Instituto Politécnico Nacional

Escuela Superior de Cómputo

Bioinformatics

Practice 8 - Protein Structure Prediction

Group: 3CV9

Student:
Ramos Diaz Enrique

Professor:
Rosas Trigueros Jorge Luis

Development Date:
December 2nd 2020

Due Date:
December 9th 2020

1 Theoretical Framework

1.1 Critical Assessment of protein Structure Prediction - CASP

Critical Assessment of protein Structure Prediction is a community-wide, worldwide experiment for protein structure prediction taking place every two years since 1994. CASP provides research groups with an opportunity to objectively test their structure prediction methods and delivers an independent assessment of the state of the art in protein structure modeling to the research community and software users [1].

If a given unknown sequence is found to be related by common descent to a protein sequence of known structure (called a *template*), comparative protein modeling may be used to predict the tertiary structure. Templates can be found using sequence alignment methods (e.g. **BLAST** or HHsearch) or **protein threading methods**, which are better in finding distantly related templates. Otherwise, **de novo protein structure prediction** must be applied (e.g. Rosetta), which is much less reliable but can sometimes yield models with the correct fold [1].

Evaluation of the results is carried out in the following prediction categories [1]:

- Tertiary structure prediction (all CASPs)
- Secondary structure prediction (dropped after CASP5)
- Prediction of structure complexes (CASP2 only; a separate experiment — CAPRI — carries on this subject)
- Residue-residue contact prediction (starting CASP4)
- Disordered regions prediction (starting CASP5)
- Model quality assessment (starting CASP7)
- Model refinement (starting CASP7)
- High-accuracy template-based prediction (starting CASP7)

Tertiary structure prediction category was further subdivided into [1]:

1. **Homology modeling**.
2. **Fold recognition**: Also called protein **threading**; Note, this is incorrect as threading is a method.
3. **De novo structure prediction:** Now referred to as 'New Fold' as many methods apply evaluation functions that are biased by knowledge of native protein structures, such as an artificial neural network.

2 Material and Equipment

- UniProt web page [2].
- Swiss-Model web page [3].
- RaptorX web page [4].
- VMD - Visual Molecular Dynamics Software [5].
- CLT75_5425: Uncharacterized protein YukE sequence [6].

3 Practice Development

The objective of this practice it to search for an uncharacterized or unknown protein sequence and apply it the following analyses:

1. **BLAST:** To identify similar well known proteins according to their sequence comparisons, in order to determine in what they might lookalike.
2. **Homology:** To predict the protein structure through their sequence.
3. **Threading:** If the homology process doesn't seems to get satisfactory results (incomplete structure due to incomprehensible parts of its structure), threading is applied to get the full protein structure.

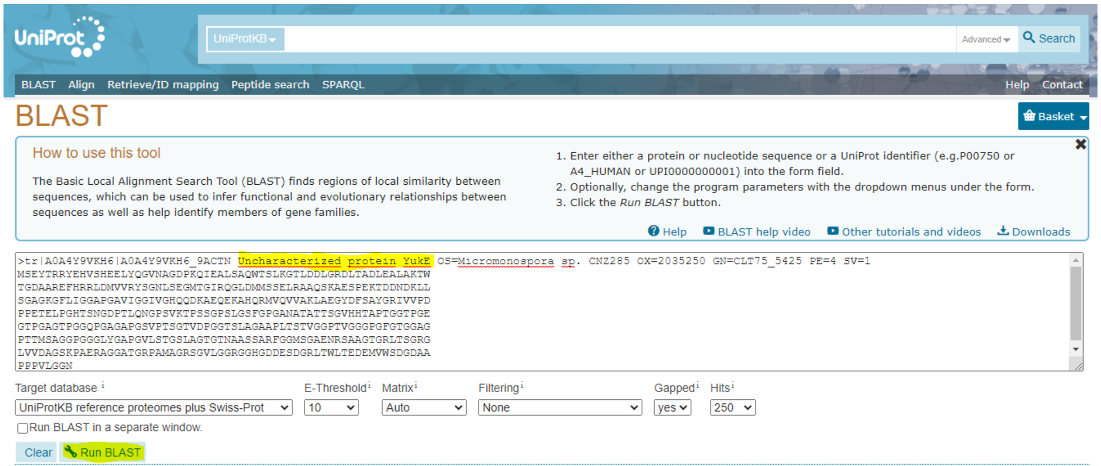
The protein selected for this is the **Uncharacterized protein YukE** with 428 amino acids, located in the CLT75_5425 gene from the **Micromonospora sp. CNZ285** organism [6].

The sequence of this uncharacterized protein in FASTA (canonical) format is obtained from the UniProt web page [2], and is the following:

```
>tr|A0A4Y9VKH6|A0A4Y9VKH6_9ACTN Uncharacterized protein YukE OS=Micromonospora sp. CNZ285 OX=2035250 GN=CLT75_5425 PE=4 SV=1
MSEYTRRYEHVSHEELYQGVNAGDPKQIEALSAQWTSLKGTLDDLGRDLTADLEALAKTW
TGDAAREFHRRLDMVVRYSGNLSEGMTGIRQGLDMMSSELRAAQSKAESPEKTDDNDKLL
SGAGKGFLIGGAPGAVIGGIVGHQQDKAEQEKAHQRMVQVVAKLAEGYDFSAYGRIVVPD
PPETELPGHTSNGDPTLQNGPSVKTPSSGPSLGSFGPGANATATTSGVHHTAPTGGTPGE
GTPGAGTPGGQPGAGAPGSVPTSGTVDPGGTSLAGAAPLTSTVGGPTVGGGPGFGTGGAG
PTTMSAGGPGGGLYGAPGVLSTGSLAGTGTNAASSARFGGMSGAENRSAAGTGRLTSGRG
LVVDAGSKPAERAGGATGRPAMAGRSGVLGGRGGHGDDESDGRLTWLTEDEMVWSDGDAA
PPPVLGGN
```

3.1 BLAST

In the BLAST analysis tool on UnitProt [7], enter the **Uncharacterized protein YukE** sequence in the text area (see Figure 1), run the process by clicking the *Run BLAST* button and wait until it ends.



