

/Models/boltz\_results\_2I25\$ tree

# Carpetas generales:

- lightning\_logs # registros de ejecución
- msa # Aparecen tantos csv como cadenas proteicas únicas
- predictions # estructuras predichas
- processed # datos procesados para inferencia

---

# Carpetas extendidas:

- lightning\_logs # registros de ejecución
  - version\_0
    - hparams.yaml # todos los hiperparámetros utilizados
- msa # tantos .csv como cadenas proteicas únicas; sin orden por longitud
  - 2I25\_0.csv # key,sequence: ids: 0-27; desde 28 sólo aparecen -1
  - 2I25\_1.csv # ídem (Ver figura 1); son los MSAs per se
  - 2I25\_paired\_tmp\_pairgreedy-env # resultados de alineamientos emparejados considerando relaciones evolutivas entre pares
    - out.tar.gz # (contiene los otros 2 archivos)
    - pair.a3m # alineamiento de secuencias emparejadas
    - pair.sh # script que genera ese alineamiento
  - 2I25\_unpaired\_tmp\_env # alineamientos no emparejados (sin asumir relaciones)
    - out.tar.gz # (contiene los otros 4 archivos)
    - bfd.mgnify30.metaeuk30.smag30.a3m # alineamiento combinado de varias bases de datos
    - uniref.a3m # alineamiento contra UniRef
    - pdb70.m8 # coincidencias con estructuras conocidas (PDB70)
    - msa.sh # script que genera estos alineamientos
- predictions # estructuras predichas (tantas como se hayan pedido)
  - 2I25
    - 2I25\_model\_0.cif # modelo legible por visualizadores pdb/cif
    - confidence\_2I25\_model\_0.json # Q características (Ver Cod. 1)
    - plddt\_2I25\_model\_0.npz # (Ver figura 2)
- processed # datos procesados para inferencia
  - manifest.json # da los id de cadenas, longitud y mas
  - constraints # las restricciones específicas del modelo
    - 2I25.npz # probablemente distancias o ángulos
  - msa # representación estructurada en binario de los msa
    - 2I25\_0.npz # (secuencias), (deleciones), (residuos) de los msa
    - 2I25\_1.npz # (index, id\_seq, start, end, X, Y)
  - structures
    - 2I25.npz # (Ver Fig. 3)

