Hidden errors in cryo-EM models of macromolecules

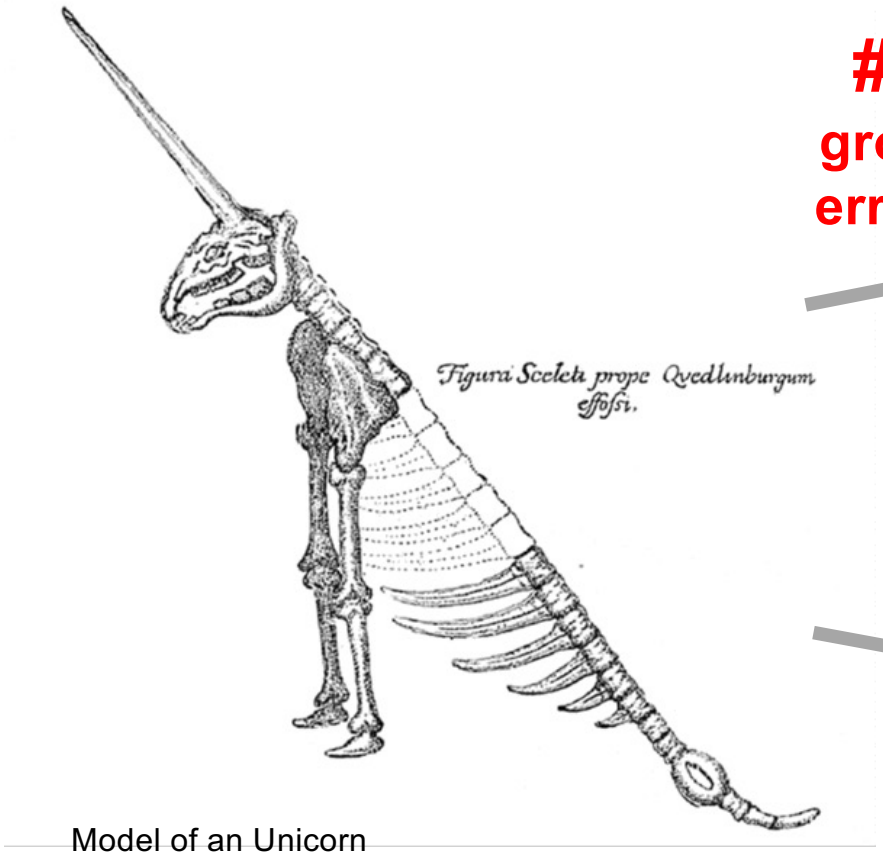
Grzegorz Chojnowski
EMBL Hamburg

CCP-EM Icknield Model Building Workshop

09.10.2023

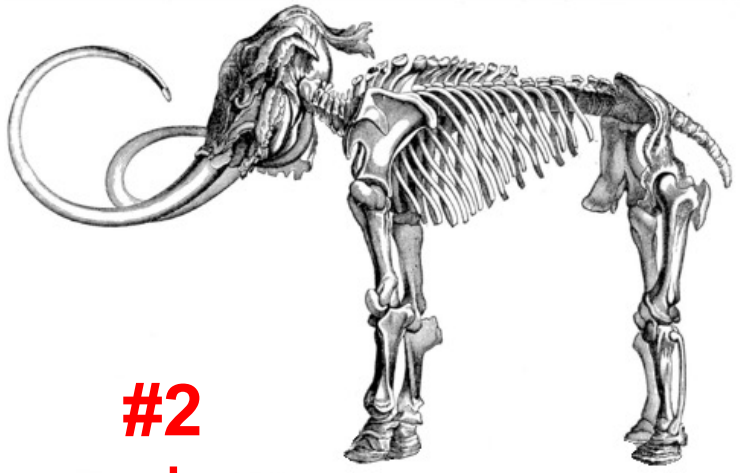
EMBL

Model building traps in EM and MX

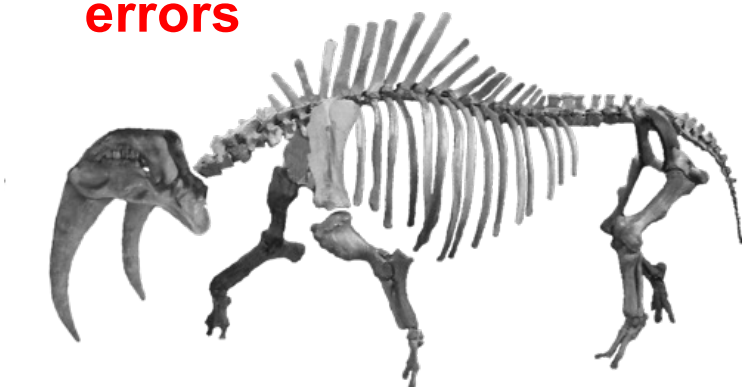


Model of an Unicorn
Gottfried Leibniz after Otto von Guericke, *Protogaea* (1719)

#1
gross errors

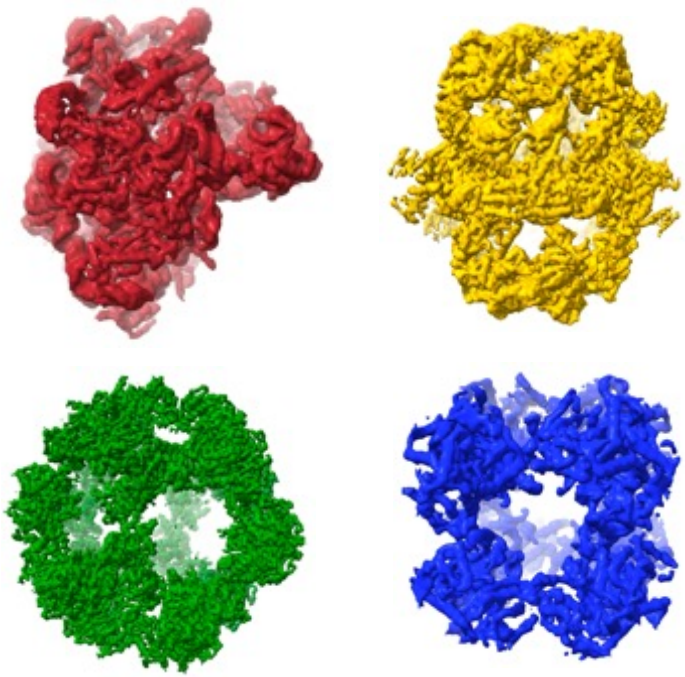


#2
random errors

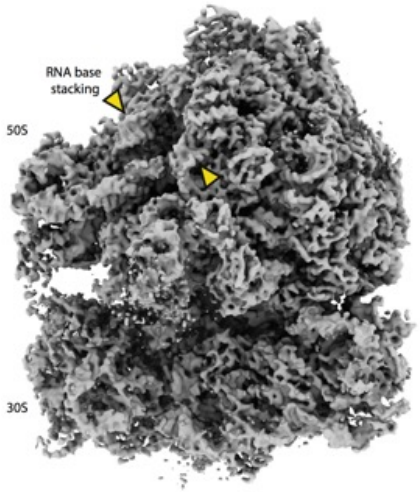


EMBL

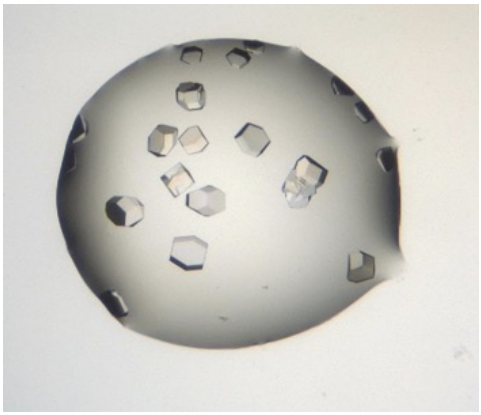
Unknown proteins and gross errors in EM and MX



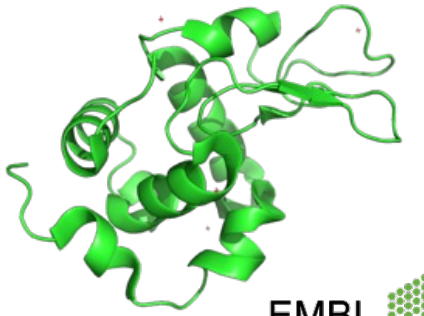
Cryo-EM and artificial intelligence visualize endogenous protein community members
Skalidis et al. Structure 2022



M. pneumoniae 70S ribosome at 3.5 Å refined from in situ tilt-series data
Tegunov et al. 2021

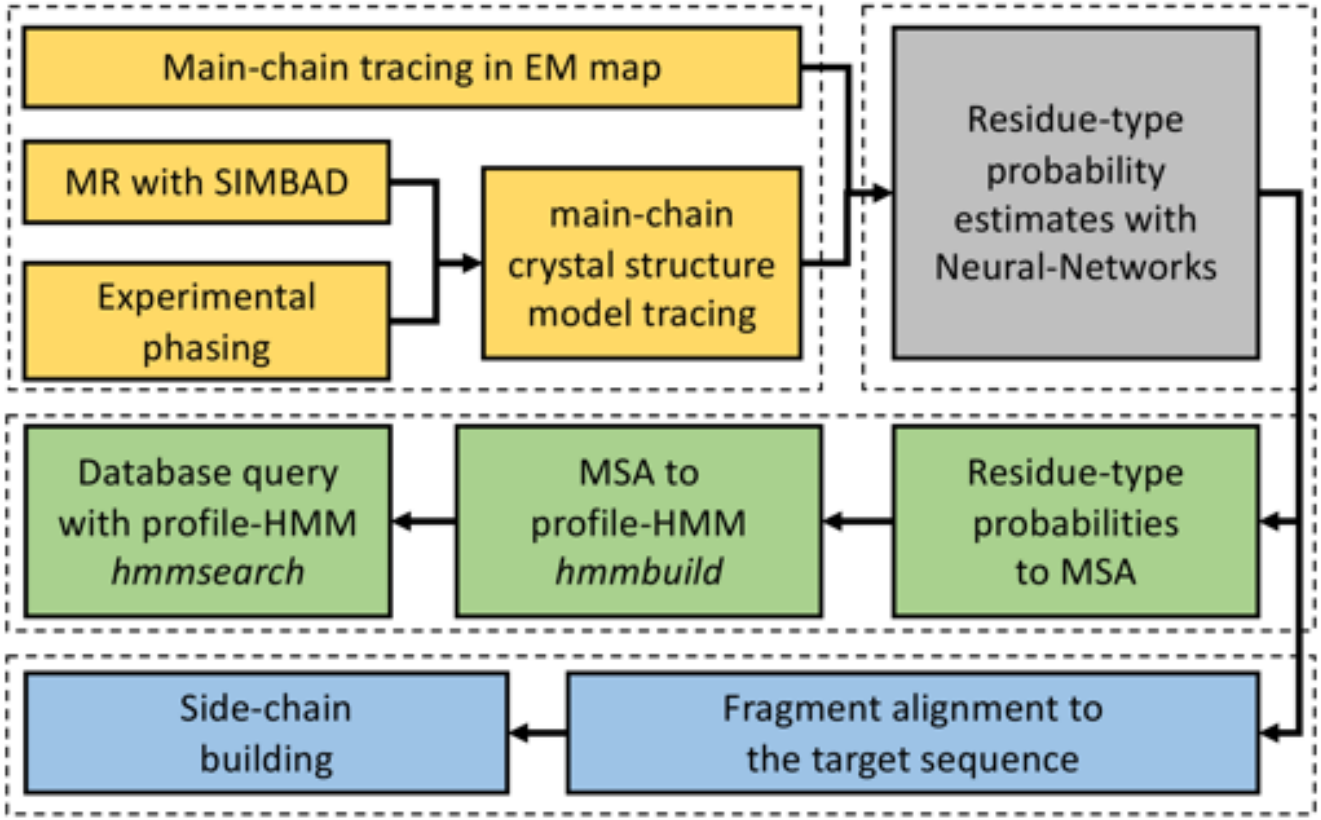


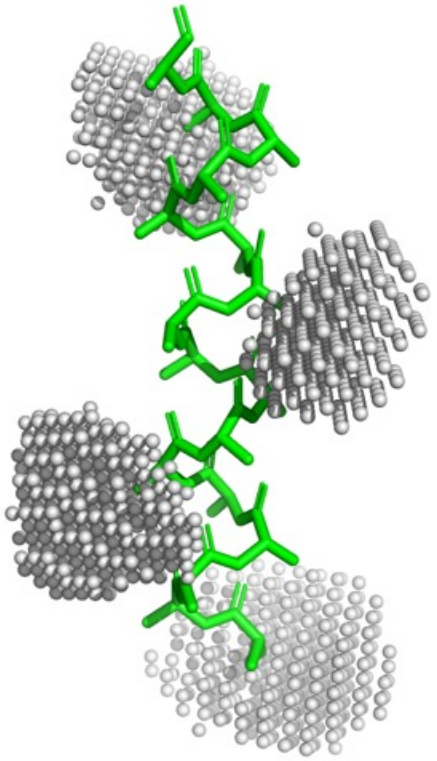
?



EMBL

Protein sequence identification with findMySequence





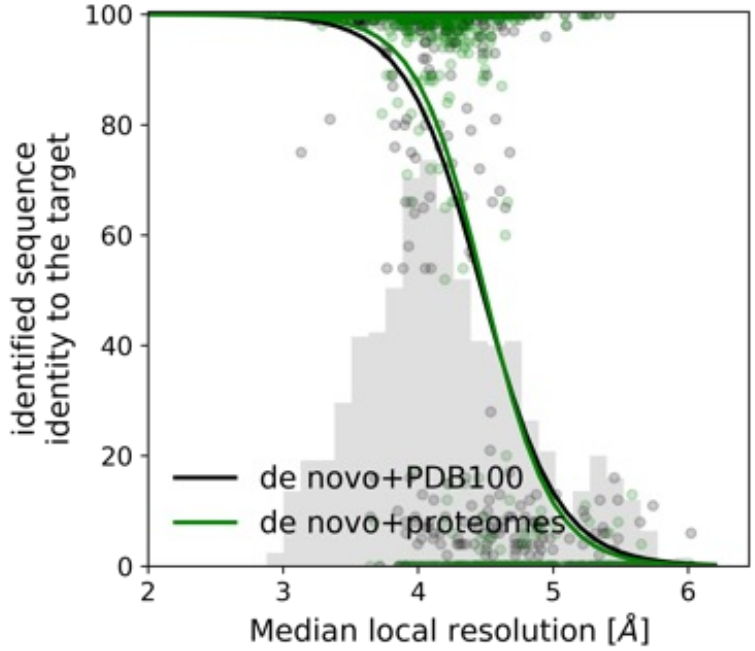
residue-type NN classifier
328 input features
111,800 parameters

Chojnowski et al IUCrJ 9.1 (2022)

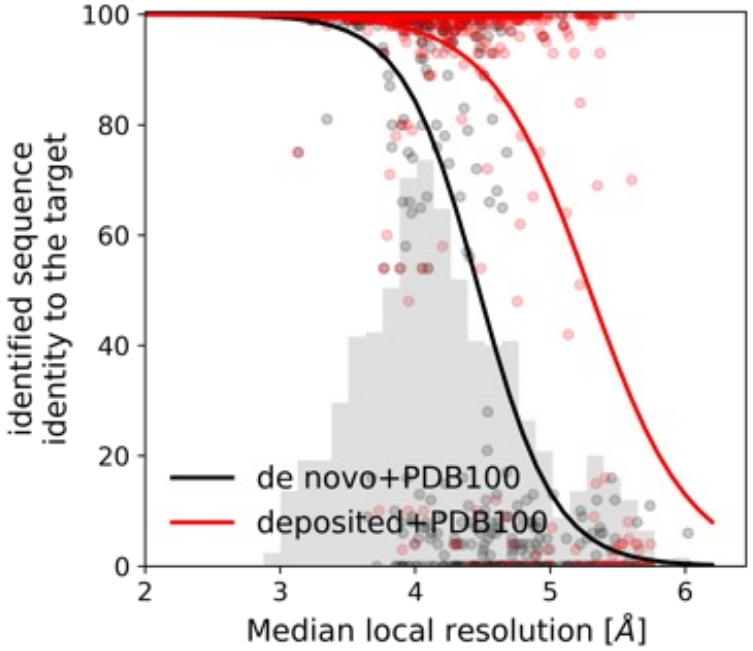
EMBL

Resolution, model quality, and sequence DB size

models built *de-novo* with ARP/wARP
with different sequence DB sizes



models built *de-novo* with ARP/wARP
and deposited coordinates

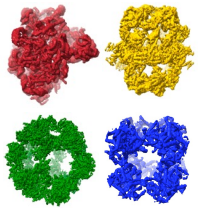


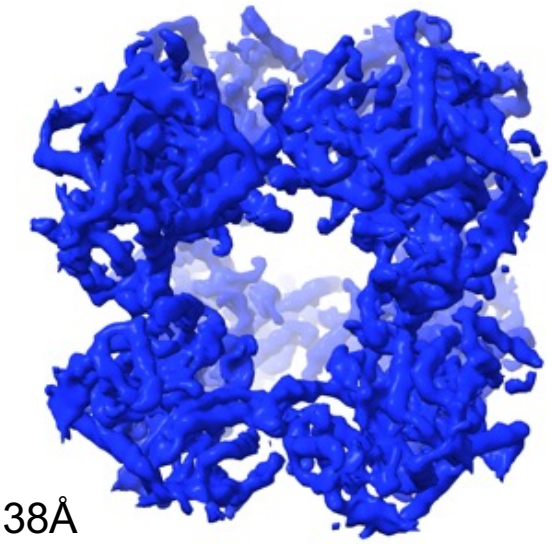
Chojnowski et al IUCrJ 2022

EMBL

findMySequence: sequence validation in EM

A tale of two dehydrogenases from *Chaetomium thermophilum* native cell extracts





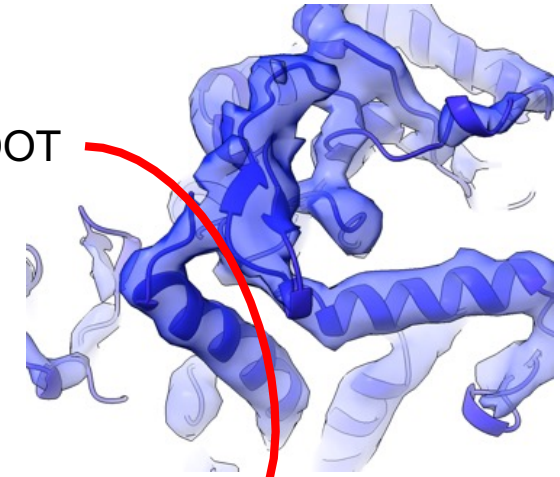
4.38Å

omokage search

>E2b ketoacid
MRSWFF
SRVLRQ
QWRVVR

(~20% SI)