BLAST

How to use this tool

The Basic Local Alignment Search Tool (BLAST) finds regions of local similarity between sequences, which can be used to infer functional and evolutionary relationships between sequences as well as help identify members of gene families.

1. Enter either a protein or nucleotide sequence or a UniProt identifier (e.g.P00750 or A4_HUMAN or UPI0000000001) into the form field.
2. Optionally, change the program parameters with the dropdown menus under the form.
3. Click the *Run BLAST* button.

Help · BLAST help video · Other tutorials and videos · Downloads

```
>tr|A0A4Y9VKH6|A0A4Y9VKH6_9ACTN Uncharacterized protein YukE OS=Micromonospora sp. CNZ285 OX=2035250 GN=CLT75_5425 PE=4 SV=1
MSEYTRRYEHVSHEELYQGVNAGDPKQIEALSAQWTSLKGTLDDLGRDLTADLEALAKTW
TGDAAREFHRRLDMVVRYSGNLSEGMTGIRQGLDMMSSELRAAQSKAESPEKTDDNDKLL
SGAGKGFLIGGAPGAVIGGIVGHQQDKAEQEKAHQRMVQVVAKLAEGYDFSAYGRIVVPD
PPETELPGHTSNGDPTLQNGPSVKTPSSGPSLGSFGPGANATATTSGVHHTAPTGGTPGE
GTPGAGTPGGQPGAGAPGSVPTSGTVDPGGTSLAGAAPLTSTVGGPTVGGGPGFGTGGAG
PTTMSAGGPGGGLYGAPGVLSTGSLAGTGTNAASSARFGGMSGAENRSAAGTGRLTSGRG
LVVDAGSKPAERAGGATGRPAMAGRSGVLGGRGGHGDDESDGRLTWLTEDEMVWSDGDAA
PPPVLGGN
```

Target database: UniProtKB reference proteomes plus Swiss-Prot · E-Threshold: 10 · Matrix: Auto · Filtering: None · Gapped: yes · Hits: 250

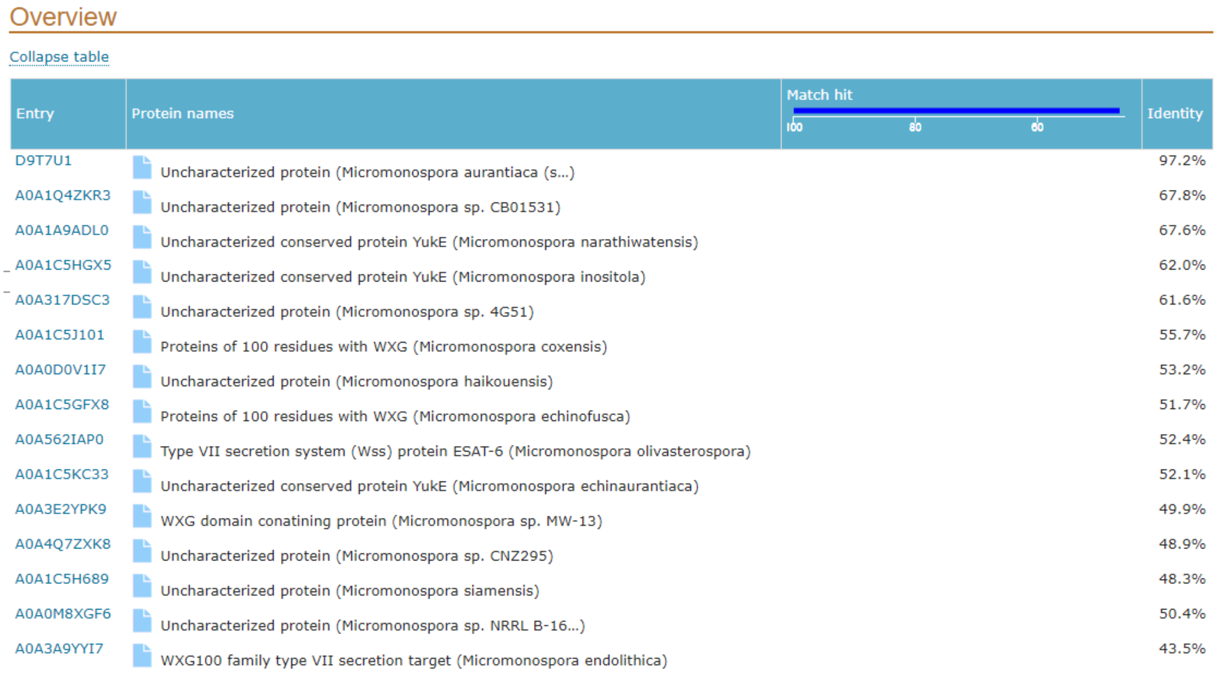
□Run BLAST in a separate window.