12 directories, 19 files

## ANEXOS

```

1  key, sequence
2  0, RVDQTPQRIKTESLTINCVRDRCVLTSGYWRKPPGSRNEESISDGGRYVETVNRGSKSFSRLINDLTVKDSGTYRCKPESRYGSYDAVCAALNDQYGC
3  1, ---QTPSLSRKECESVSIQCVFEASGVTFKSMEFFRQIGKGTQLVSIIPVDGRFVVKADRVSAFSLEINELEVQDGATYYCR-----
4  2, -----VSVQNGEKVTLNCTYETASSEPSL-HWYIQRPGVRphfillrhrqfSKEEDEPGGKYSSKNKESKSIDLRISGVSVSDSGLYYCAVSS-----
5  3, -----EYGEKVTLNCTYETTN---SDPYlhWYIQRPGERphfillrhrqfSKEEDEPGGKYSAKLNKESKSVDLRISGVSESDSGLYYCaIRPTVSLSA
6  4, ----PHSVSESPGKTVTISCTRSSGSIASNYVQWYQQRPGSSpttviyeyNQRPSSGVDPDRFSGSIDRSSNSASLTISGLQTEDEADYYCQ-----SYD----
7  5, KVTQTPDISVRDTEVTTLNCMYdTSDSNYL----FWYKQPPSRemiliirgeayRQENATNN--RFSVNFQKGNKSFSRLISDSRLEDAATYFC-----
8  6, KVIQTQNTVTKQEGEAATFNCNYETTWSDYYTLWYKQPPGGemifliyQDENKPNKQGRLSFSFQKAASISLTISLQLTDSATYFC-----
9  7, -VTQSPSSVSKQEGETVTLHCsfTLNYYVYVMS--WFRQLPNGKMTETIIylyssstnSKEGRYSVSHQRGNAISLTITGLMPTDSGTYFC-----
10 8, -----TGGSLRLTCVASGFTFSDYSMMWVRQAPGKGqlwvayIRHDGISTGyadsvkGRFTISRDSAKNTLYLQMNSLKTEDSATYVC---ARDGDY
11 9, KVTQDQPTVSSQVGEVTLNC---RYEAPWNTNYlfWYKQPPSGEmtflihqdTskTNAKDGRYSVNSRRGEKSLSLTISALQLKDSATYFC-----
1210, -----EAGGSLRLSCVVTGSSFSSTMAWYRQPPGKQREwvASFtSGGaiKYTDSVkgRftmsrDNAKMTYLMENLKPEDTAVYYCALHN-----
1311, -----QAGETLRLSCTASGFTFDDSDMGWYRQAPGNECElvSSIsSDgstyyadsvkGRFTISRQNAKNTVYLMNSLKPEDTGVYYCAAEGH--RY
1412, KVIQDQPDISSRVGESVTLNCRYETSRSI----FWYKQLP-SGEMIFLTRGGYYSINVERSHKSSSLTISTLQLEDsAKYFC-----
1513, -VEQHPSTLSVQEGDSSVINCTYSDS---FSDYFpWYKQEPGQGLqlfsmalGVKEDGRLKSTVNSQERYSTLHISATQLEDsGTYLCAEAQ-----
1614, QVRQSPQSLTVWEGETAALNCSYEDS--TFNYFPWYQQFPGEgallisIRSVSDKKEDGRFTIFFNKREKKLSLHITDSQPGDSATYFCAASETSGSWQLI-
1715, KVEQSPLSFEVQEGGNVTNCTYTDTAfVYL--FWYQYQPGKGPQllmrmlsNMVTDhQGRFTIHLNKAQLFSLHISASQSEDSGTYFCAASTR--SSLCTC
1816, KVEQT-QFLRVQEGDSATLNCNYTDS--IASYFLWYHQSPGKgpvlriAGYSSGDEEGRTISVNKEAKQFSLHISASQPVDSGTYFC-----
1917, -----GGSLRLACVASESVFEMYTVAWYRQAPGKQRElvAGITDEGRtnYADfvkgrfTISRdnSKKTVHLQMDNLNPEDTAVYYCKLEHDLGY
2018, KVIQTKNTVTKQEGEAATFNCNYETTWSDYYTLWYKQPPGGemifliyQDENKPNKQGRLSINQKAASISLTISLQLTDSATYFC-----
2119, RVEQTPTTTTKETGESLTINCVRDSSCALDSTYWFYTKKGATKESLNGGRYAEVTKNASKSFSRLISDLRVEDSGTYHCR---AYSLSAGMCAWMGYIEC
2220, -VEQSPASLPVPEGASASLNCNYSDS---FSTDftWYRQYPGKgpvllyTSSEKPIED-GRFTVQVNRASKHVSLLIRDSQPSDSATYLC-----
2321, -VEQHPESLSVPEGAMASLNCYTRDS--ASQNfGwYRQYPGKGLKLLVStfssgnkDEGRYTAHLNKASRYFSLHIRGSQPSDSATYLC-----
2422, RVAQDQPDISSQVGEVTLNCQY-DTRLSSFRIFWYKQLPsgemiyligQVSSSQKARDGRYTTDLQRSRKSISLTISDLQLEDsAKYFC-----
2523, KVTQAQTIITRREGEAVTMGCTYEISWGSYYDLWYKQPPSGemvfllyqNRDKPSAQ-QGRFSVNFQKGKKSISLTISPLQLADsATYFC-----
2624, KVIQPEKAVSVRAGESVTLNCTVTSR-PLGSIKWFRGTGQSRqliysfTEEFPRVTYVTDIAIKRYNLDFsIRISNVTIHDSGTYC---VKFQKADADKEF
2725, -VEQTPRTLTKETGESLTINCVLKDASYALRGTSWYLTkLNATTWSRISIGGRYSETVNKGLKSFSRLRIDLRLVEDSGAYHCE-----
2826, KVQQNPEFLSVPEGAMTSLNCTFTDS--ASQSFwWYRQQPGkspkalisifsnGEKEEGTKEDGRLKSTFDSKERYSTLHIRDAQLEDsGTYLCAEAQ---
2927, -VNQTPAVMNGLEGDTLTIECrfkVVKDSSTYSV-TWYKK-GADGHEKELKNGTGLVTTAlDsSKGLRSLSLTVKKADVTDSGTYVCGVGTKHGMYP-----
30-1, RVDQTPQRIKTESLTINCVRDRCVLTSGYWRKPPGSRNEESISDGGRYVETVNRGSKSFSRLINDLTVKDSGTYRCKPESRYGSYDAVCAALNDQYGC
31-1, RVDQTPQITKETGESLTINCVRDSSCALDSTYWFYTKKGATKESLNGGRYAEVTKNASKSFSRLISDLRVEDSGTYHCR---AYSLSAGMCAWMGYIEC
32-1, RVDQTPQITKETGESLTINCVRDSSCALDSTYWFYTKKGATKESLNGGRYAEVTKNASKSFSRLISDLRVEDSGTYHCR---AYSLSAGMCAWMGYIEC

```

Figura 1. Interior del archivo msa/2I25\_1.csv.

```

data = load('icvv/home/csic/sbf/vmh/BOLTZ_1x/AB1/Models/boltz_results_2I25/predictions/2I25/plddt_2I25_model_0.npz')
lst = data.files
for item in lst:
    print(item)
    print(data[item])