AF2+COOT



>E2o oxo-acid
MLYRGL
RMAARA
APKFAV

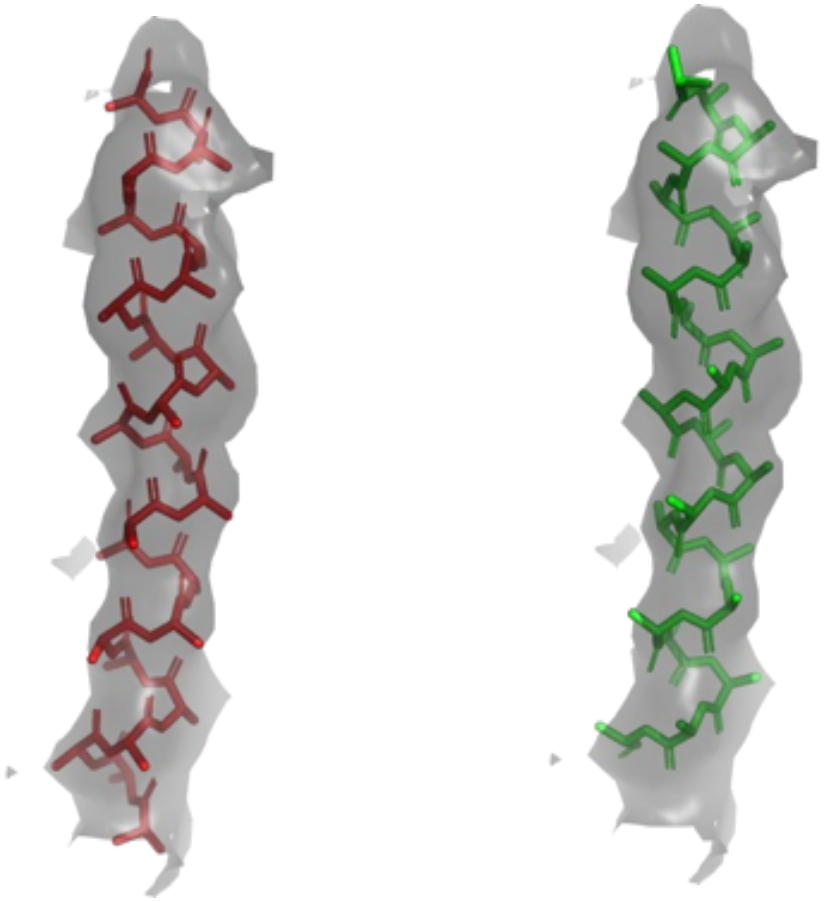
AF2+COOT

findMySequence

with **Panos Kastritis**
Chojnowski et al IUCrJ 2022
Skalidis et al Structure 2022

EMBL

findMySequence: model building



ESX-5 hexameric pore complex at 3.4 Å resolution

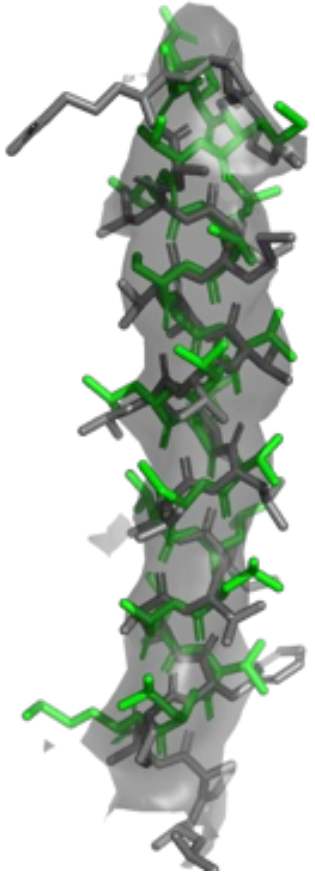
22aa α-helix models
built with COOT
(two alternative directions)

Beckham, Ritter, Chojnowski et al *SciAdv* (2021)

EMBL

findMySequence: model building

p-value = 0.1



p-value = 0.99

Beckham, Ritter, Chojnowski et al *SciAdv* (2021)

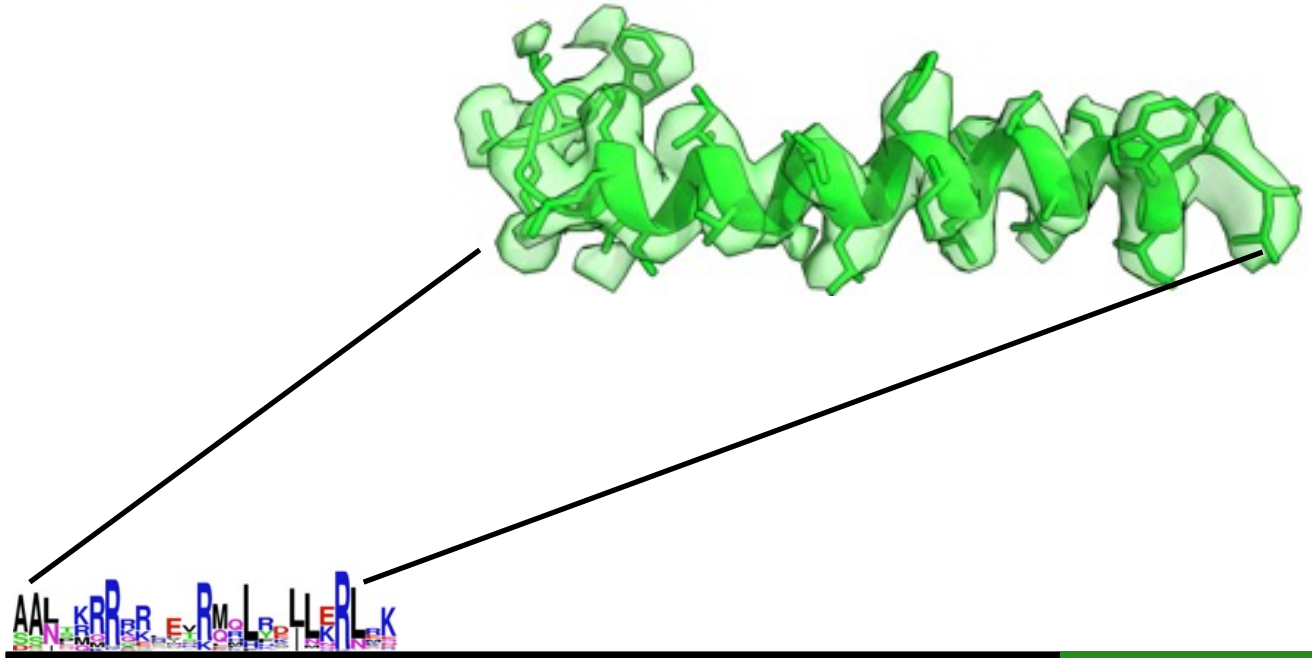
ESX-5 hexameric pore complex at 3.4 Å resolution

22aa α-helix models
built with COOT
(two alternative directions)

... and assigned to the target
sequence with findMySequence

EMBL

Sequence re-assignment as a model validation



target sequence

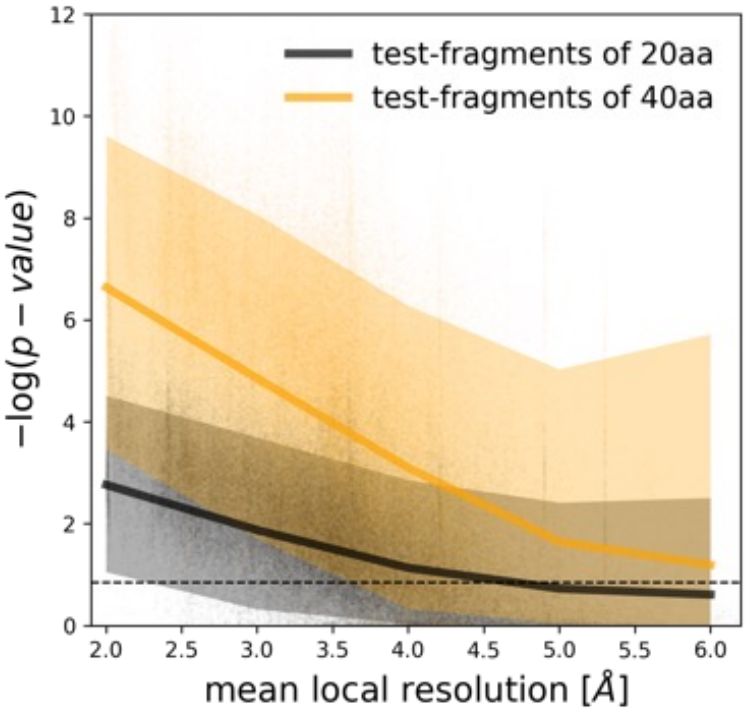
alignment p-value

	p-value small	p-value ~ 1
sequences OK	✓	?
sequences differ	⚠	?

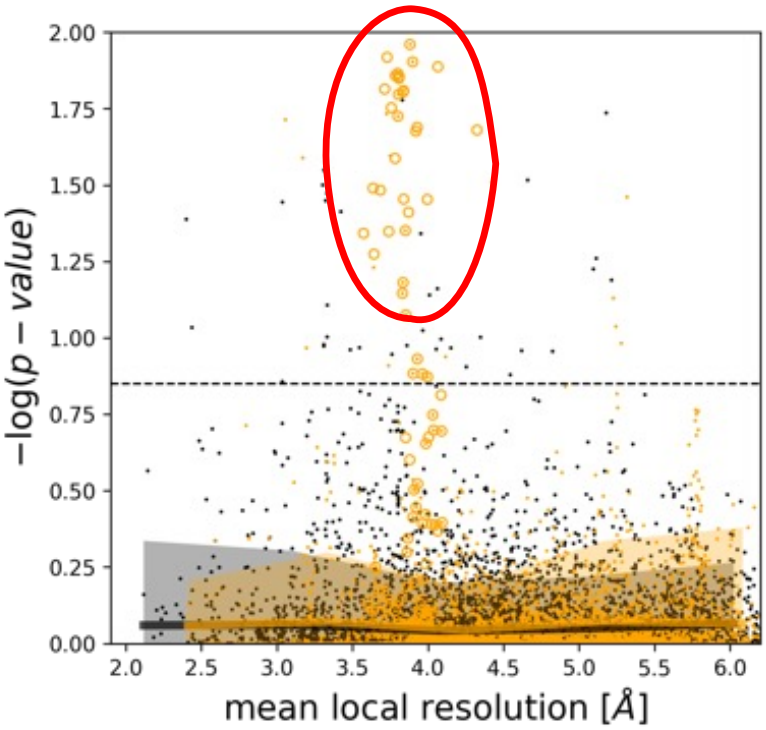
EMBL

Is the sequence assignment p-value a reliable score?

sequence OK



sequences differ



	p-value small	p-value ~ 1
sequences OK	✓	?
sequences differ	⚠	?

p-values for 30k protein chain fragments re-assigned to target sequence

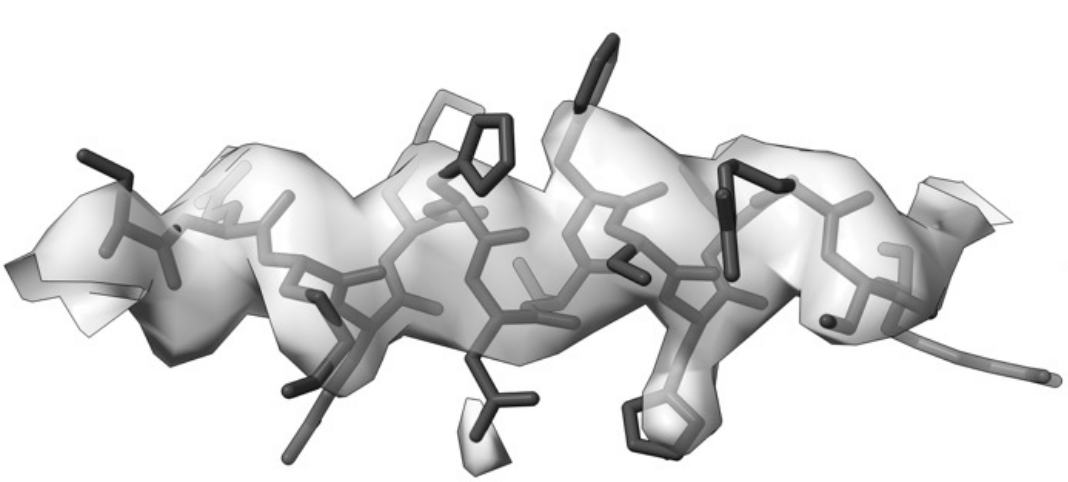
Chojnowski ActaD 2022

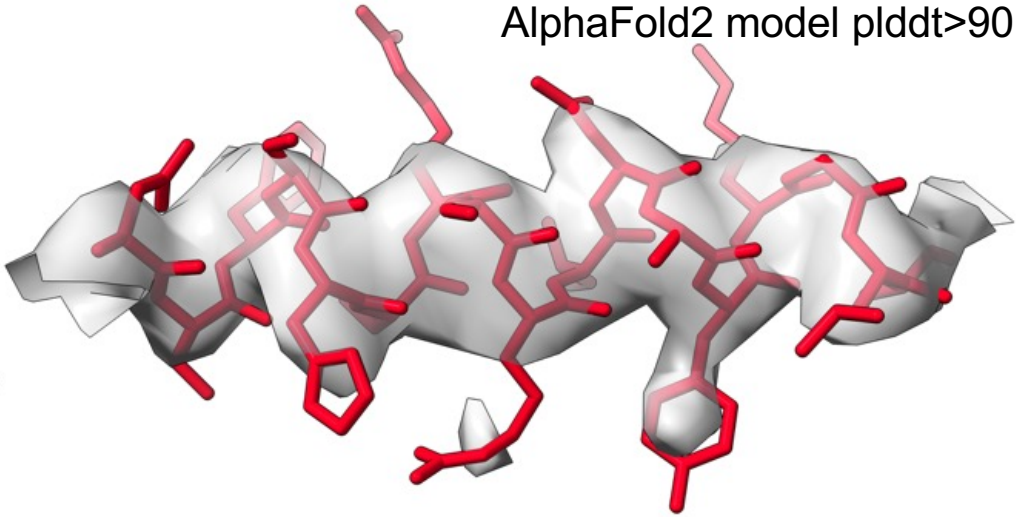
EMBL

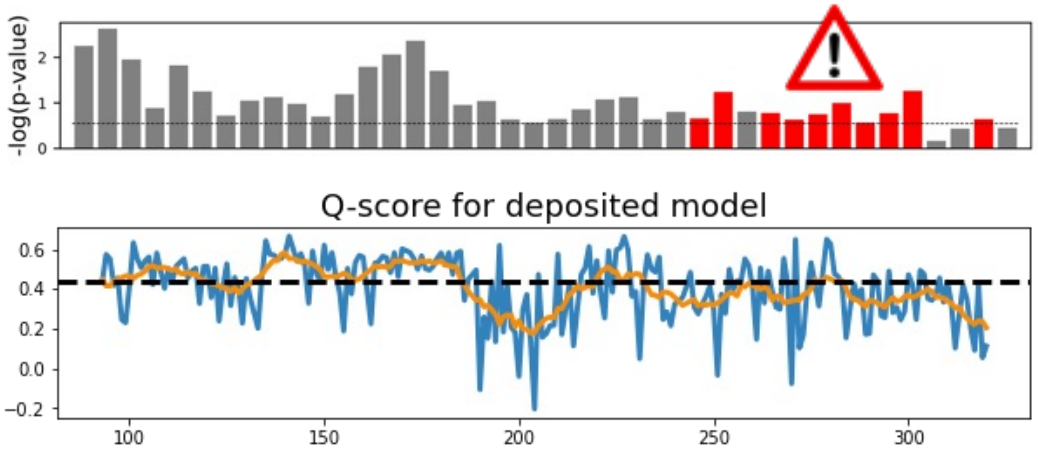
hidden errors in EM models: finding a better hypothesis

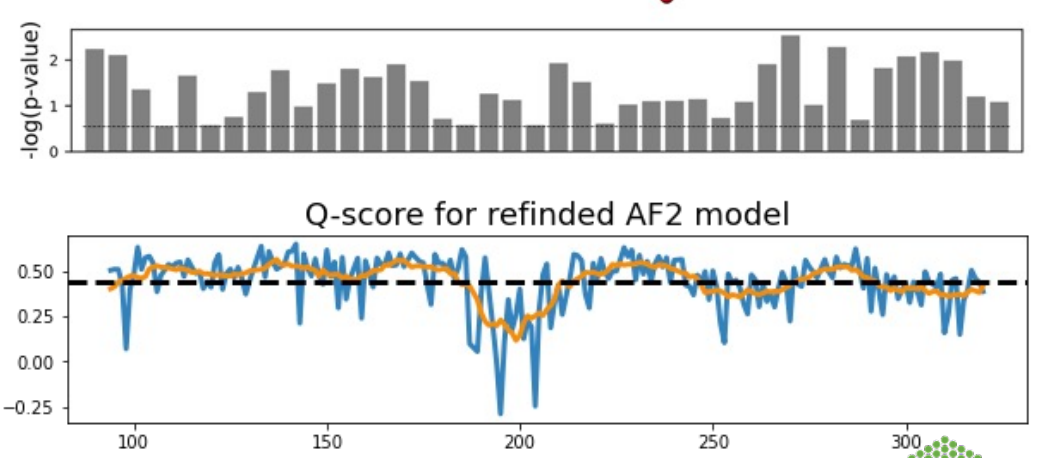
cytoplasmic domain of a cation channel at 3.8Å resolution