Clear · Run BLAST

Figure 1: BLAST tool interface on UnitProt [7].

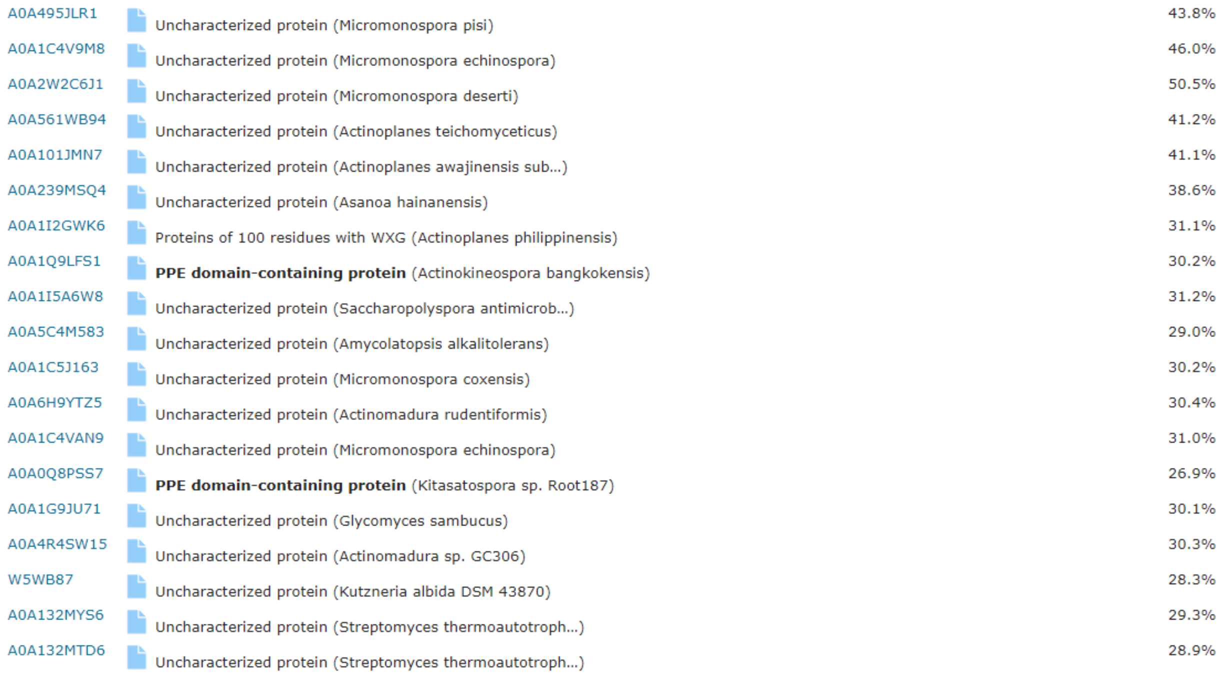
Figures 2 and 3 shows the result of the BLAST analysis, where some of the principal proteins infered by homology that could be related to the uncharacterized one based on their matching sequences are:

- PPE domain-containing protein - *Actinokineospora bangkokensis* **30.2%**
- PPE domain-containing protein - *Kitasatospora sp. Root187* **26.9%**
- PPE domain-containing protein - *Alloactinosynnema sp. L-07* **28.2%**
- PPE domain-containing protein - *Actinokineospora spheciospongiae* **27.2%**



Entry	Protein names	Match hit	Identity
D9T7U1	Uncharacterized protein (Micromonospora aurantiaca (s...)		97.2%
A0A1Q4ZKR3	Uncharacterized protein (Micromonospora sp. CB01531)		67.8%
A0A1A9ADL0	Uncharacterized conserved protein YukE (Micromonospora narathiwatensis)		67.6%
A0A1C5HGX5	Uncharacterized conserved protein YukE (Micromonospora inositola)		62.0%
A0A317DSC3	Uncharacterized protein (Micromonospora sp. 4G51)		61.6%
A0A1C5J101	Proteins of 100 residues with WXG (Micromonospora coxensis)		55.7%
A0A0D0V1I7	Uncharacterized protein (Micromonospora haikouensis)		53.2%
A0A1C5GFX8	Proteins of 100 residues with WXG (Micromonospora echinofusca)		51.7%
A0A562IAP0	Type VII secretion system (Wss) protein ESAT-6 (Micromonospora olivasterospora)		52.4%
A0A1C5KC33	Uncharacterized conserved protein YukE (Micromonospora echinaurantiaca)		52.1%
A0A3E2YPK9	WXG domain conatining protein (Micromonospora sp. MW-13)		49.9%
A0A4Q7ZXK8	Uncharacterized protein (Micromonospora sp. CNZ295)		48.9%
A0A1C5H689	Uncharacterized protein (Micromonospora siamensis)		48.3%
A0A0M8XGF6	Uncharacterized protein (Micromonospora sp. NRRL B-16...)		50.4%
A0A3A9YYI7	WXG100 family type VII secretion target (Micromonospora endolithica)		43.5%

Figure 2: BLAST results for the **Uncharacterized protein YukE** - *Micromonospora sp. CNZ285*.