```

✓ 0.1s Python

plddt					
0.7612842	0.9068771	0.94824696	0.9725381	0.9680596	0.9706423
0.9602358	0.97132635	0.958546	0.9758326	0.97511446	0.97169536
0.97061616	0.97756326	0.97286	0.96228313	0.944122	0.97577107
0.9796241	0.97534835	0.96418667	0.95607495	0.8958936	0.8604226
0.7027489	0.7194738	0.56114805	0.7026833	0.7345652	0.87450594
0.8281967	0.8850028	0.9365904	0.9386676	0.9746771	0.9660869
0.9623868	0.9495667	0.9378345	0.9244611	0.9004381	0.90303594
0.90400964	0.90901744	0.8972373	0.9096116	0.9380691	0.95937145
0.8932995	0.89719	0.9390491	0.9571451	0.9639738	0.97227156
0.96591437	0.9540947	0.95512444	0.93892765	0.946565	0.85553265
0.87580013	0.9167514	0.8996631	0.96005255	0.9686413	0.9711722
0.9823778	0.97832435	0.97725844	0.97713983	0.9703944	0.977808
0.9639206	0.9690422	0.9495832	0.9726666	0.9749501	0.97603434
0.96915096	0.98388654	0.9576441	0.96926254	0.94238716	0.9154589
0.8655267	0.7678782	0.6415212	0.5454715	0.487513	0.41095215
0.3361078	0.5040483	0.41799015	0.48809415	0.6300902	0.63070846
0.5987817	0.7430789	0.7713666	0.74821305	0.8264088	0.91259193
0.9555098	0.95976186	0.9724419	0.9803446	0.9756682	0.97940326

Figura 2. Interior del archivo plddt\_2I25\_model\_0.npz.csv.

```

atoms
[[[46, 0, 0, 0], 7, 0, [0., 0., 0.], [-2.0170422, 0.6717798, -1.1794233], True, 0)
 ([35, 33, 0, 0], 6, 0, [0., 0., 0.], [-2.0503085, -0.5735037, -0.40972203], True, 2)
 ([35, 0, 0, 0], 6, 0, [0., 0., 0.], [-3.4694405, -1.0612813, -0.27558324], True, 0)
 ...
 ([35, 39, 0, 0], 6, 0, [0., 0., 0.], [-0.99162674, 0.52349573, 0.06723011], True, 0)
 ([35, 36, 17, 0], 6, 0, [0., 0., 0.], [-2.4228058, 0.29949337, 0.5730421], True, 0)
 ([35, 36, 18, 0], 6, 0, [0., 0., 0.], [-1.0282856, 1.1250265, -1.3460144], True, 0)]
bonds
[]
residues
[('ARG', 3, 0, 0, 11, 1, 4, True, True)
 ('VAL', 21, 1, 11, 7, 12, 15, True, True)
 ('ASP', 5, 2, 18, 8, 19, 22, True, True)
 ('GLN', 7, 3, 26, 9, 27, 30, True, True)
 ('THR', 18, 4, 35, 7, 36, 39, True, True)
 ('PRO', 16, 5, 42, 7, 43, 46, True, True)
 ('GLN', 7, 6, 49, 9, 50, 53, True, True)
 ('ARG', 3, 7, 58, 11, 59, 62, True, True)
 ('ILE', 11, 8, 69, 8, 70, 73, True, True)
 ('THR', 18, 9, 77, 7, 78, 81, True, True)
 ('LYS', 13, 10, 84, 9, 85, 88, True, True)
 ('GLU', 8, 11, 93, 9, 94, 97, True, True)
 ('THR', 18, 12, 102, 7, 103, 106, True, True)
 ('GLY', 9, 13, 109, 4, 110, 110, True, True)
 ...
interfaces
[]
mask
[ True  True]
Output is truncated. View as a scrollable element or open in a text editor. Adjust cell output settings...

```

Figura 3. Interior del archivo constraints/2I25.npz. Al ser archivo NumPy se ha abierto con el mismo código.

```

''' json
{
  "confidence_score": 0.7400132417678833,
  "ptm": 0.5552749633789062,
  "iptm": 0.08926191180944443,
  "ligand_iptm": 0.0,
  "protein_iptm": 0.08926191180944443,
  "complex_plddt": 0.9027010202407837,
  "complex_ipddt": 0.7742644548416138,
  "complex_pde": 0.6638683080673218,
  "complex_ipde": 9.526134490966797,
  "chains_ptm": {
    "0": 0.9617125391960144,
    "1": 0.9778867959976196
  },
  "pair_chains_iptm": {
    "0": {
      "0": 0.9617125391960144,
      "1": 0.07789546996355057
    },
    "1": {
      "0": 0.08926191180944443,
      "1": 0.9778867959976196
    }
  }
}
'''

```

```
    }  
  }  
}
```

Código 1. Características (estadísticos) globales de calidad del modelo.  
Claves destacadas en amarillo.