AlphaFold2 model plddt>90









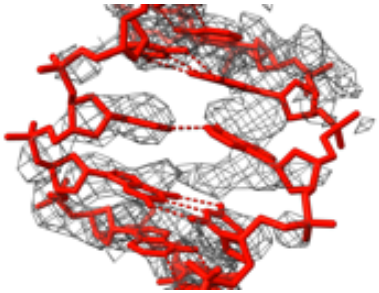
Chojnowski ActaD 2022

EMBL

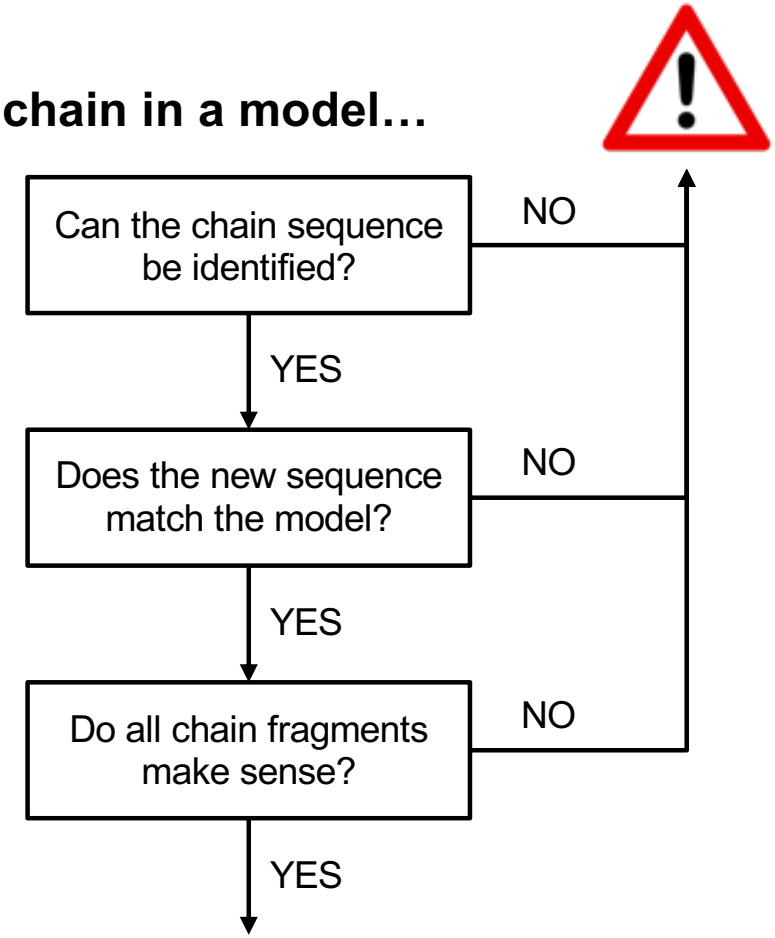
checkMySequence: complete sequence assignment validation

checkMySequence dependencies

- findMySequence (aa, HMM)
- doubleHelix (na, CM/HMM)



∀ chain in a model…



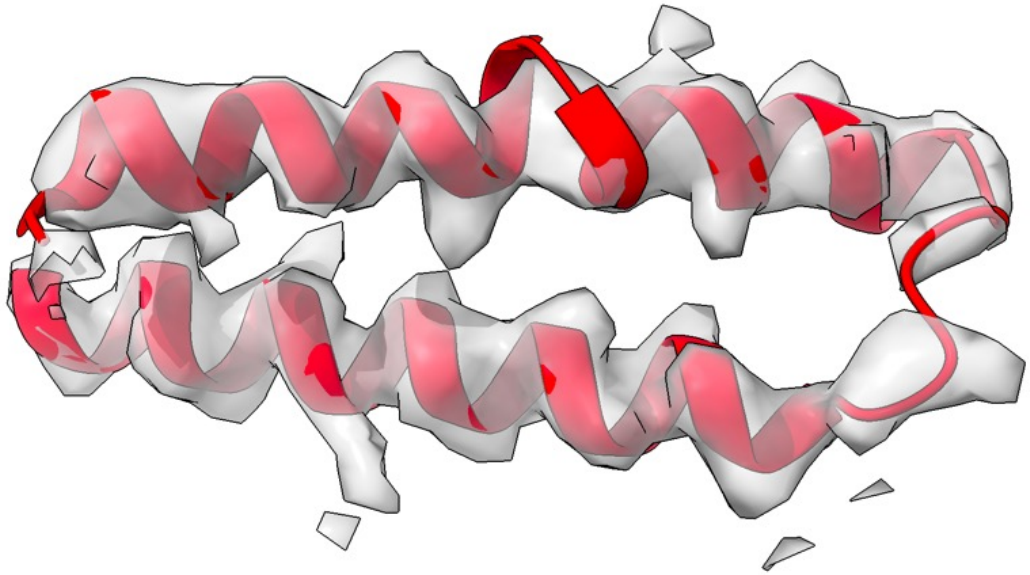
checkMySequence (proteins, EM) Chojnowski ActaD 2022
checkMySequence (proteins, MX) Chojnowski ActaD 2023
findMySequence (proteins, EM and MX) Chojnowski et al ActaD 2022
doubleHelix (NA, EM and MX) Chojnowski NAR 2023

EMBL

checkMySequence: complete sequence assignment validation

```
*********************************************************************************
*********************************   SUMMARY   *********************************
*********************************************************************************

 ==> Unidentified chains; check input sequences and model-to-map fit

             e/2:51
             g/3:39

 ==> Chains with sequence mismatches; you will have to fix them first!

  model       KDNVVQMMNEKKSFDVSDFPKVYLTTAVEEDLDT--
              |||||||||||||||||||||||||*|||||||
  refseq      KDNVVQMMNEKKSFDVSDFPKVYLTTTVEEDLDTRG

 ==> Possible sequence assignment issues

  - Fragment N/5-24 has a chain break at N/11 and no residue indexing gap, check!

  model       ---------VELTEEE-LYISKKNLLFKRF----------
                       ||||||| |||||||||||||
  refseq      AAAAAAMPRVELTEEEKLYISKKNLLFKRFVEPGRLCLIE


- Fragment F/356-395 is shifted by -4 residues [-log(p-value)=1.99]
  model seq 356-395
      sknkkEKRVQKQIQKKELQKINHDYYKGVAKAVKKKKKREEKKAKskktanqavi
    new seq 360-399
      sknkkekrvQKQIQKKELQKINHDYYKGVAKAVKKKKKREEKKAKSKKTanqavi
```

checkMySequence for 6,000 cryo-EM models at 5Å or better

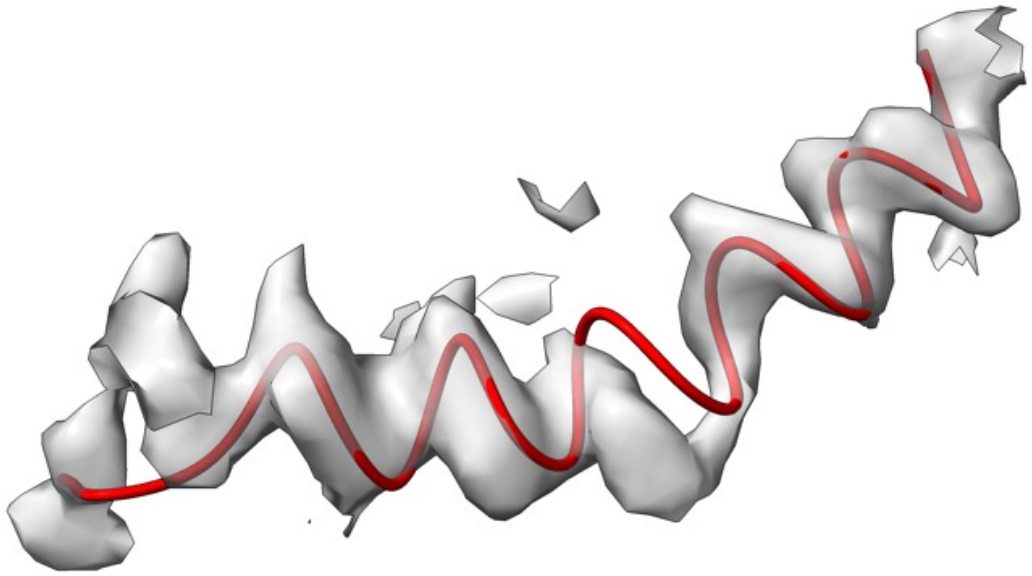


```
- Fragment y/1-20 is shifted by -1 residue [-log(p-value)=2.95]
  model seq 1-20
      MKAKELREKSVEELNTELLNllreqfn
    new seq 2-21
      mKAKELREKSVEELNTELLNLlreqfn
```

70S ribosomal protein 2.9 Å

EMBL

checkMySequence for 6,000 cryo-EM models at 5Å or better

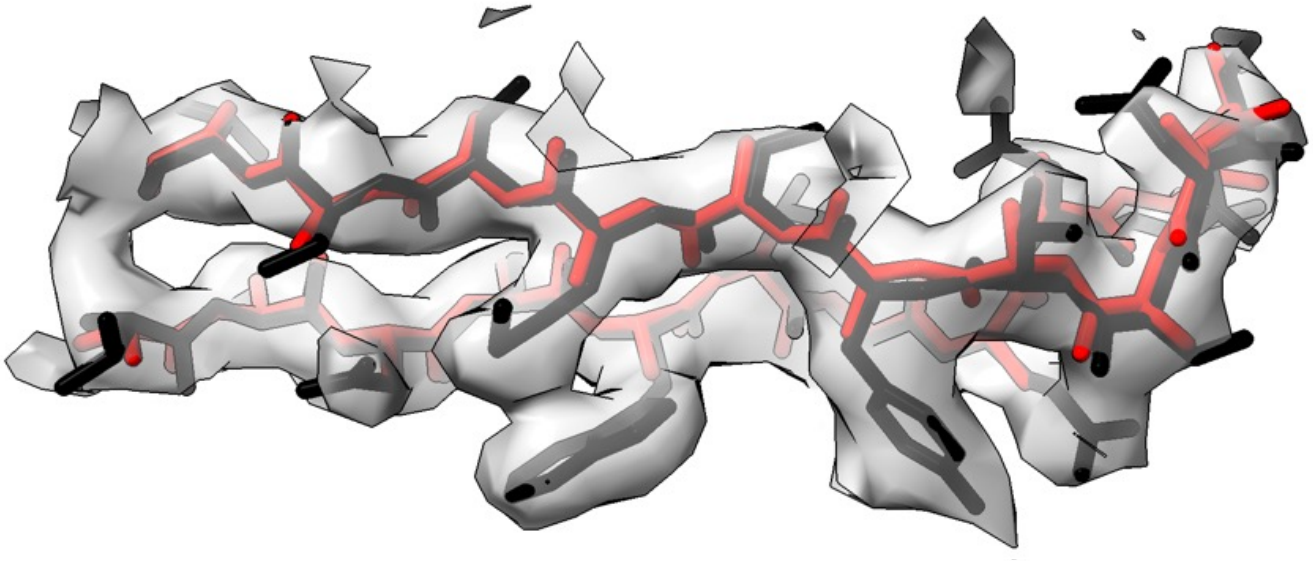


```
- Fragment u/43-62 is shifted by -3 residues [-log(p-value)=2.40]
  model seq 43-62
      ekpttERKRAKASAVKRHAKKLARENarrt
    new seq 46-65
      ekptterkRAKASAVKRHAKKLARENARrt
```

70S ribosomal protein at 3 Å

EMBL

checkMySequence for 6,000 cryo-EM models at 5Å or better



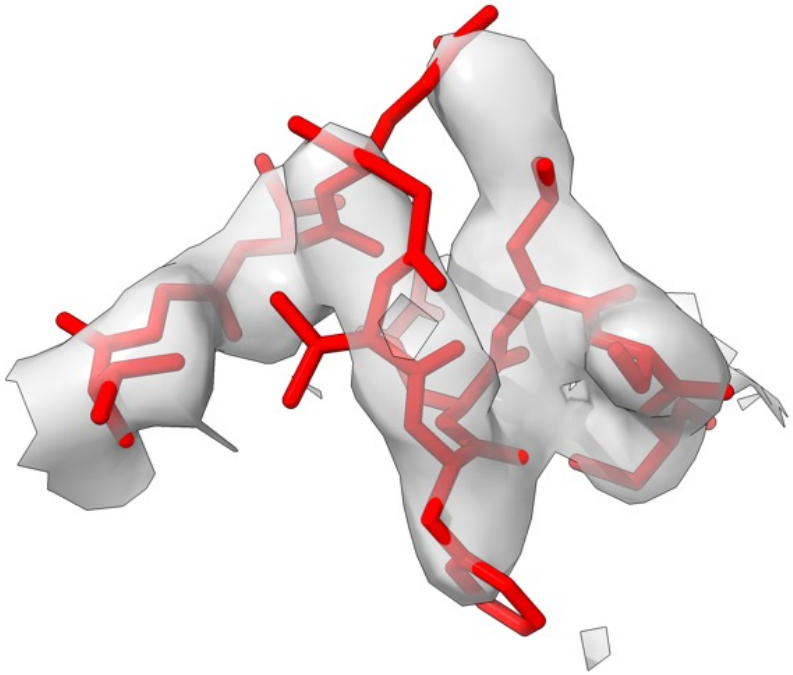
deposited model
rebuilt with findMySequence

```
- Fragment C/161-200 is shifted by 2 residue [-log(p-value)=3.11]
  model seq 161-200
      lnnfypkDINVKWKIDGSERQNGVLNSWTDQDSKDSTYSMSSTLTLTkdeyer
    new seq 159-198
      lnnfyPKDINVKWKIDGSERQNGVLNSWTDQDSKDSTYSMSSTLTltkdeyer
```

ferroportin at 3 Å

EMBL

checkMySequence for 6,000 cryo-EM models at 5Å or better

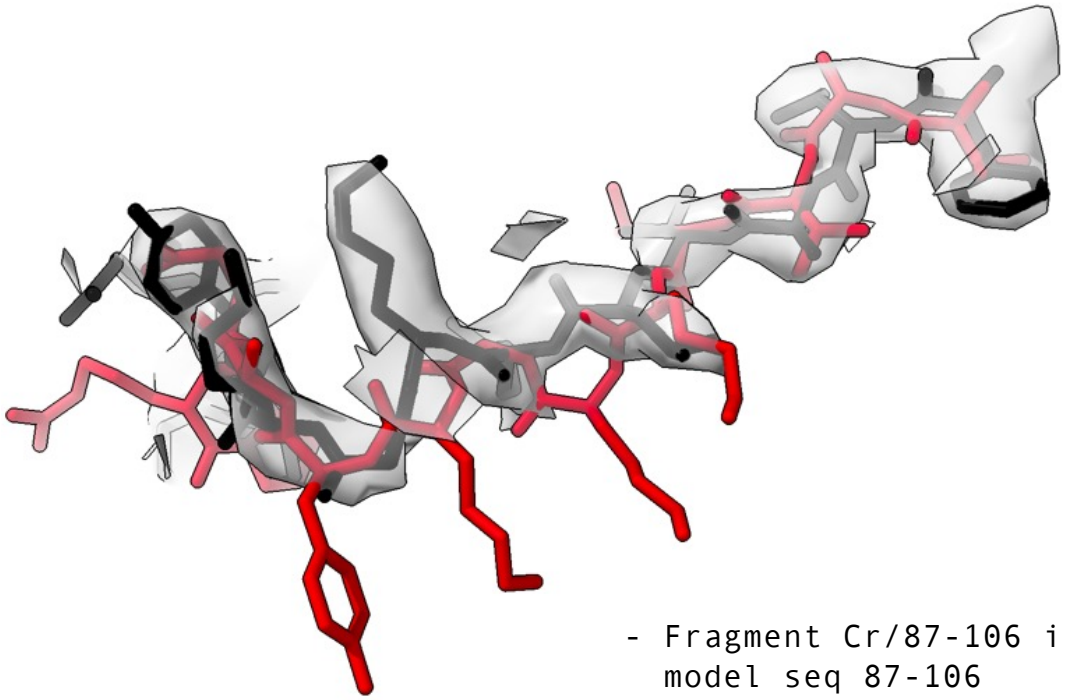


```
- Fragment A/2367-2426 is shifted by 1 residue [-log(p-value)=3.02]
  model seq 2367-2426
      kgkslrVPEKVPFRMTQNIETALGVTGVEGVFRLSCEQVLHIMRRGRETLLTLLEAFVYDPLVDWTaggeag
    new seq 2366-2425
      kgkslRVPEKVPFRMTQNIETALGVTGVEGVFRLSCEQVLHIMRRGRETLLTLLEAFVYDPLVDWtaggeag
```

a kinase at 3.6 Å

EMBL

checkMySequence for 6,000 cryo-EM models at 5Å or better



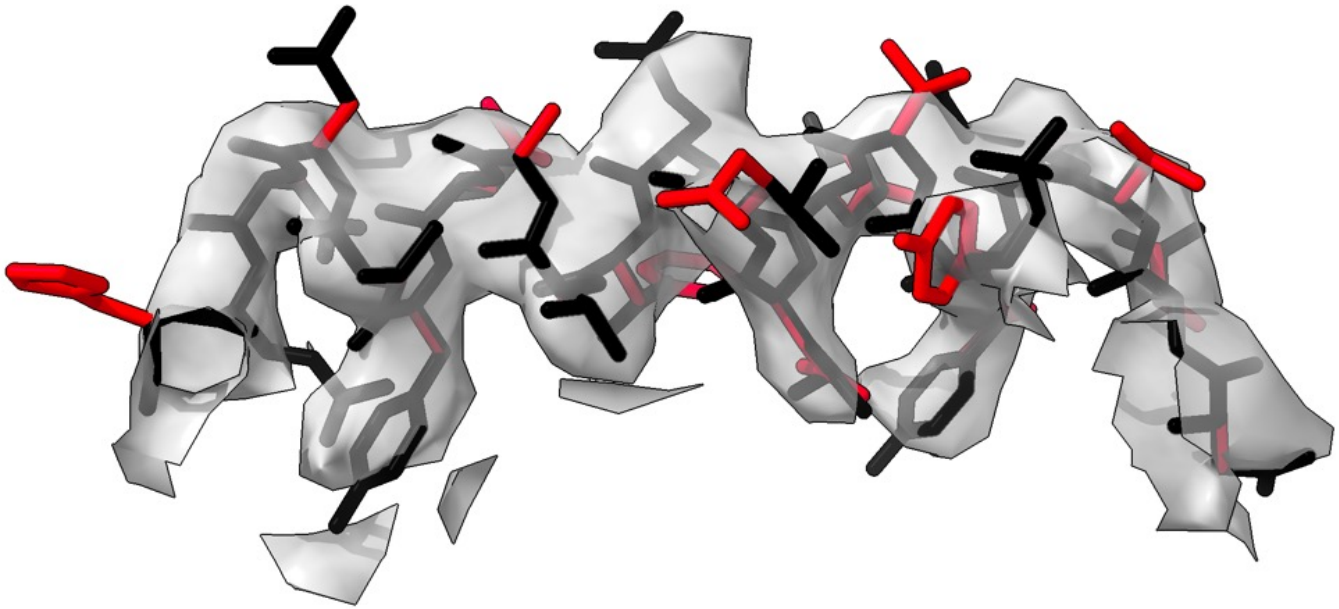
deposited model
rebuilt with findMySequence

```
- Fragment Cr/87-106 is shifted by 4 residue [-log(p-value)=2.95]
  model seq 87-106
      akntvrvdfKAGPRRSLKKLKNLLIGSKYrkdltq
    new seq 83-102
      akntvRVDFKAGPRRSLKKLKNLLIgskyrkdltq
```