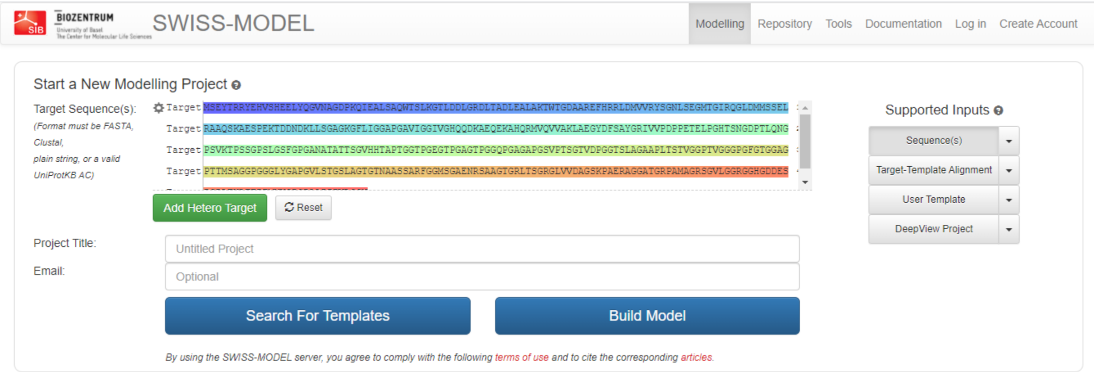


A0A495JLR1	Uncharacterized protein (Micromonospora pisi)	43.8%
A0A1C4V9M8	Uncharacterized protein (Micromonospora echinospora)	46.0%
A0A2W2C6J1	Uncharacterized protein (Micromonospora deserti)	50.5%
A0A561WB94	Uncharacterized protein (Actinoplanes teichomyceticus)	41.2%
A0A101JMN7	Uncharacterized protein (Actinoplanes awajinensis sub...)	41.1%
A0A239MSQ4	Uncharacterized protein (Asanoa hainanensis)	38.6%
A0A1I2GWK6	Proteins of 100 residues with WXG (Actinoplanes philippinensis)	31.1%
A0A1Q9LFS1	**PPE domain-containing protein** (Actinokineospora bangkokensis)	30.2%
A0A1I5A6W8	Uncharacterized protein (Saccharopolyspora antimicrob...)	31.2%
A0A5C4M583	Uncharacterized protein (Amycolatopsis alkalitolerans)	29.0%
A0A1C5J163	Uncharacterized protein (Micromonospora coxensis)	30.2%
A0A6H9YTZ5	Uncharacterized protein (Actinomadura rudentiformis)	30.4%
A0A1C4VAN9	Uncharacterized protein (Micromonospora echinospora)	31.0%
A0A0Q8PSS7	**PPE domain-containing protein** (Kitasatospora sp. Root187)	26.9%
A0A1G9JU71	Uncharacterized protein (Glycomyces sambucus)	30.1%
A0A4R4SW15	Uncharacterized protein (Actinomadura sp. GC306)	30.3%
W5WB87	Uncharacterized protein (Kutzneria albida DSM 43870)	28.3%
A0A132MYS6	Uncharacterized protein (Streptomyces thermoautotroph...)	29.3%
A0A132MTD6	Uncharacterized protein (Streptomyces thermoautotroph...)	28.9%

Figure 3: BLAST results for the **Uncharacterized protein YukE** - *Micromonospora sp. CNZ285* (cont.).

3.2 Homology

Go to the Swiss-Model web page [3] and enter the **Uncharacterized protein YukE** sequence in the *Target Sequence(s)* field (see Figure 4), then click the *Build Model* button to perform the homology process. The results should be ready in a few minutes.



Start a New Modelling Project

Target Sequence(s): (Format must be FASTA, Clustal, plain string, or a valid UniProtKB AC)

```
Target MSEYTRRYEHVSHEELYQGVNAGDPKQIEALSAQWTSLKGTLDDLGRDLTADLEALAKTWTGDAAREFHRRLDMVVRYSGNLSEGMTGIRQGLDMMSSEL
Target RAAQSKAESPEKTDDNDKLLSGAGKGFLIGGAPGAVIGGIVGHQQDKAEQEKAHQRMVQVVAKLAEGYDFSAYGRIVVPDPPETELPGHTSNGDPTLQNG
Target PSVKTPSSGPSLGSFGPGANATATTSGVHHTAPTGGTPGEGTPGAGTPGGQPGAGAPGSVPTSGTVDPGGTSLAGAAPLTSTVGGPTVGGGPGFGTGGAG
Target PTTMSAGGPGGGLYGAPGVLSTGSLAGTGTNAASSARFGGMSGAENRSAAGTGRLTSGRGLVVDAGSKPAERAGGATGRPAMAGRSGVLGGRGGHGDDES
```

