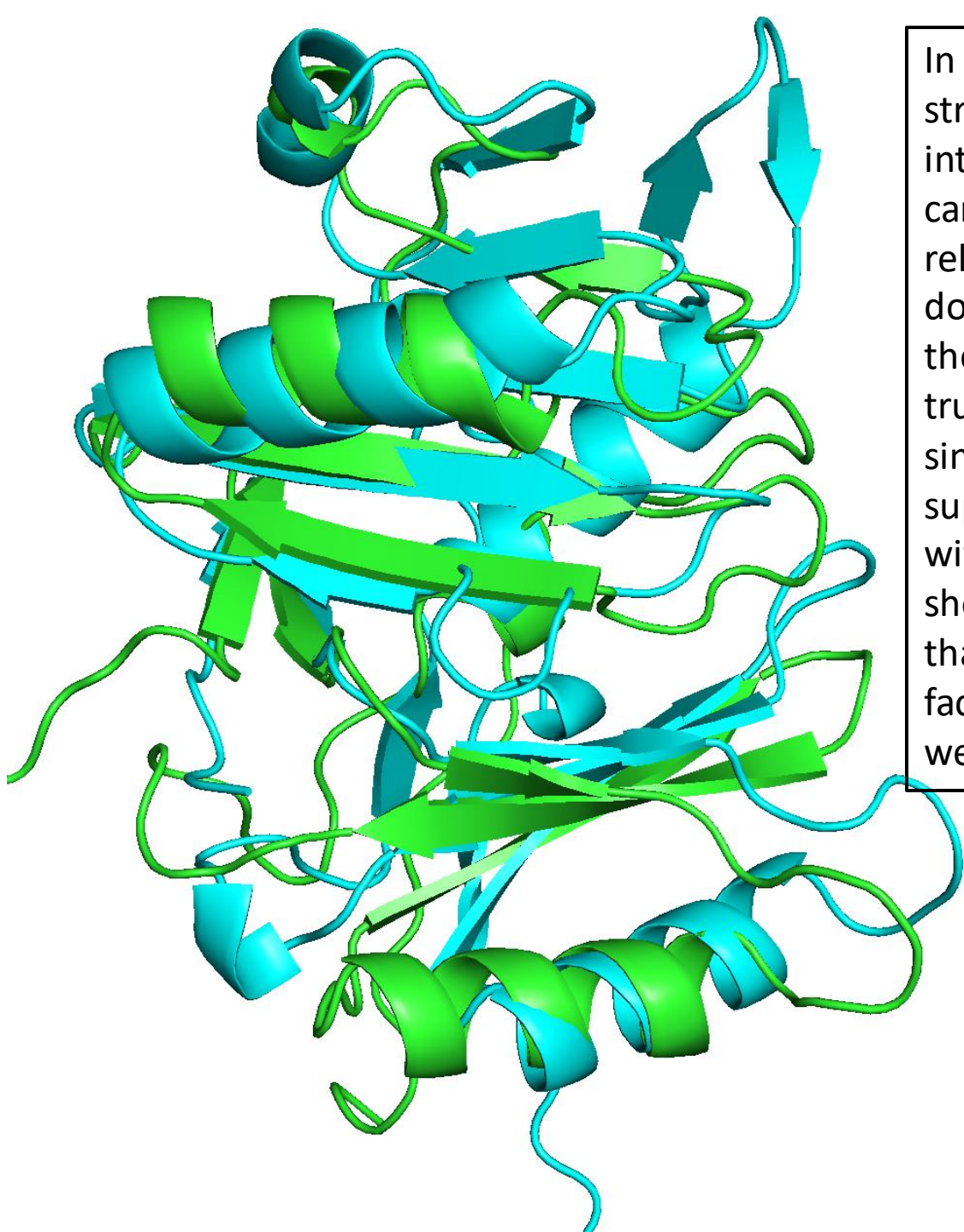
Predicted aligned error (PAE)



Predicted aligned error (PAE)



In both cases, since the structure of the interdomain region cannot be trusted, the relative position of one domain with respect to the other cannot be trusted either. This way, since each domain superpose very well with its counterpart, it should be considered that both structures, in fact, superpose very well.



In both cases, since the structure of the interdomain region cannot be trusted, the relative position of one domain with respect to the other cannot be trusted either. This way, since each domain superpose very well with its counterpart, it should be considered that both structures, in fact, superpose very well.