50S ribosomal protein at 3 Å

EMBL

checkMySequence for 6,000 cryo-EM models at 5Å or better



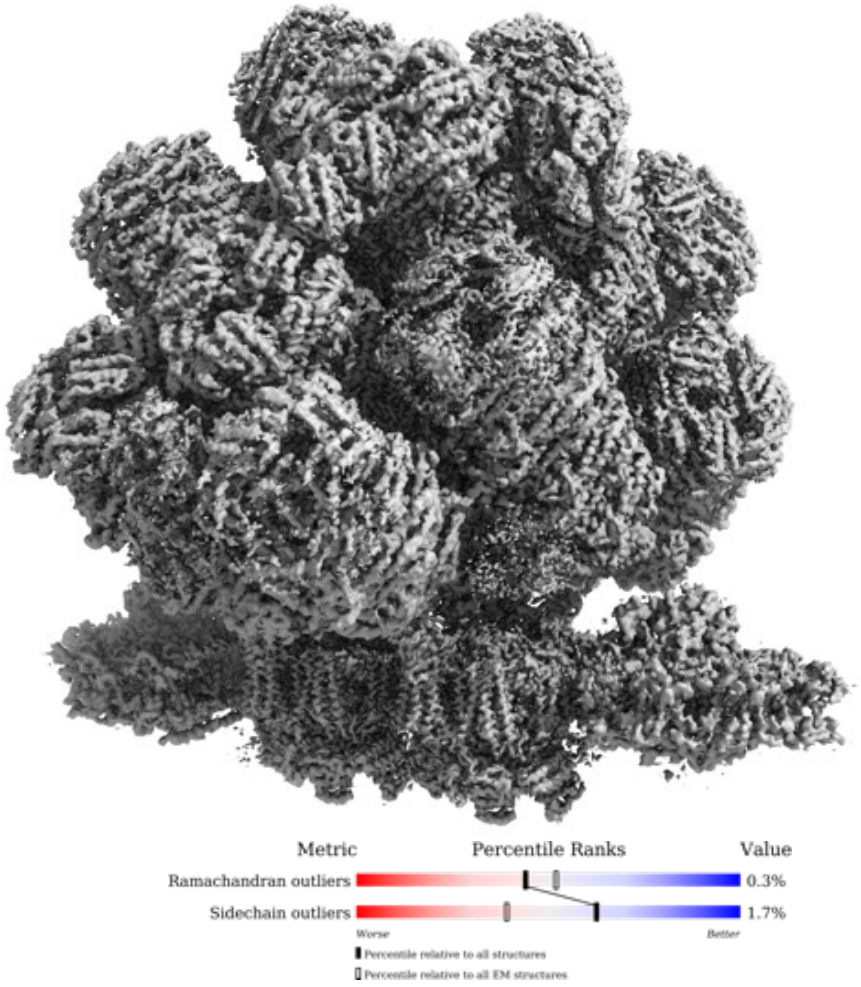
deposited model
rebuilt with findMySequence

```
- Fragment C/145-204 is shifted by 10 residue [-log(p-value)=1.94]
  model seq 145-204
      gfaelarryahnlanARFLWRNRVGAEAVEVRINHIRQGEVARAWRFDALAIGLRDFKADAELDALAELIASGLSgsghvl
    new seq 135-194
      gfaelARRYAHNLANARFLWRNRVGAEAVEVRINHIRQGEVARAWRFDALAIGLRDFKADAELDAlaeliasglsgsghvl
```

CRISPR-associated protein at 3.8 Å

EMBL

recent EM use case: In situ photosystem megacomplex at 3.3Å



18MDa/153,000 residues/864 aa chains

Panagiotis Kastritis
@3Dstructure

Replying to @emeKato

PDB validation report of *only* 4691 pages! Jaw-dropping work!

Tristan Croll
@CrollTristan

Wow. At a rate of 1 second per residue, viewing every residue in this monster would take...

... checks notes...

... blinks...

... just under 43 *hours*.

Tristan Croll
@CrollTristan

Tristan Croll
@CrollTristan

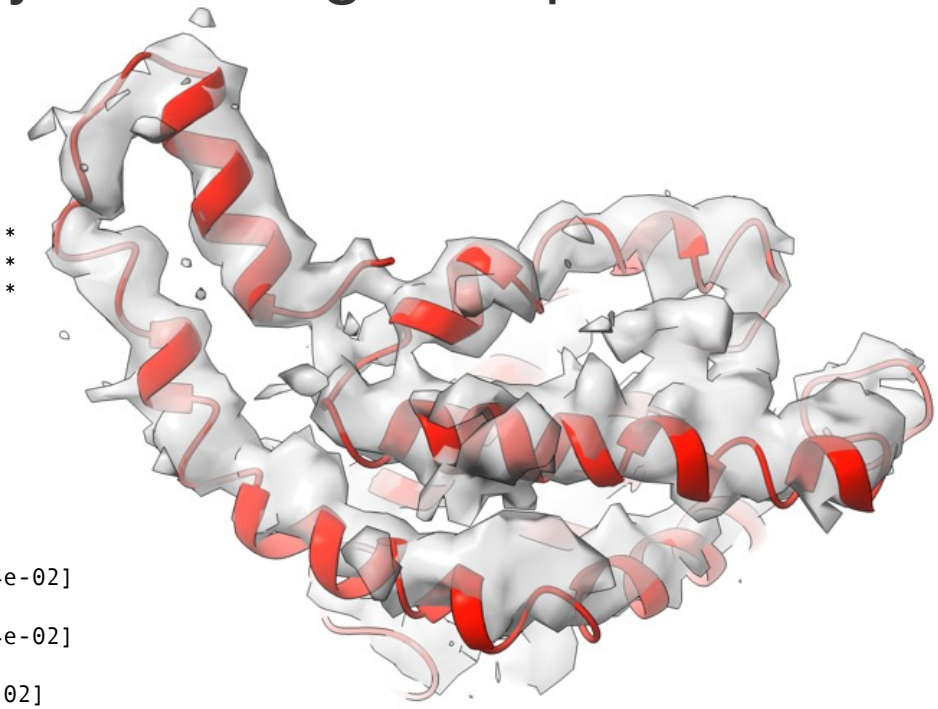
Replying to @ktototak

OK, I'm happy to stand corrected - the example you showed me is clearly a bona fide register error.

7y5e/EMD-33618
You, X., Zhang, X., Cheng, J., Xiao, Y.N., Sui, S.F. *Nature 2023*

EMBL

recent EM use case: In situ photosystem megacomplex

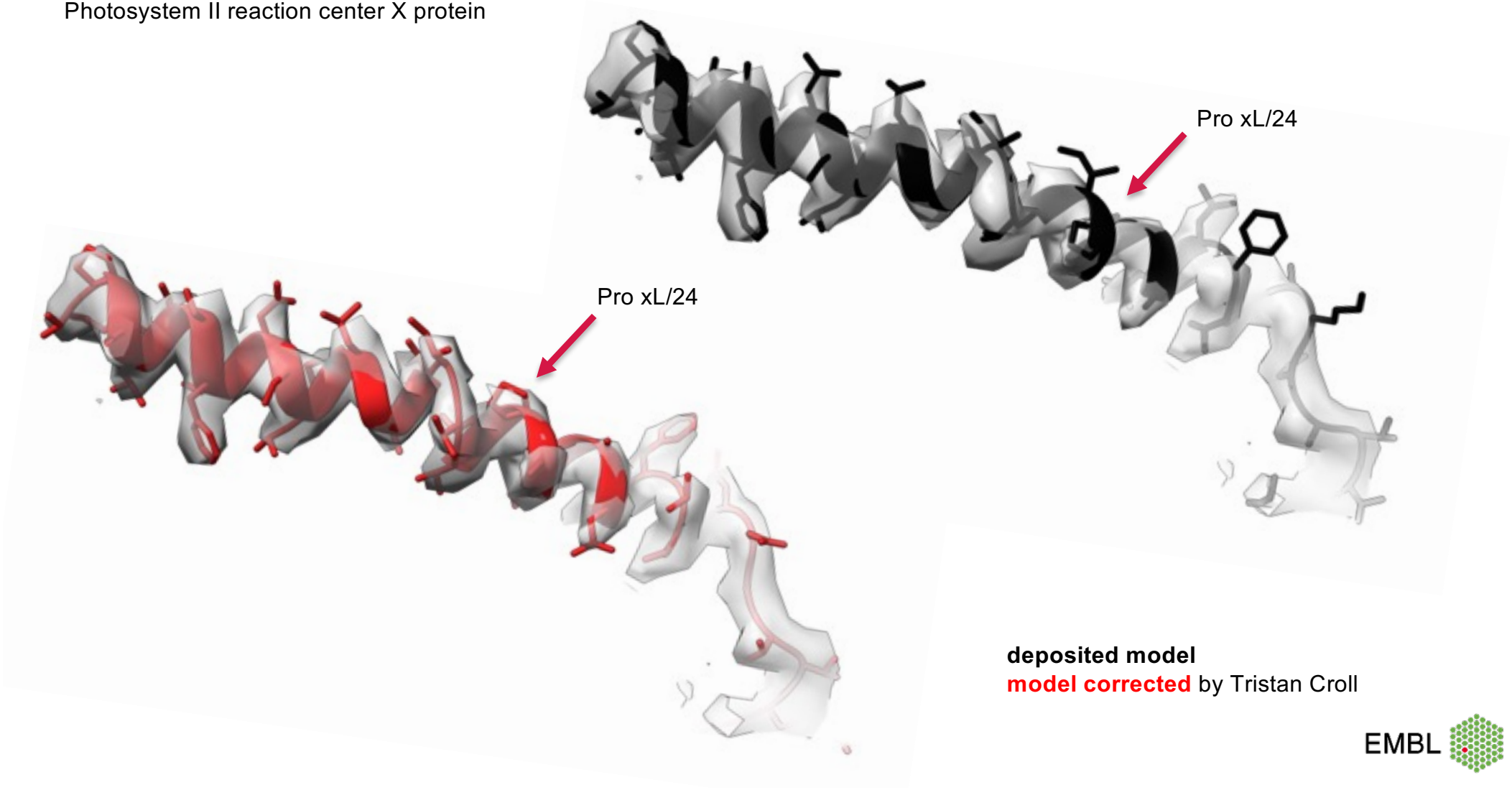


```
*********************************************************************************
*********************************   SUMMARY   *********************************
*********************************************************************************

 ==> Unidentified chains; check input sequences and model-to-map fit

AA/1:164
… [144 removed for the presentation clarity]
Z9/25:299

 ==> Sequence register shifts

 - protein chain fragment YP/227-266 may be shifted by 2 residues [p-value=4.14e-02]

 - protein chain fragment bP/227-266 may be shifted by 2 residues [p-value=4.04e-02]

 - protein chain fragment xL/18-41 may be shifted by -1 residue [p-value=1.19e-02]

 - protein chain fragment XL/18-37 may be shifted by -1 residue [p-value=7.30e-03]

 - protein chain fragment X6/18-37 may be shifted by -1 residue [p-value=5.30e-03]

 - protein chain fragment x6/18-37 may be shifted by -1 residue [p-value=5.35e-02]


Time elapsed 0:43:23
```

EMBL

recent EM use case: In situ photosystem megacomplex

Photosystem II reaction center X protein

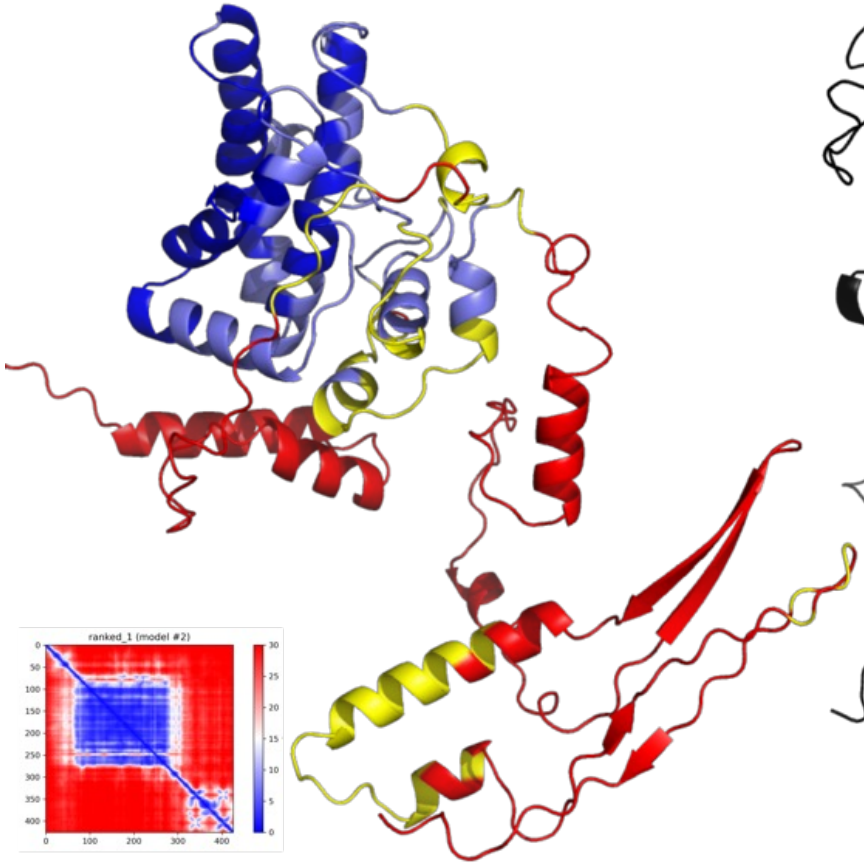


deposited model
model corrected by Tristan Croll

EMBL

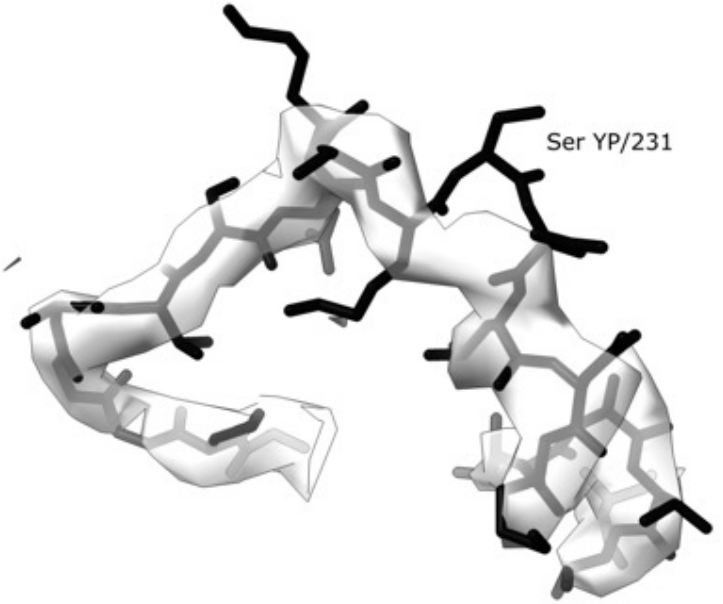
recent EM use case: In situ photosystem megacomplex

Phycobilisome linker polypeptide



AlphaFold2 model

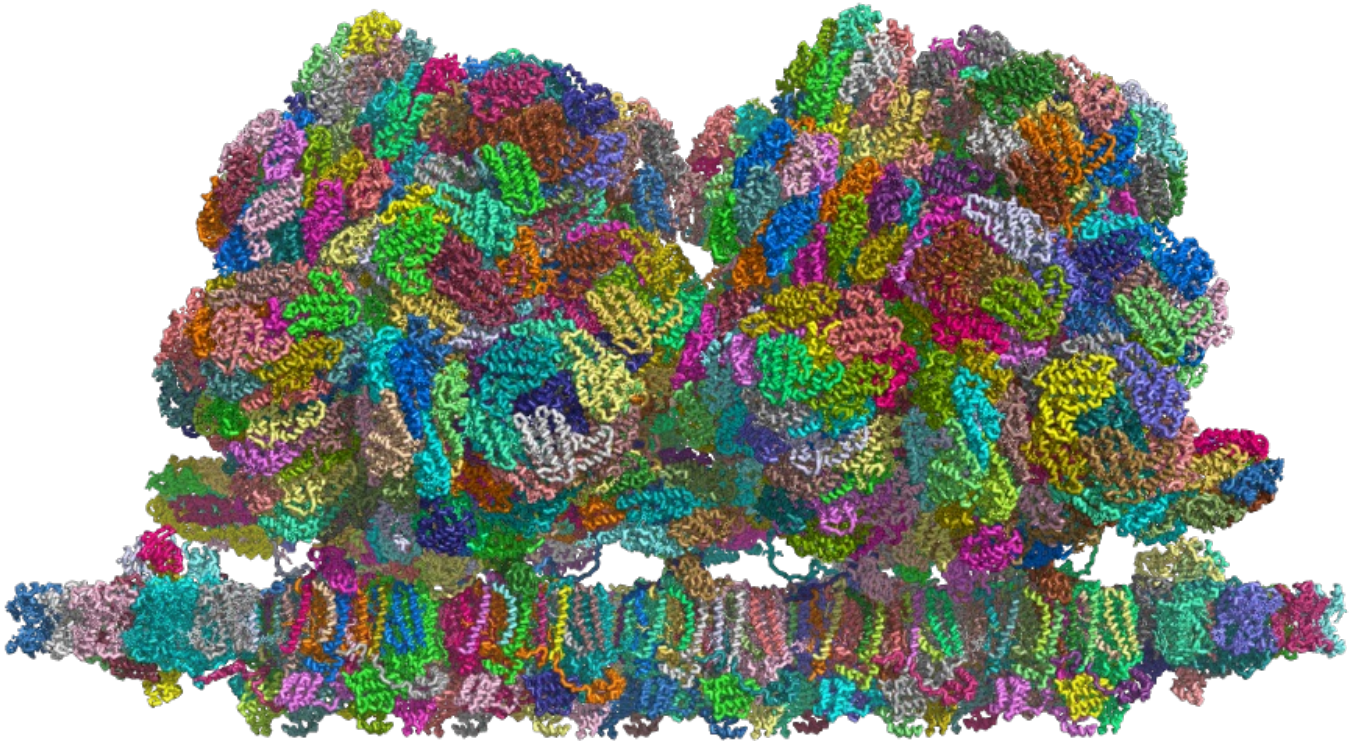
deposited model



EMBL

When you think you've seen it all…

In situ **double**-PBS-PSII-PSI-LHCs megacomplex at 4.3Å



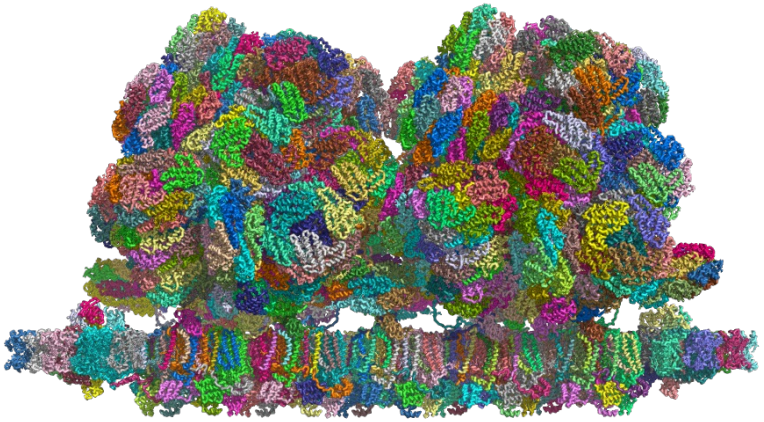
37MDa/305,000 residues/1,792 aa chains

7y5e/EMD-33618
You, X., Zhang, X., Cheng, J., Xiao, Y.N., Sui, S.F. *Nature 2023*