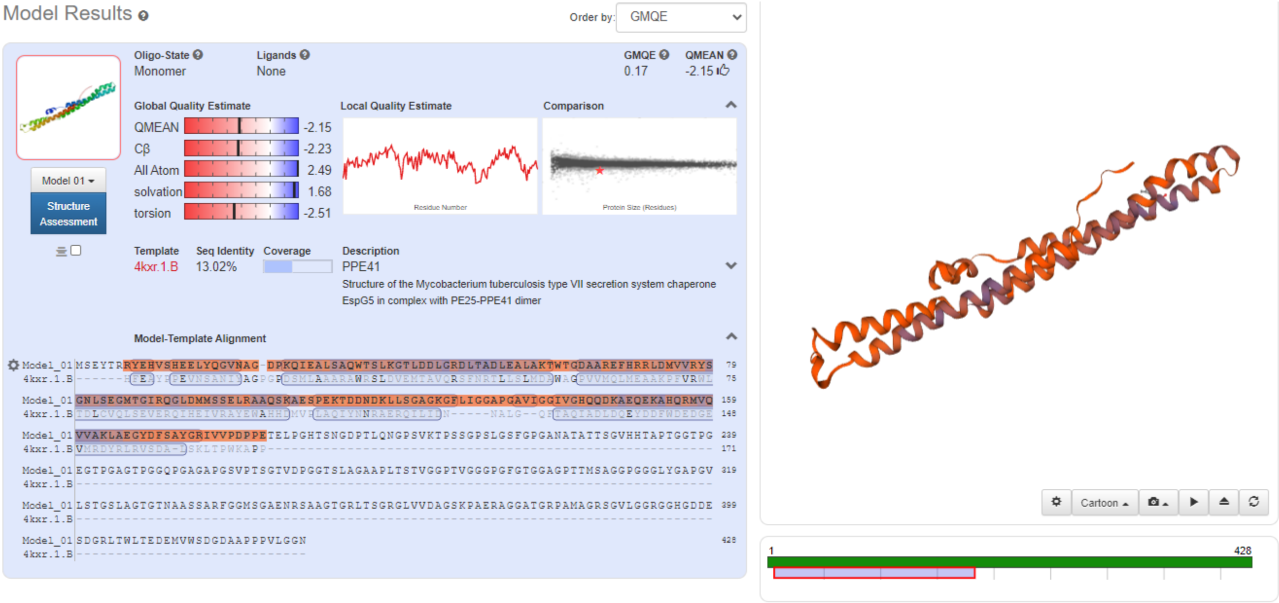
Project Title: Untitled Project

Email: Optional

By using the SWISS-MODEL server, you agree to comply with the following terms of use and to cite the corresponding articles.

Figure 4: Swiss-Model interface [3]

When the homology process is done, a page that is shown in Figure 5 is displayed with the generated models of this sequence. Also, a lot of templates are generated too, Figure 6 shows some of them.



Model Results

Order by: GMQE

Oligo-State	Ligands	GMQE	QMEAN
Monomer	None	0.17	-2.15

Global Quality Estimate

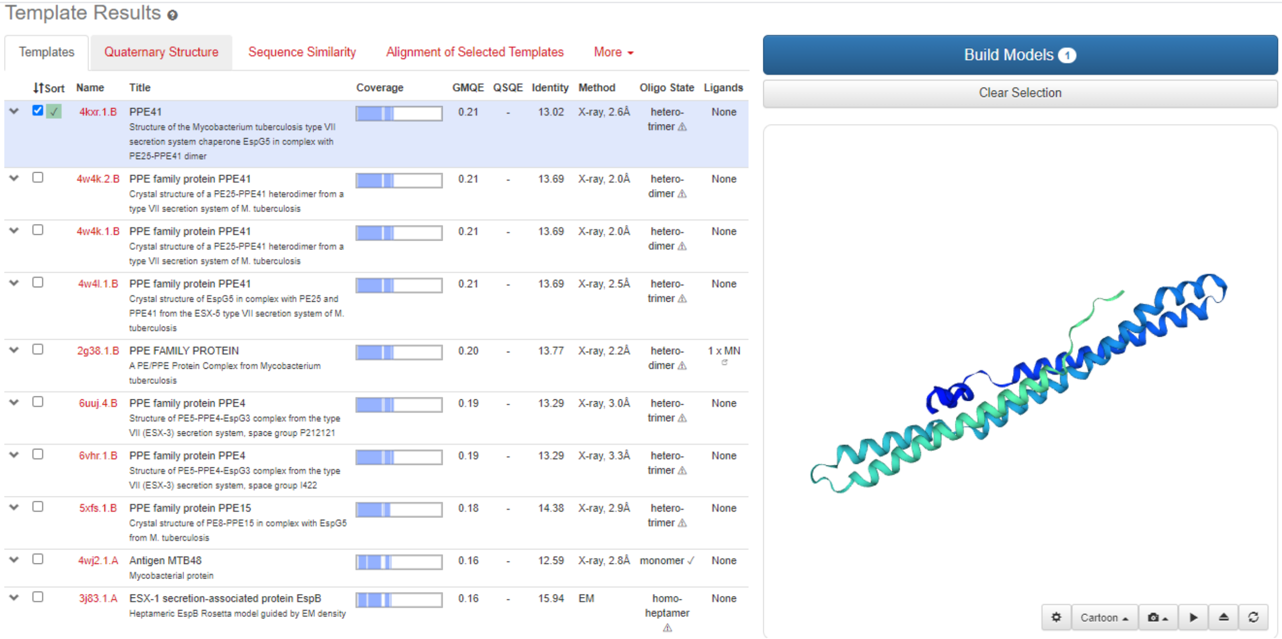
QMEAN	-2.15
Cβ	-2.23
All Atom	2.49
solvation	1.68
torsion	-2.51

Template	Seq Identity	Coverage	Description
4kxr.1.B	13.02%		PPE41 Structure of the Mycobacterium tuberculosis type VII secretion system chaperone EspG5 in complex with PE25-PPE41 dimer

Model-Template Alignment

```
Model_01 MSEYTRRYEHVSHEELYQGVNAG-DPKQIEALSAQWTSLKGTLDDLGRDLTADLEALAKTWTGDAAREFHRRLDMVVRYS  79
4kxr.1.B ------HFEAYPPEVNSANIYAGPGPDSMLAAARAWRSLDVEMTAVQRSFNRTLLSLMDAWAGPVVMQLMEAAKPFVRWL  75

Model_01 GNLSEGMTGIRQGLDMMSSELRAAQSKAESPEKTDDNDKLLSGAGKGFLIGGAPGAVIGGIVGHQQDKAEQEKAHQRMVQ 159
4kxr.1.B TDLCVQLSEVERQIHEIVRAYEWAHHDMVRLAQIYNNRAERQILIDN-----NALG--QFTAQIADLDQEYDDFWDEDGE 148

Model_01 VVAKLAEGYDFSAYGRIVVPDPPETELPGHTSNGDPTLQNGPSVKTPSSGPSLGSFGPGANATATTSGVHHTAPTGGTPG 239
4kxr.1.B VMRDYRLRVSDA-DSKLTPWKAPP-------------------------------------------------------- 171

Model_01 EGTPGAGTPGGQPGAGAPGSVPTSGTVDPGGTSLAGAAPLTSTVGGPTVGGGPGFGTGGAGPTTMSAGGPGGGLYGAPGV 319
4kxr.1.B --------------------------------------------------------------------------------

Model_01 LSTGSLAGTGTNAASSARFGGMSGAENRSAAGTGRLTSGRGLVVDAGSKPAERAGGATGRPAMAGRSGVLGGRGGHGDDE 399
4kxr.1.B --------------------------------------------------------------------------------

Model_01 SDGRLTWLTEDEMVWSDGDAAPPPVLGGN                                                    428
4kxr.1.B -----------------------------
```

Figure 5: Homology result for the **Uncharacterized protein YukE** - *Micromonospora sp. CNZ285*.



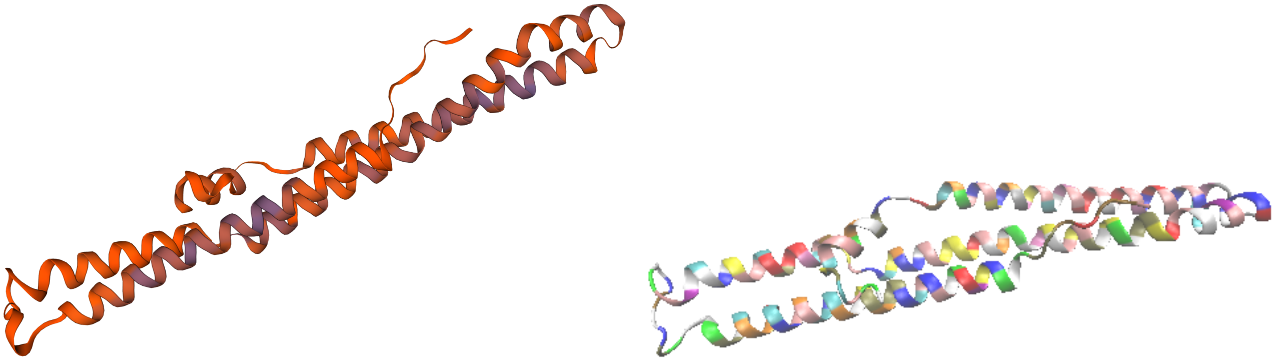
Template Results

Templates | Quaternary Structure | Sequence Similarity | Alignment of Selected Templates | More

	Sort	Name	Title	Coverage	GMQE	QSQE	Identity	Method	Oligo State	Ligands
	☑ ✓	4kxr.1.B	PPE41 Structure of the Mycobacterium tuberculosis type VII secretion system chaperone EspG5 in complex with PE25-PPE41 dimer		0.21	-	13.02	X-ray, 2.6Å	hetero-trimer ⚠	None
	☐	4w4k.2.B	PPE family protein PPE41 Crystal structure of a PE25-PPE41 heterodimer from a type VII secretion system of M. tuberculosis		0.21	-	13.69	X-ray, 2.0Å	hetero-dimer ⚠	None
	☐	4w4k.1.B	PPE family protein PPE41 Crystal structure of a PE25-PPE41 heterodimer from a type VII secretion system of M. tuberculosis		0.21	-	13.69	X-ray, 2.0Å	hetero-dimer ⚠	None
	☐	4w4l.1.B	PPE family protein PPE41 Crystal structure of EspG5 in complex with PE25 and PPE41 from the ESX-5 type VII secretion system of M. tuberculosis		0.21	-	13.69	X-ray, 2.5Å	hetero-trimer ⚠	None
	☐	2g38.1.B	PPE FAMILY PROTEIN A PE/PPE Protein Complex from Mycobacterium tuberculosis		0.20	-	13.77	X-ray, 2.2Å	hetero-dimer ⚠	1 x MN
	☐	6uuj.4.B	PPE family protein PPE4 Structure of PE5-PPE4-EspG3 complex from the type VII (ESX-3) secretion system, space group P212121		0.19	-	13.29	X-ray, 3.0Å	hetero-trimer ⚠	None
	☐	6vhr.1.B	PPE family protein PPE4 Structure of PE5-PPE4-EspG3 complex from the type VII (ESX-3) secretion system, space group I422		0.19	-	13.29	X-ray, 3.3Å	hetero-trimer ⚠	None
	☐	5xfs.1.B	PPE family protein PPE15 Crystal structure of PE8-PPE15 in complex with EspG5 from M. tuberculosis		0.18	-	14.38	X-ray, 2.9Å	hetero-trimer ⚠	None
	☐	4wj2.1.A	Antigen MTB48 Mycobacterial protein		0.16	-	12.59	X-ray, 2.8Å	monomer ✓	None
	☐	3j83.1.A	ESX-1 secretion-associated protein EspB Heptameric EspB Rosetta model guided by EM density		0.16	-	15.94	EM	homo-heptamer ⚠	None

Figure 6: Generated templates for the **Uncharacterized protein YukE** - *Micromonospora sp. CNZ285*.