EMBL

When you think you've seen it all…

In situ **double**-PBS-PSII-PSI-LHCs megacomplex at 4.3Å



37MDa/305,000 residues/1,792 aa chains

```
- Protein chain Y5/62:298  - sequence identity to reference 65.22% [E-value=4.00e-10]

             12345678901234567890123456789012345678901234567890123456789012345678901234567890

 model       ---------------------------------RRTVSFNARVA-------------RN-KSQAKKILEKADEFFARSVTM
                                                | |     ||             |  |  ||    | ||  ||||
 refseq      AAAAAAAAAAAMAAFVGSAASAFTGASAVKANEKRSVCSLQMVAMPQTGLVNSKFSARMAKKTAKQTKNKVDEYMARSVQR

 model       QYKAFACPNGVYDIQCTEGTVKGAAYEKRAMAVSAAFRAKQASPAAKARALFENRRHAIIASHECQHEEDLFVRFPKLSA
             |||  |   |||  ||||||||||   |  | |  || || |   ||  ||| | |||||   |  ||    ||||| |
 refseq      QYKQAAVATGVYGTQCTEGTVKGAAEASRSAALSRQFRIKQRSAFSKAHDLFEFRKHAIIAAAGCSYEEKMVTRFPKLAA

 model       AYMMGKTEAMRTCSRYVVPDSLEEEYMAASVDRQMKERACPGGVYASSCVEGNAKGQAEQARVAALATAFRSAQKSASKT
             |   | || |||||||||| | |||||||||| ||| |  |||||  || || |||||| ||| ||  | | | ||||
 refseq      AMVLGQTEMMRTCSRYVVPESVEEEYMAASVDKQMKRRGAPGGVYSLSCAEGVAKGQAEIARVSALGAAYRAASKSASAV

 model       TAERYSSAAYGRDHFAHGCSYEESVFNTYPATAAAMRSKSYNY----------
             ||||| | |||| ||||||||||  || ||| |||||| || |
 refseq      TAERYNSMAYGRVHFAHGCSYEEQQFNKYPAAAAAMRSDSYGYAAAAAAAAAA

Time elapsed 1:46:54
```

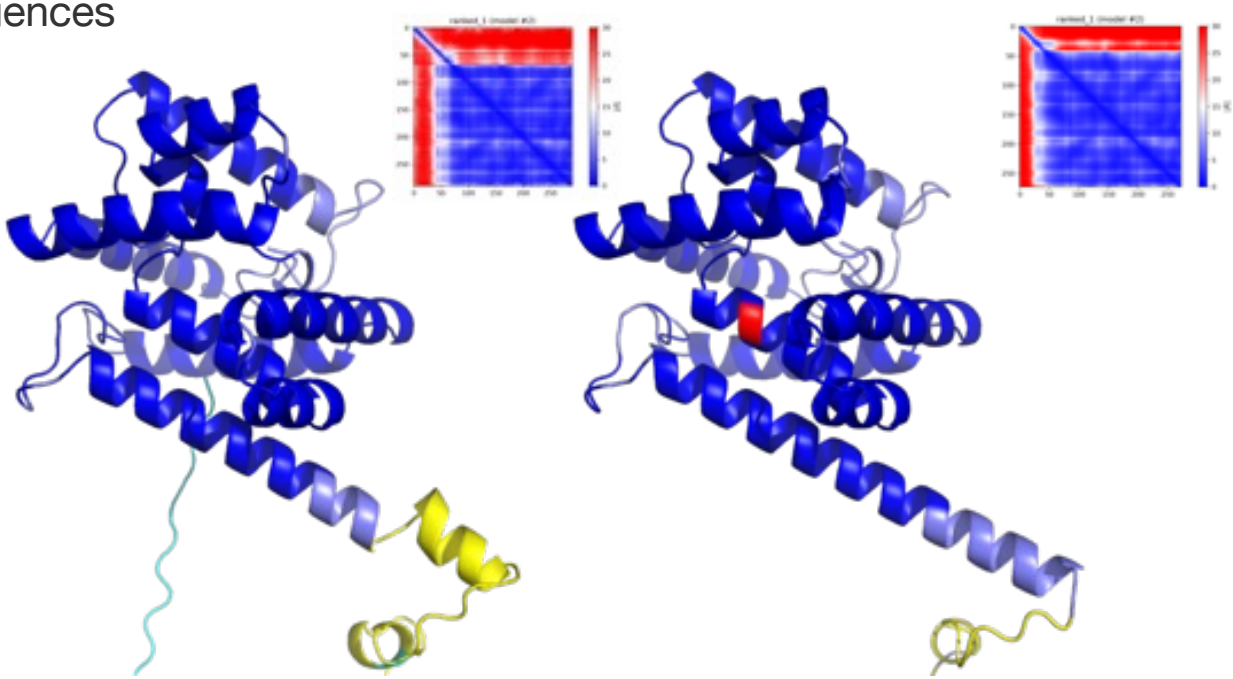
7y5e/EMD-33618
You, X., Zhang, X., Cheng, J., Xiao, Y.N., Sui, S.F. *Nature 2023*

EMBL

When you think you've seen it all…

two R-phycoerythrin gamma chain sequences

deposited model
(7y7a/Y5)



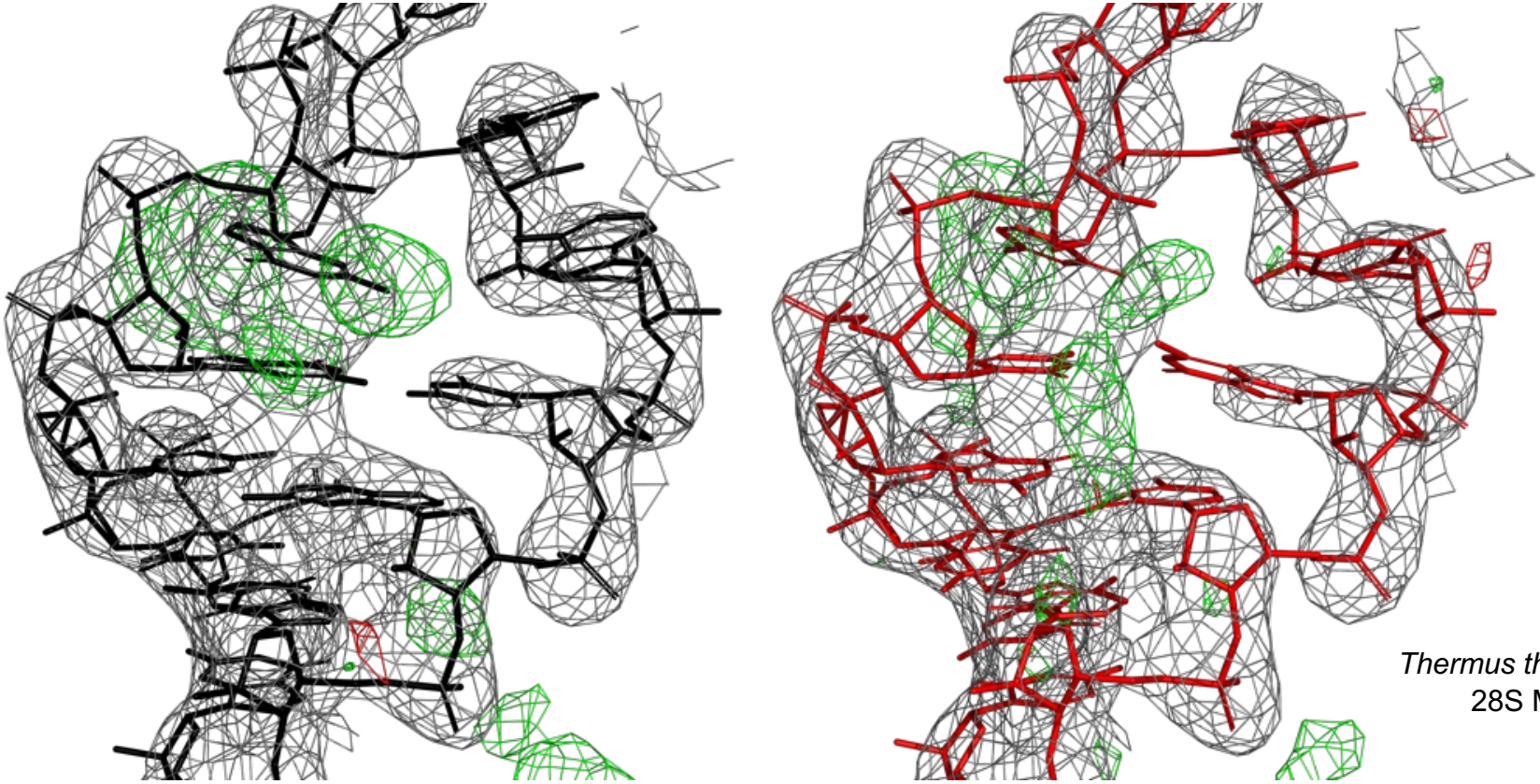
AlphaFold2 prediction for
deposited model sequence
(A0A5J4YX19)

AlphaFold2 model for
alternative sequence
(A0A5J4YZM7)

65% sequence identity
0.5Å CA RMSD

EMBL

Nucleic acid MX/EM model building is challenging



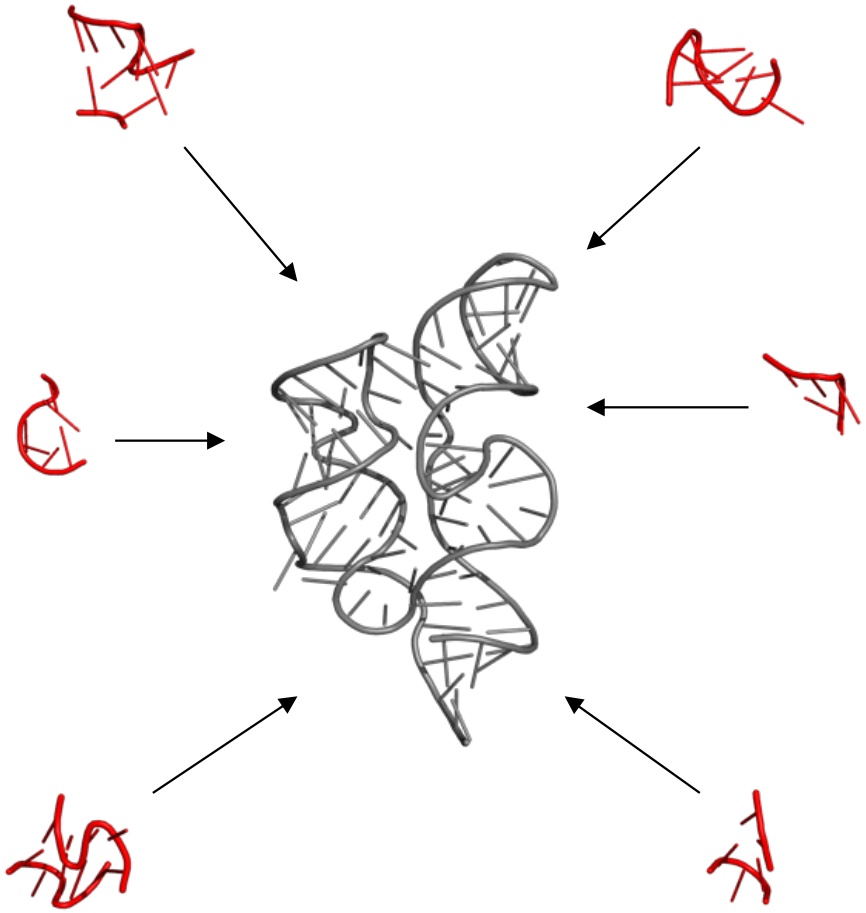
Thermus thermophilus
28S MX @ 2.8Å

Different types of purines and pyrimidines are (usually) indistinguishable in MX or EM maps

Chojnowski NAR, 2023

EMBL

Base-pairs from backbone geometry

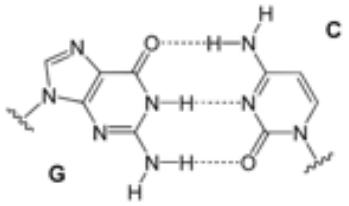


input RNA model backbone is “covered” with motifs with known secondary structure (from genesilico.pl/rnabricks2)

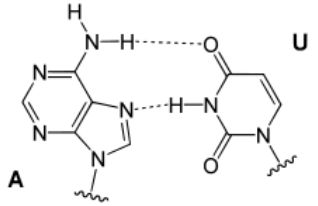
Chojnowski et al. NAR 42 D123-D131 2014
Chojnowski at al. ActaD 71 697-705 2015