The resulted model from the homology for the **Uncharacterized protein YukE** can be seen in Figure 7, where its corresponding PDB file is visualized with the same Swiss-Model web page [3] and with the VMD software [5].



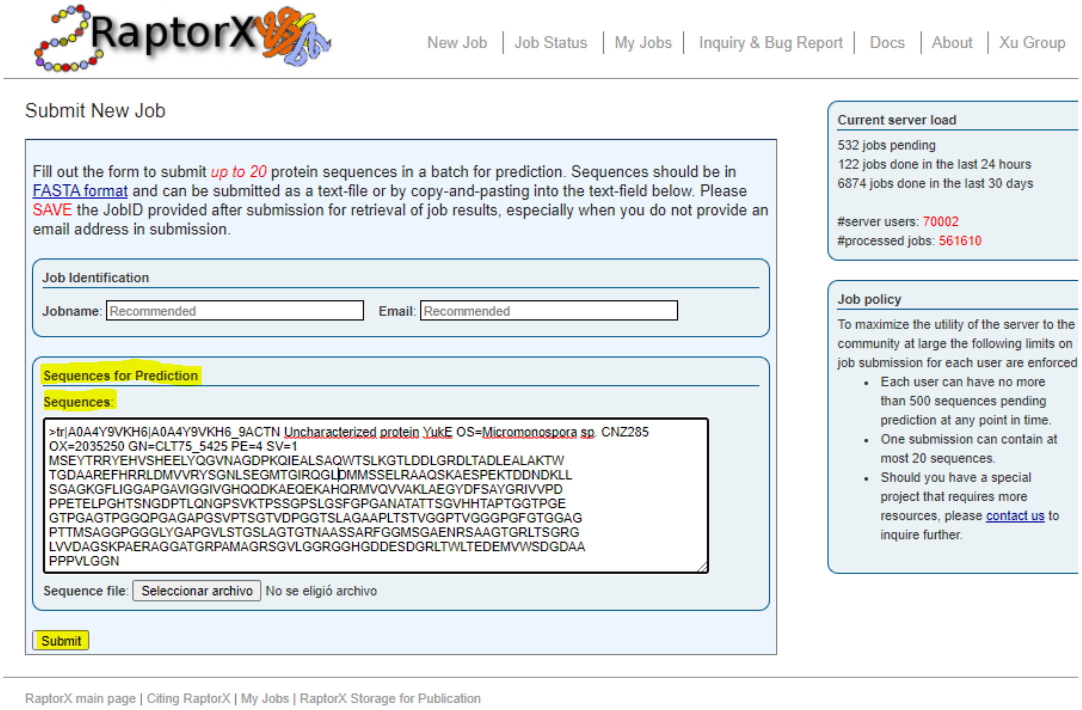
(a) Homology Model in Swiss-Model [3].

(b) Homology Model in VMD [5].

Figure 7: PDB file for the **Uncharacterized protein YukE** - *Micromonospora sp. CNZ285* by the Homology process.

3.3 Threading

Go to the RaptorX web page [4] and enter the **Uncharacterized protein YukE** sequence in the Sequences for Prediction field (see Figure 8), then click the *Submit* button to perform the threading process. Due the high demand and plenty of pending jobs in queue (RaptorX [4] is open free for anyone who wants to perform their sequence analyses, from biology scientists to students and even common people), this process may take hours or even days to complete.



RaptorX
New Job | Job Status | My Jobs | Inquiry & Bug Report | Docs | About | Xu Group

Submit New Job

Fill out the form to submit *up to 20* protein sequences in a batch for prediction. Sequences should be in FASTA format and can be submitted as a text-file or by copy-and-pasting into the text-field below. Please SAVE the JobID provided after submission for retrieval of job results, especially when you do not provide an email address in submission.

Job Identification

Jobname: Recommended   Email: Recommended

Sequences for Prediction

Sequences:

```
>tr|A0A4Y9VKH6|A0A4Y9VKH6_9ACTN Uncharacterized protein YukE OS=Micromonospora sp. CNZ285
OX=2035250 GN=CLT75_5425 PE=4 SV=1
MSEYTRRYEHVSHEELYQGVNAGDPKQIEALSAQWTSLKGTLDDLGRDLTADLEALAKTW
TGDAAREFHRRLDMVVRYSGNLSEGMTGIRQGLDMMSSELRAAQSKAESPEKTDDNDKLL
SGAGKGFLIGGAPGAVIGGIVGHQQDKAEQEKAHQRMVQVVAKLAEGYDFSAYGRIVVPD
PPETELPGHTSNGDPTLQNGPSVKTPSSGPSLGSFGPGANATATTSGVHHTAPTGGTPGE
GTPGAGTPGGQPGAGAPGSVPTSGTVDPGGTSLAGAAPLTSTVGGPTVGGGPGFGTGGAG
PTTMSAGGPGGGLYGAPGVLSTGSLAGTGTNAASSARFGGMSGAENRSAAGTGRLTSGRG
LVVDAGSKPAERAGGATGRPAMAGRSGVLGGRGGHGDDESDGRLTWLTEDEMVWSDGDAA
PPPVLGGN
```