EMBL

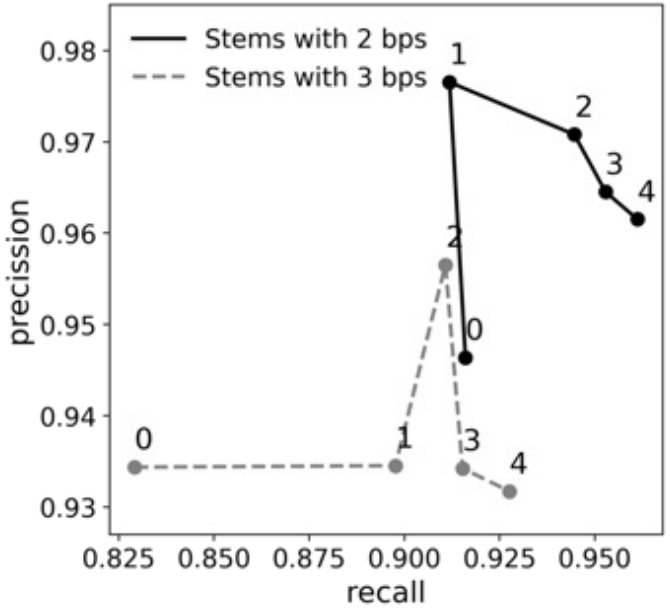
Base-pairs from backbone geometry

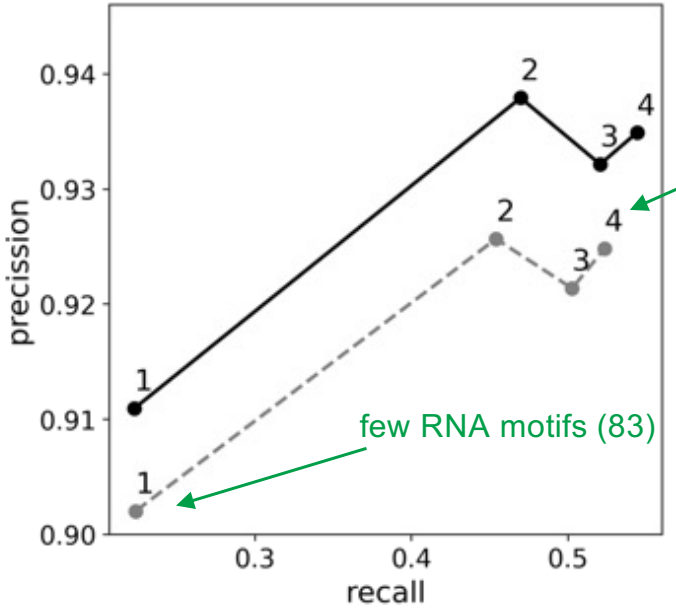


Watson-Crick (canonical) base-pairs



non-canonical base-pairs



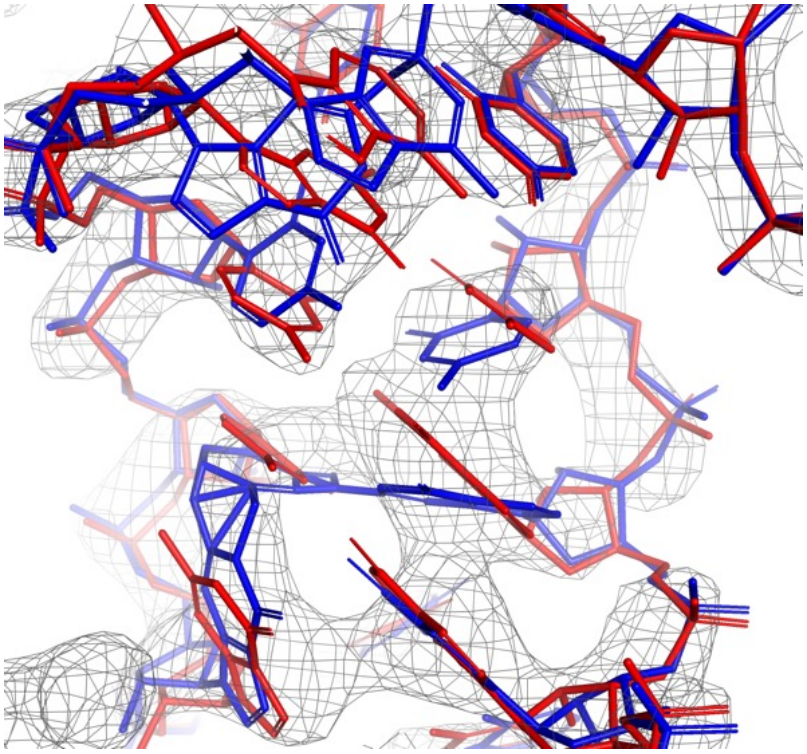


Chojnowski NAR, 2023

EMBL

Secondary structure restraints from doubleHelix

nucleic acid sequence identification and assignment tool



ARP/wARP model of 23S fragment in
cryo-EM map @ 3.2Å resolution

Standard tools

phenix.secondary_structure_restraints
LibG
ISOLDE

are very reliable if a model is good,
but cannot restrain base-pairs if

- geometry is very bad
- sequence is wrong

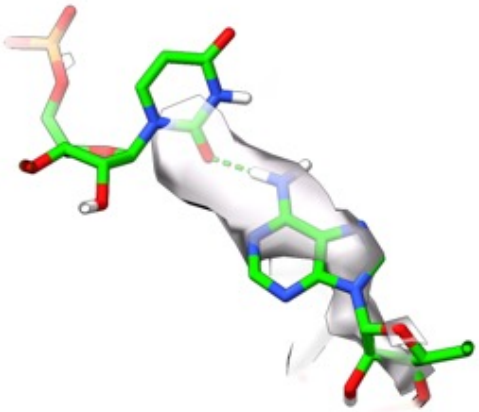
Chojnowski NAR, 2023

EMBL

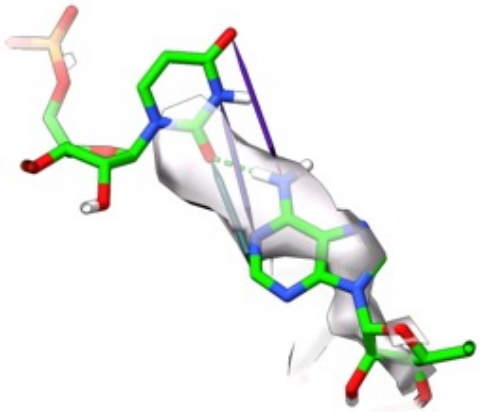
Secondary structure restraints from doubleHelix

nucleic acid sequence identification and assignment tool

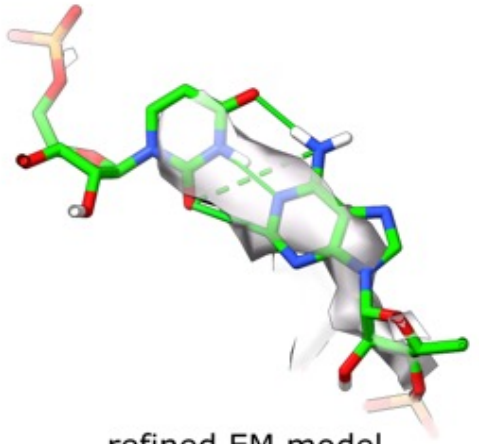
RNA polymerase
@2.5Å (7bv2/30210)



deposited EM model
(dashed line - auto-generated
H-bond restrain)



deposited EM model
with doubleHelix restraints
(violated - dark solid bars)



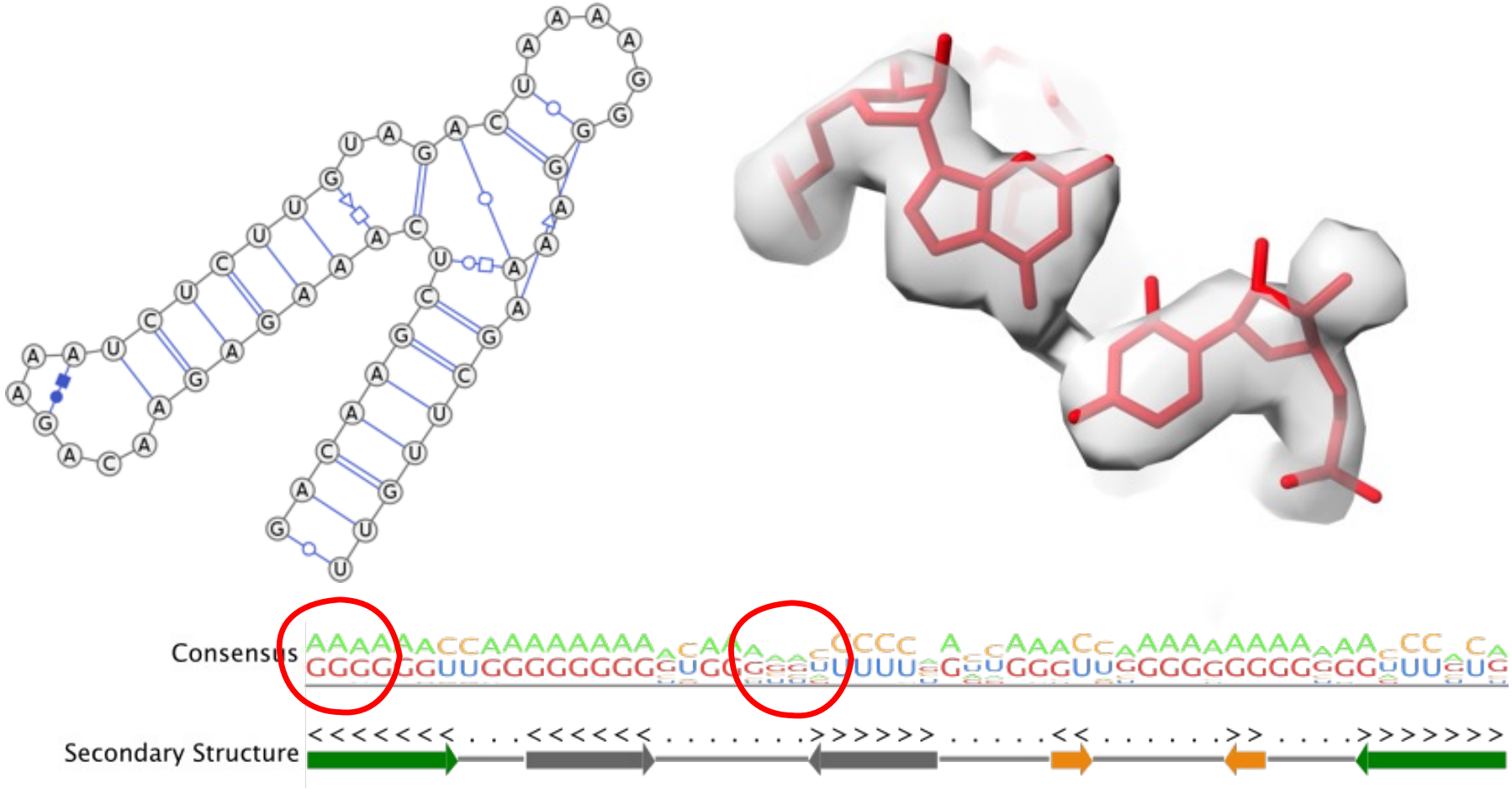
refined EM model
with doubleHelix restraints
(satisfied - green solid bars)

Chojnowski NAR, 2023

EMBL

Nucleic acid sequence identification with INFERNAL

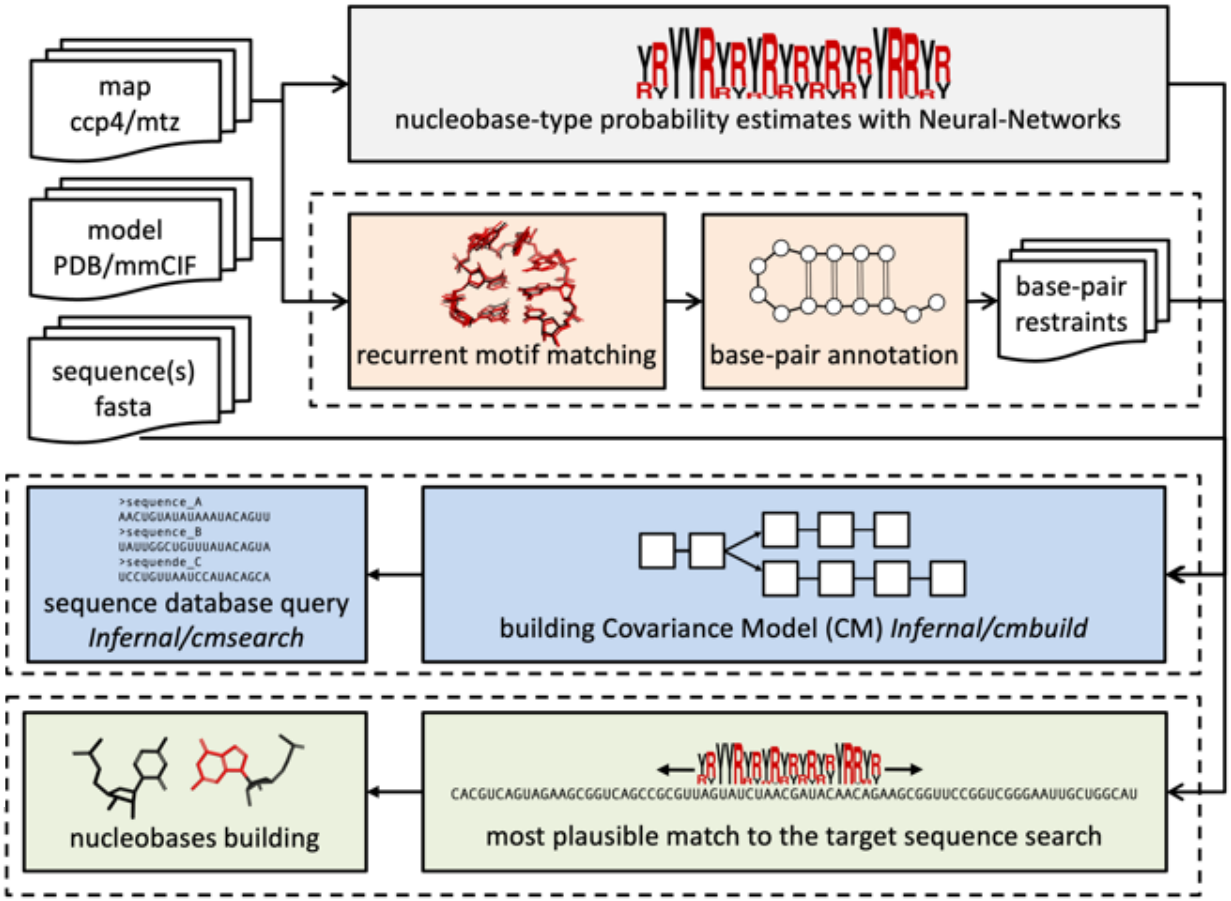
Fragment of 28S rRNA EM model at 3.2Å resolution

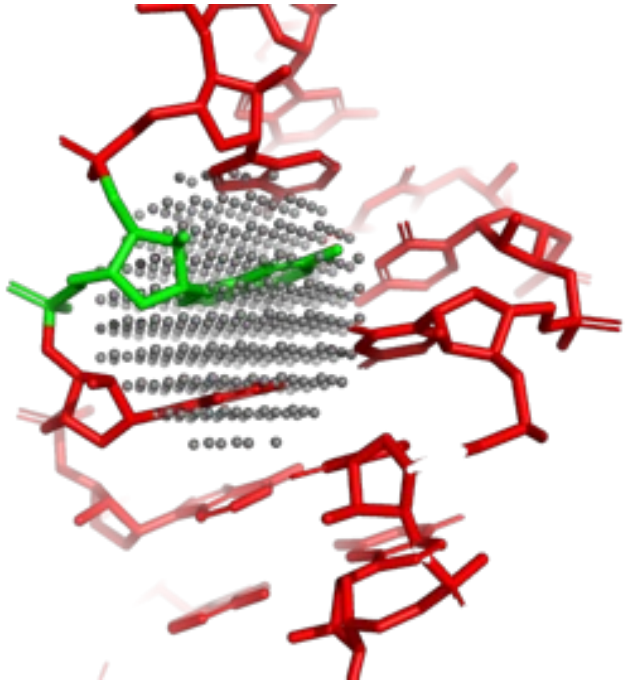


Chojnowski NAR, 2023

EMBL

doubleHelix - NA sequence assignment and validation in EM/MX

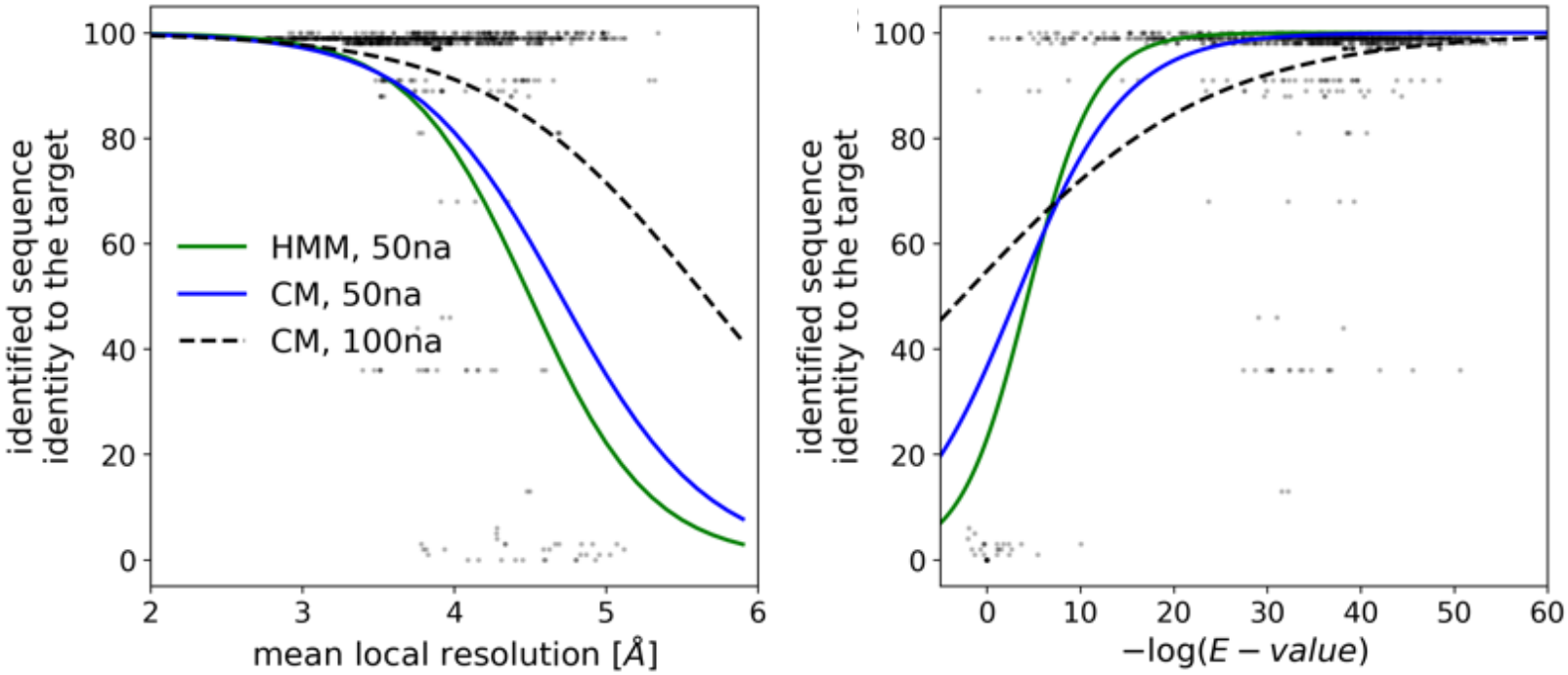




Chojnowski NAR, 2023

EMBL

Nucleic acid sequence identification in EM structures

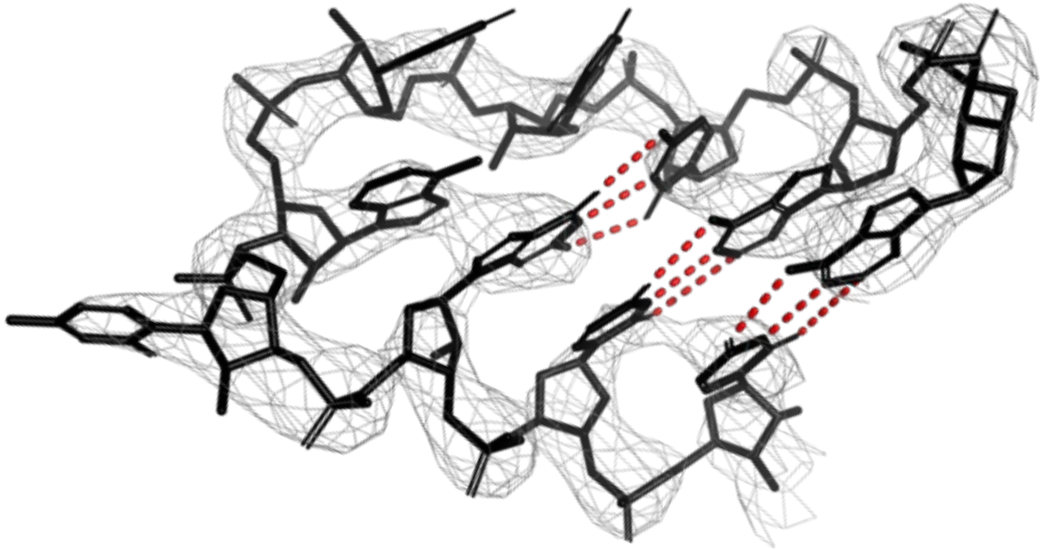


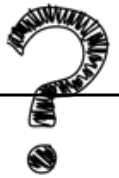
Sequence identification of short RNA fragments from
17 EM structures of ribosomes at 3.5Å or better

Chojnowski NAR, 2023

EMBL

checkMySequence and RNA/DNA sequence validation



	p-value small	p-value ~ 1
sequences OK	✓	
sequences differ		

target sequence

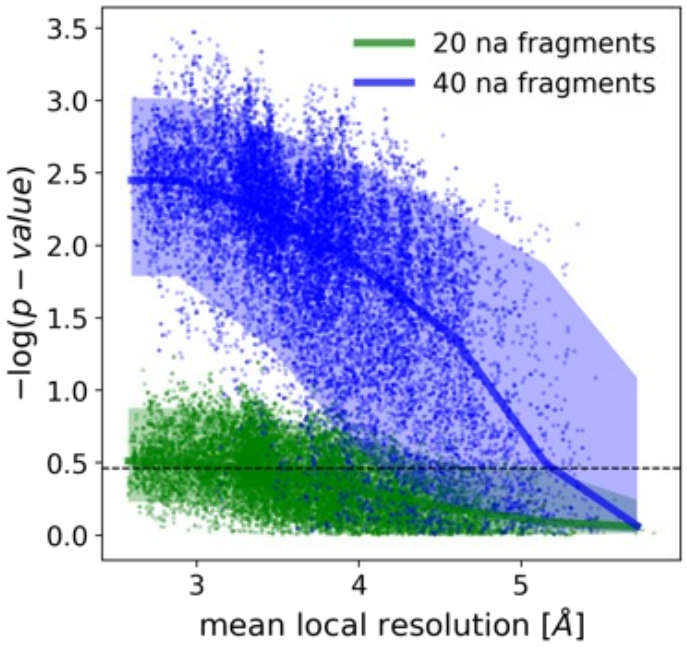
alignment p-value

Chojnowski NAR, 2023

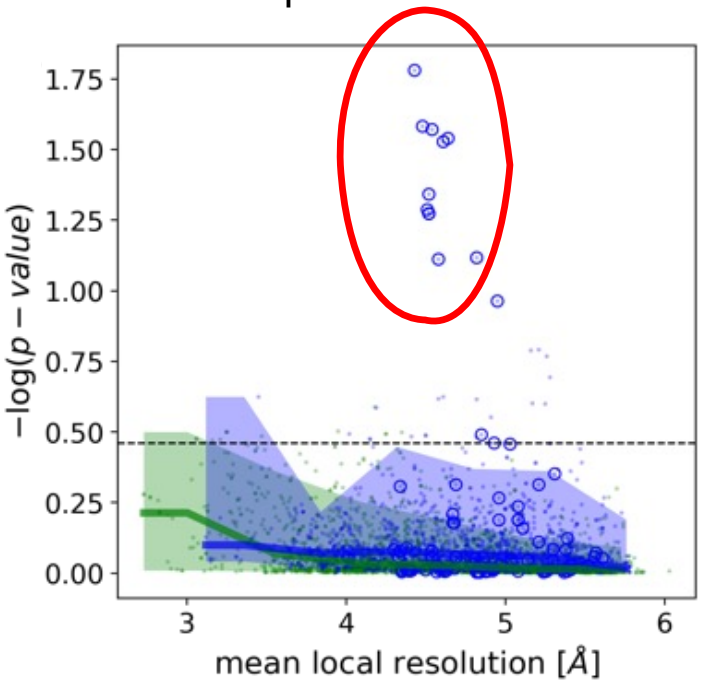
EMBL

checkMySequence and RNA/DNA sequence validation in EM

sequences OK



sequences differ



	p-value small	p-value ~ 1
sequences OK	✓	?
sequences differ	!	?

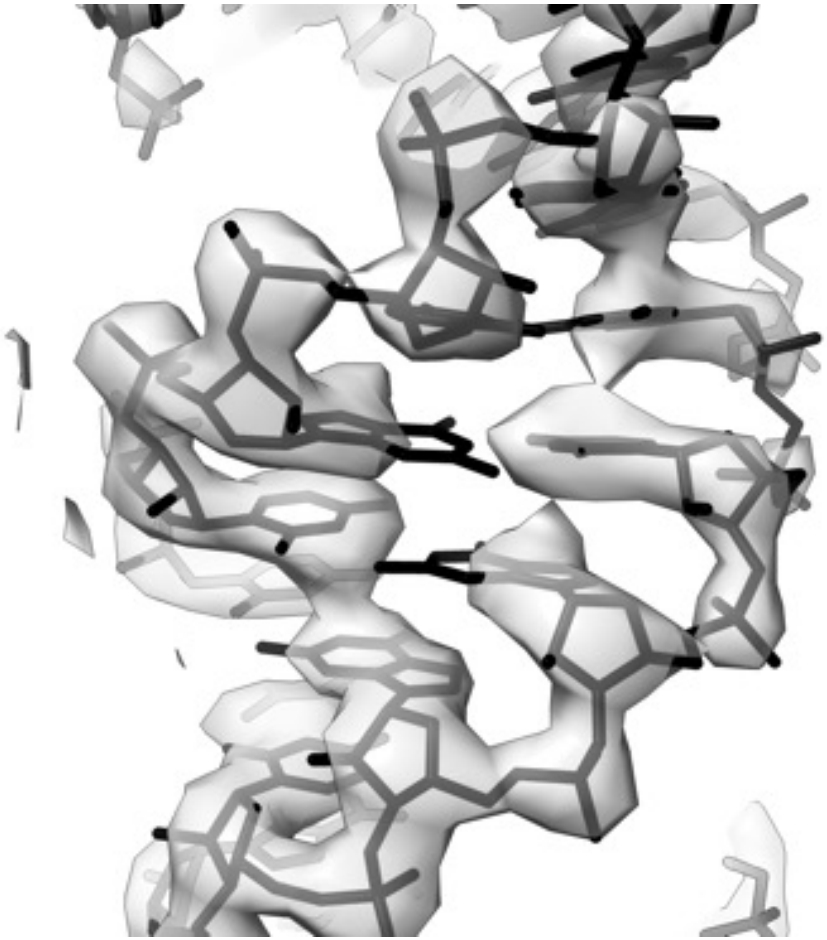
p-values for 20k RNA chain fragments re-assigned to target sequence

Chojnowski NAR, 2023

EMBL

#1 Sequence assignment issues in rRNA

expansion segment in an EM model of mammalian ribosome @3.4Å (deposited 2014)



Chojnowski NAR, 2023

EMBL

#1 Sequence assignment issues in rRNA

expansion segment in an EM model of mammalian ribosome @3.4Å (deposited 2014)

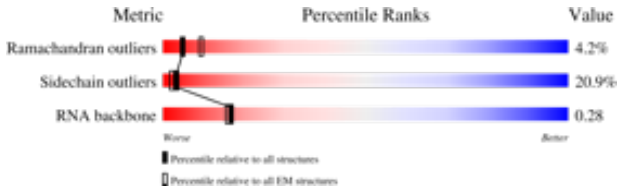

```
*********************************************************************************
*********************************    SUMMARY    *********************************
*********************************************************************************

 ==> Unidentified chains; check input sequences and model-to-map fit
2/7:68
 ==> Possible sequence assignment issues

 - Nucleic-acid chain fragment 5/442-491 may be shifted by 1 residue [-log(p-value)=0.72]
   model seq 442-491
       cgggugGGGUCCGCGCAGUCCGCCCGGAGGAUUCAACCCGGCGGCGGGUCCGGCCGugucgg
     new seq 441-490
       cggguGGGGUCCGCGCAGUCCGCCCGGAGGAUUCAACCCGGCGGCGGGUCCGGCCgugucgg

 - Nucleic-acid chain fragment 5/657-706 may be shifted by -1 residue [-log(p-value)=1.04]
   model seq 657-706
       gucccCCGACCGGCGACCGGCCGCCGCCGGGCGCAUUUCCACCGCGGCGGUGCGCcgcgacc
     new seq 658-707
       gucccCGACCGGCGACCGGCCGCCGCCGGGCGCAUUUCCACCGCGGCGGUGCGCCgcgacc

[…]

Time elapsed 0:05:53
```

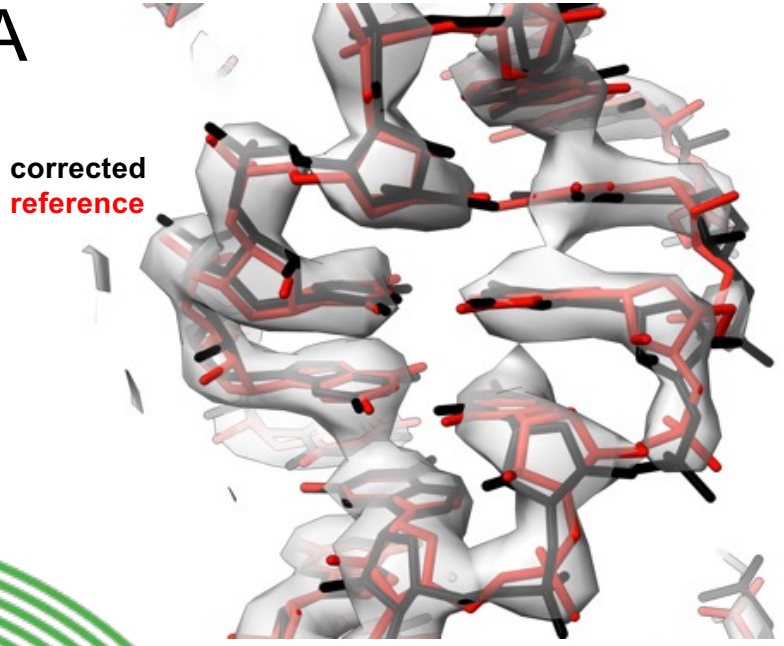
Chojnowski NAR, 2023

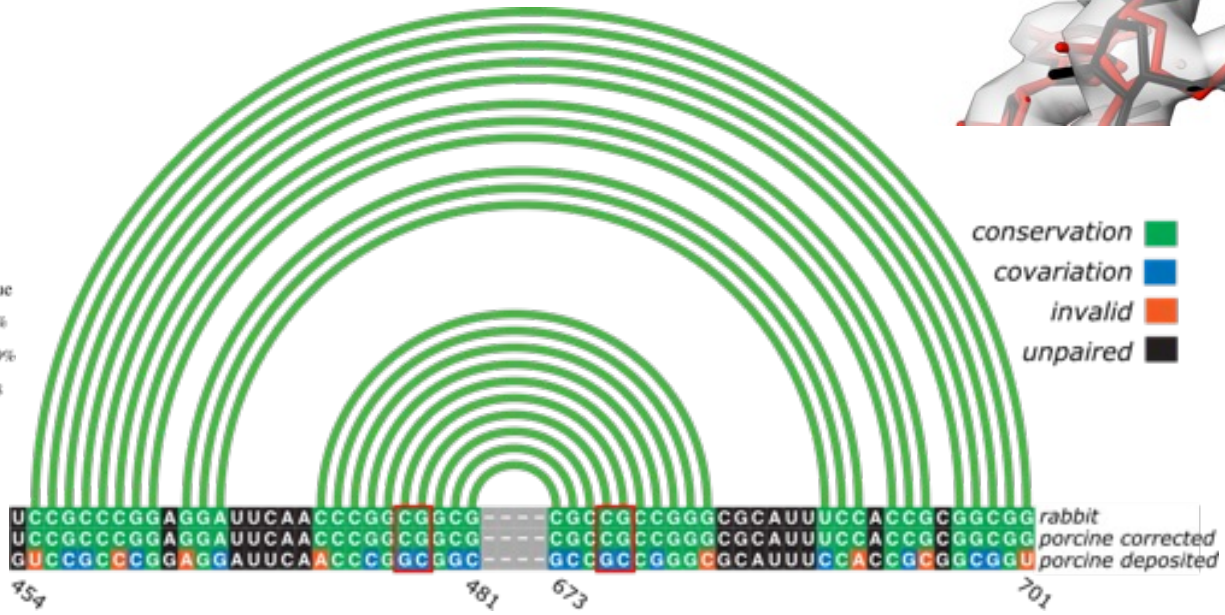
EMBL

#1 Sequence assignment issues in rRNA

expansion segment in an EM model of mammalian ribosome @3.4Å (deposited 2014)







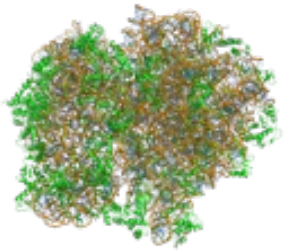
Chojnowski NAR, 2023

EMBL

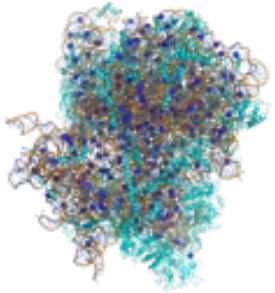
#1 Sequence assignment issues in rRNA

expansion segment in an EM model of mammalian ribosome @3.4Å (deposited 2014)

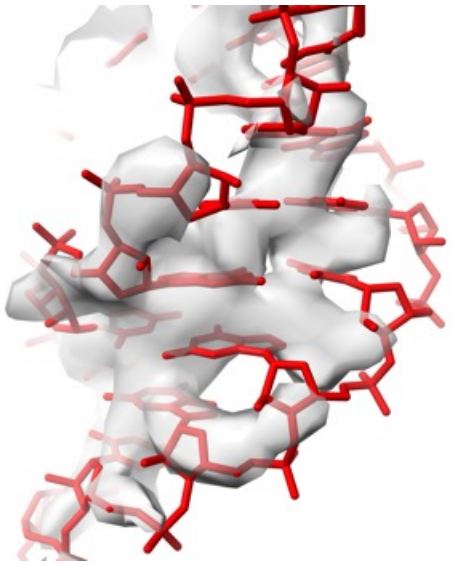
Assemble
RNAfold
COOT



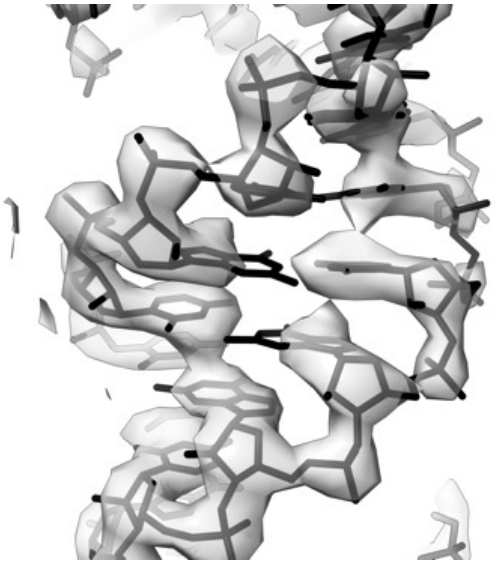
70S MX@3Å



80S MX@4Å



human 28S EM @ 5Å



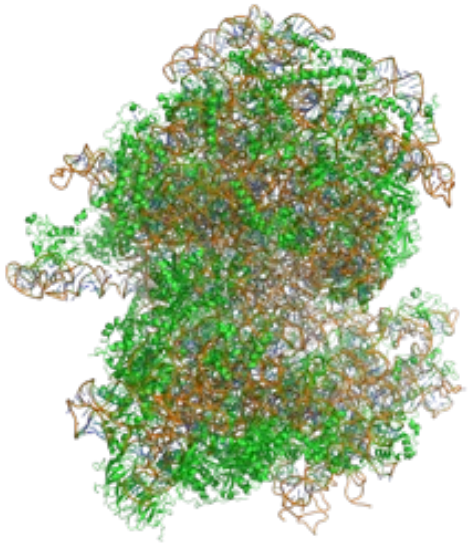
porcine 28S EM @ 3.4Å

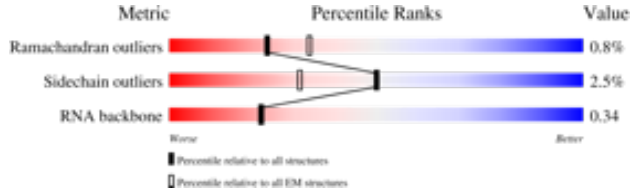
Chojnowski NAR, 2023

EMBL

#2 Sequence assignment issues in rRNA

EM structure model of eukaryotic ribosome @3.5Å (deposited 2017)





model-reference alignment of a 28S rRNA fragment

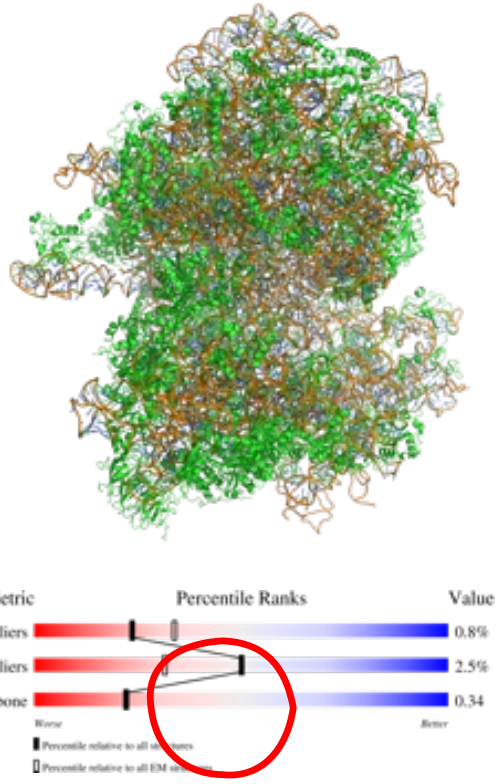
```
model/1-453      1 GGGAUGAACCAAACGUAAUGUUACG-UGCCCAAAUUAACAACUCAUGCAGAUACCAUGAAA  60
reference/1-480  1 GGGAUGAACCAAACGUAAUGUUACGGUGCCCAAAUUAACAACUCAUGCAGAUACCAUGAAA  61

model/1-453     61 GGCGUUGGUUGCUUAAAACAGCAGGACGGUGAUCAUGGAAGUCGAAAUCCGCUAAGGAGUG 121
reference/1-480 62 GGCGUUGGUUGCUUAAAACAGCAGGACGGUGAUCAUGGAAGUCGAAAUCCGCUAAGGAGUG 122

model/1-453     122 UGUAACAACUCACCUGCCGAAGCAACUAGCCCUUAAAAUGGAUGGCGCUUAAGUUGUAUAC 182
reference/1-480 123 UGUAACAACUCACCUGCCGAAGCAACUAGCCCUUAAAAUGGAUGGCGCUUAAGUUGUAUAC 183

model/1-453     183 CUAUACAUUACCGCUAAAGUAGAUGAUUUAUAUUACUUGUGAUAUAAAUUUUGAAACUUUA 243
reference/1-480 184 CUAUACAUUACCGCUAAAGUAGAUGAUUUAUAUUACUUGUGAUAUAAAUUUUGAAACUUUA 244

model/1-453     244 GUGAGUAGGAAGGUACAAUGGUAUGCGUAGAAGUGUUUGGCGUAAGCCUGCAUGGAGCUGC 304
reference/1-480 245 GUGAGUAGGAAGGUACAAUGGUAUGCGUAGAAGUGUUUGGCGUAAGCCUGCAUGGAGCUGC 305

model/1-453     305 CAUUGGUACAGAUCUUGGUGGAUAGUAGCAAAUAAUCGAAUGAGAGCCUUGGAGGACUGAA 365
reference/1-480 306 CAUUGGUACAGAUCUUGGUGGAUAGUAGCAAAUAAUCGAAUGAGAGCCUUGGAGGACUGAA 366

model/1-453     366 GUGGAGAAGGGUUUCGUGUGAACAGUGGUUGAUCACGAGUUAGUCGGUCCUAAGUUCAAGG 426
reference/1-480 367 GUGGAGAAGGGUUUCGUGUGAACAGUGGUUGAUCACGAGUUAGUCGGUCCUAAGUUCAAGG 427

model/1-453     427 CGAAAGC-GAAAAUUUUCAAGUAAAACA-------------------------  453
reference/1-480 428 CGAAAGCCGAAAAUUUUCAAGUAAAACAAAAAUGCCUAACUAUAUAAACAAAG  480
```

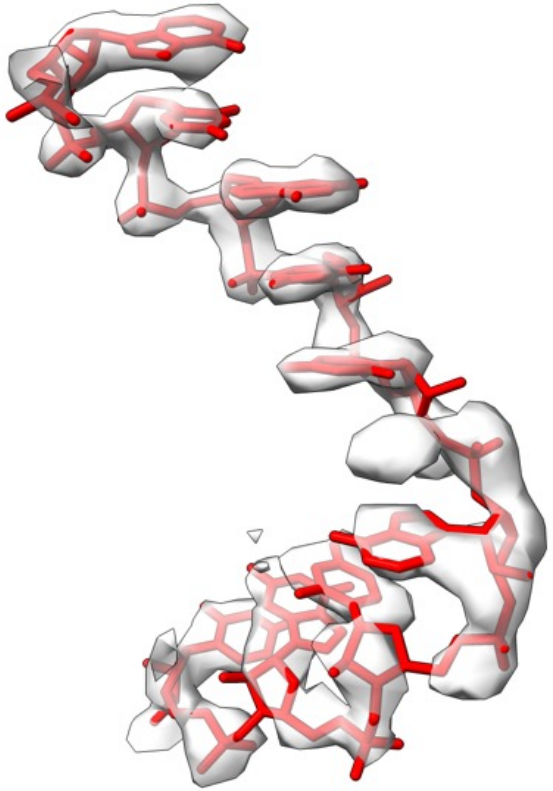
Chojnowski NAR, 2023

EMBL

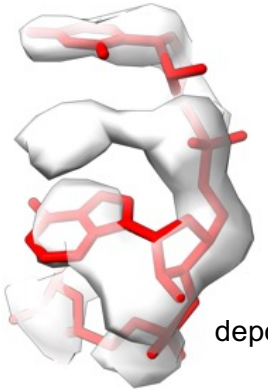
#2 Sequence assignment issues in rRNA

EM structure model of eukaryotic ribosome @3.5Å (deposited 2017)

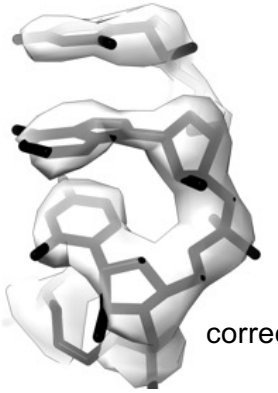




fragment of 28S rRNA



deposited



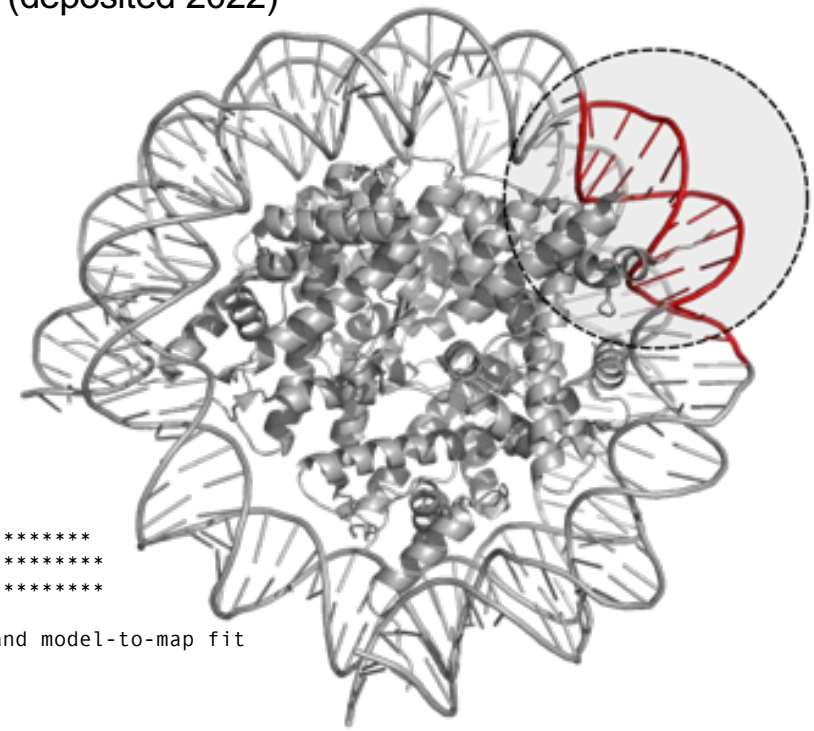
corrected

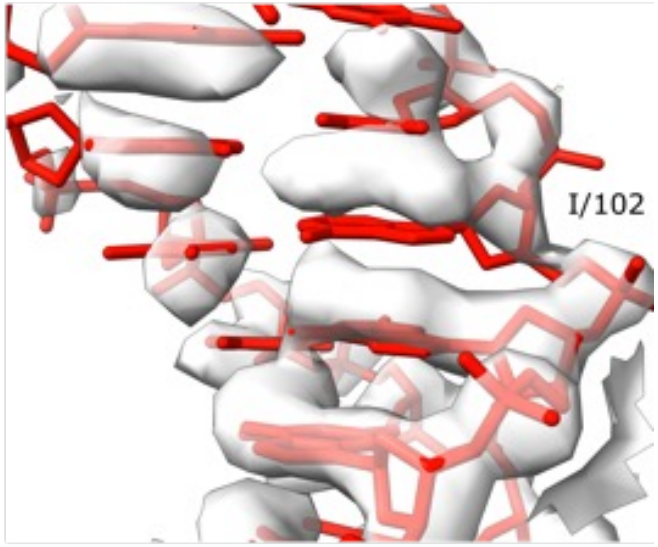
Chojnowski NAR, 2023

EMBL

#3 Sequence assignment issues in DNA

EM model of nucleosome at @3.4Å (deposited 2022)



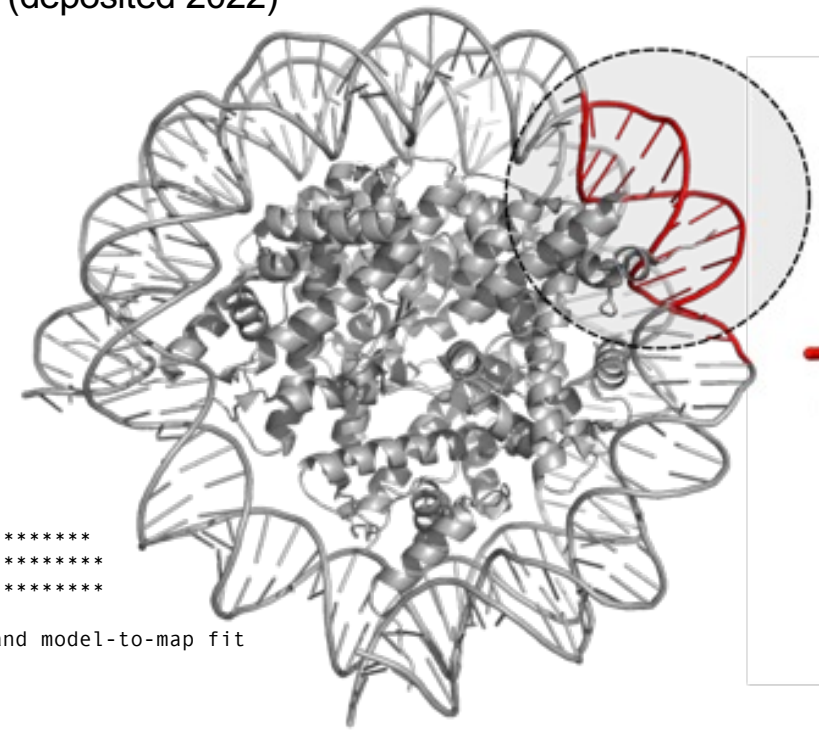


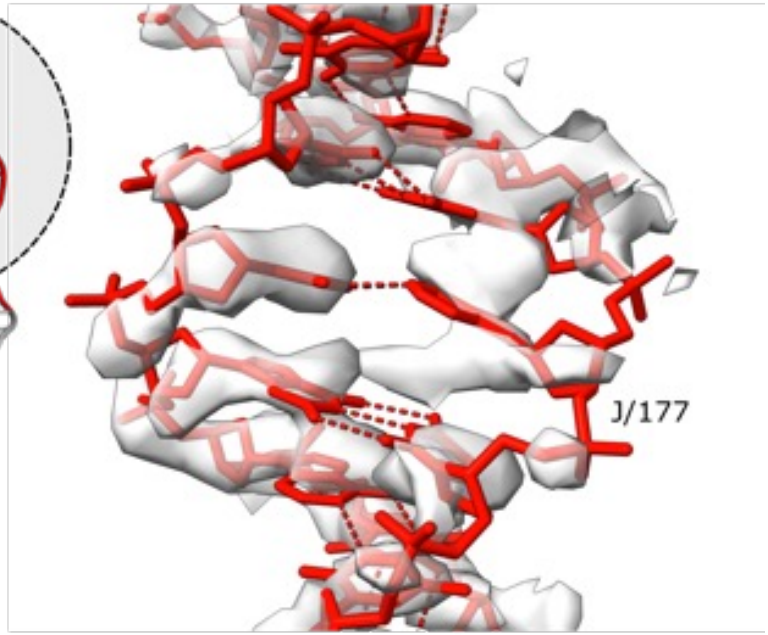

```
*********************************************************
*********************************   SUMMARY   ***********
*********************************************************

 ==> Unidentified chains; check input sequences and model-to-map fit

I/2:144

 ==> Sequence register shifts


 - nucleic-acid chain fragment J/182-211 may be shifted by -1 residue [p-value=9.28e-02]
   model seq 182-211
       actagGGAGTAATCCCCTTGGCGGTTAAAACGCGGgggacag
     new seq 183-212
       actaggGAGTAATCCCCTTGGCGGTTAAAACGCGGGgacag


Time elapsed 0:00:19
```

EMBL

#3 Sequence assignment issues in DNA

EM model of nucleosome at @3.4Å (deposited 2022)

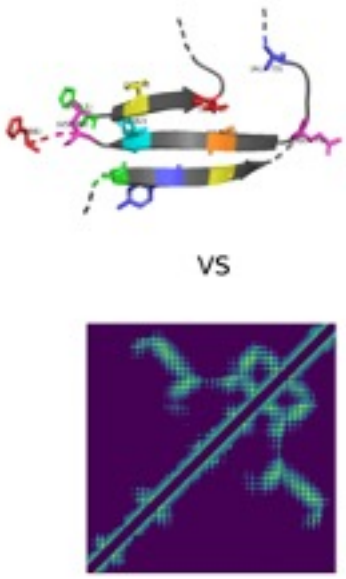


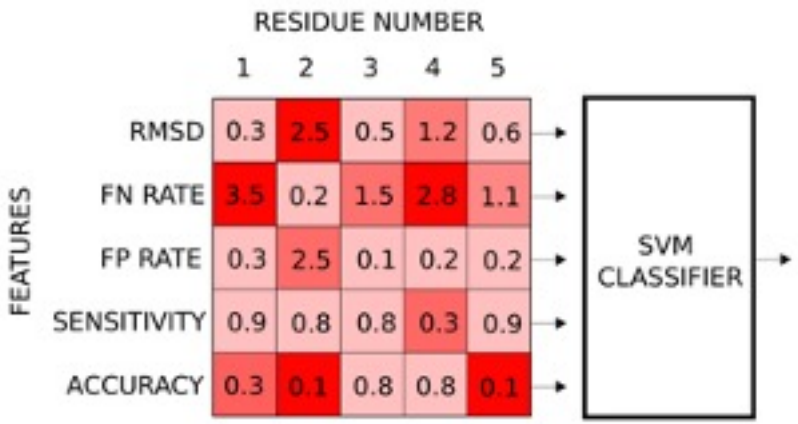


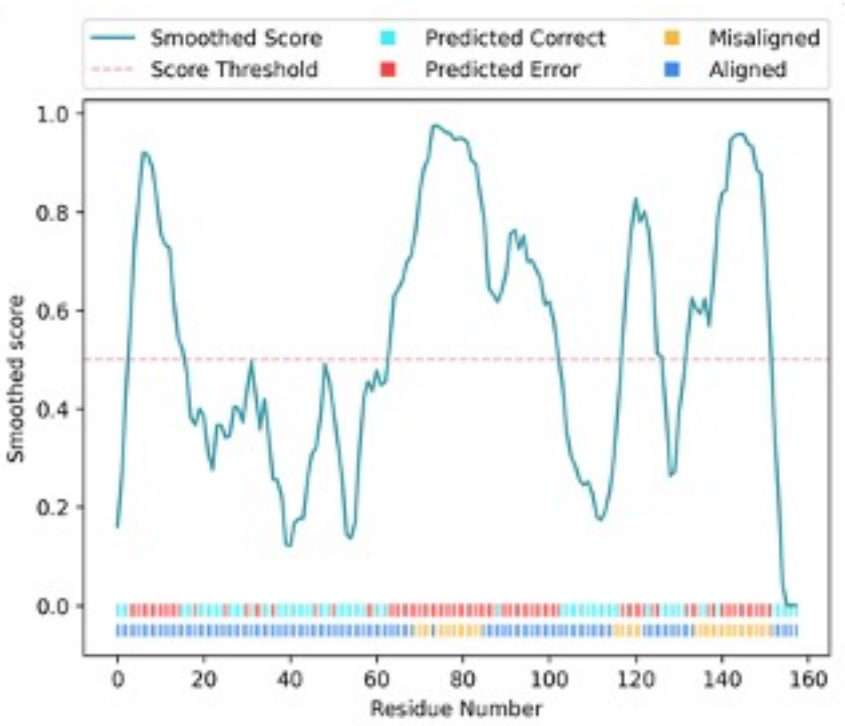

```
*********************************************************
*********************************   SUMMARY   ***********
*********************************************************

 ==> Unidentified chains; check input sequences and model-to-map fit

I/2:144

 ==> Sequence register shifts


 - nucleic-acid chain fragment J/182-211 may be shifted by -1 residue [p-value=9.28e-02]
   model seq 182-211
       actagGGAGTAATCCCCTTGGCGGTTAAAACGCGGgggacag
     new seq 183-212
       actaggGAGTAATCCCCTTGGCGGTTAAAACGCGGGgacag


Time elapsed 0:00:19
```

EMBL

Predicted inter-residue distances for model validation





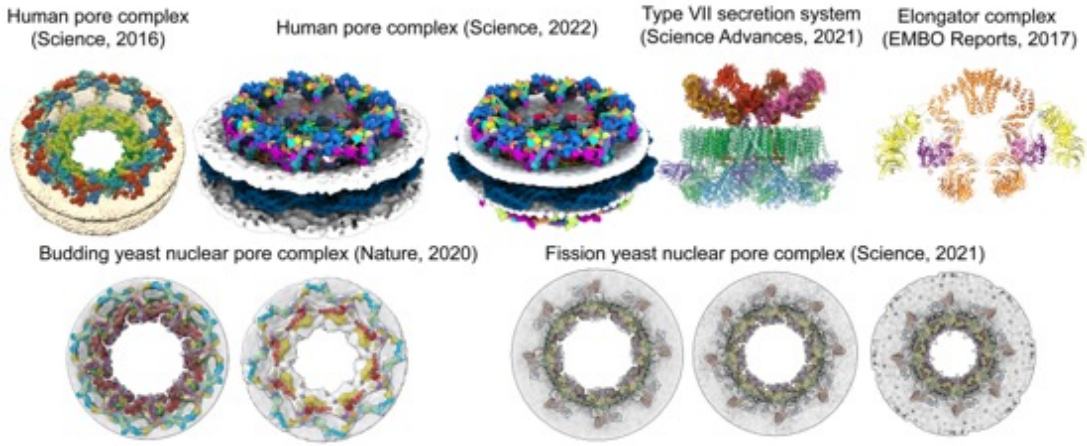


Filomeno Sanchez
Ronan Keegan
Daniel Rigden

www.conkit.org

Filomeno Sánchez Rodríguez et al Acta Cryst. (2022). D**78**

Assembline
Assembly line of macromolecular complexes



Complexes modelled using Assembline

Jan Kosinski group at EMBL Hamburg

https://www.embl-hamburg.de/Assembline/

Nature Protocols, 2022

EMBL

Acknowledgements

University of Liverpool
Daniel Rigden
Adam Simpkin
Filomeno Sánchez Rodríguez

CCP4 Core Team
Ronan Keegan
Charles Ballard
Eugene Krissinel
Kyle Stevenson
Maria Fando

CCPEM team
Agnel Joseph
Tom Burnley
Colin Palmer
Matt Iadanza

Martin Luther University
Panos Kastritis
Ioannis Skalidis

EMBL Hamburg
Matthias Wilmanns
Kate Beckham
Jan Kosiński
Christina Ritter
Edukondalu Mullapudi
Isabel Bento
Alice Bochel

findMySequence
protein sequence identification

checkMySequence
sequence validation

doubleHelix
NA sequence identification

EMBL