Sequence file: Seleccionar archivo No se eligió archivo

Submit

Current server load

532 jobs pending
122 jobs done in the last 24 hours
6874 jobs done in the last 30 days

#server users: 70002
#processed jobs: 561610

Job policy

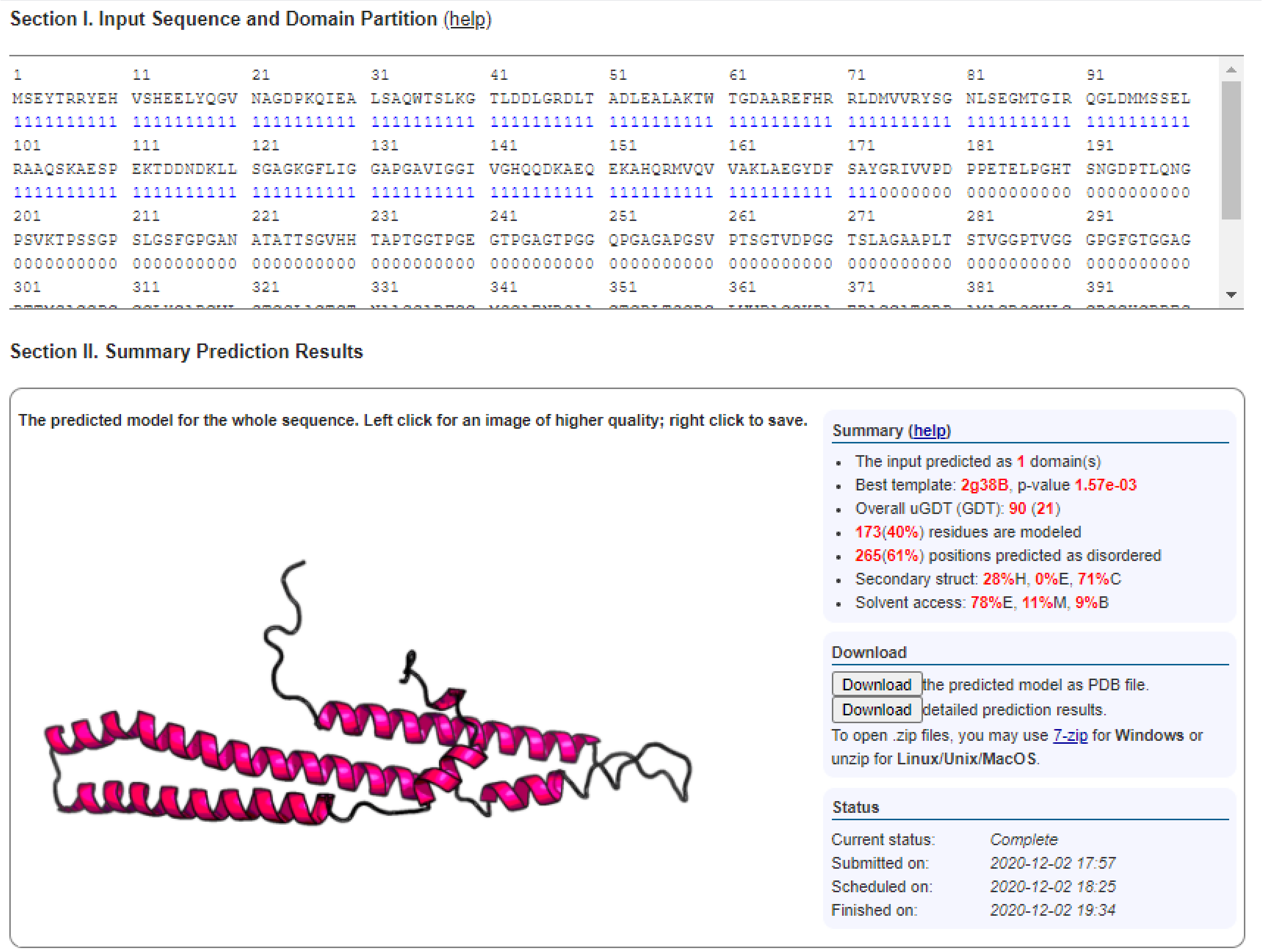
To maximize the utility of the server to the community at large the following limits on job submission for each user are enforced

- Each user can have no more than 500 sequences pending prediction at any point in time.
- One submission can contain at most 20 sequences.
- Should you have a special project that requires more resources, please contact us to inquire further.

RaptorX main page | Citing RaptorX | My Jobs | RaptorX Storage for Publication

Figure 8: RaptorX interface [4]

When the homology process is done, a page that is shown in Figure 5 is displayed with the generated model of this sequence.



Section I. Input Sequence and Domain Partition (help)

```
1          11         21         31         41         51         61         71         81         91
MSEYTRRYEH VSHEELYQGV NAGDPKQIEA LSAQWTSLKG TLDDLGRDLT ADLEALAKTW TGDAAREFHR RLDMVVRYSG NLSEGMTGIR QGLDMMSSEL
1111111111 1111111111 1111111111 1111111111 1111111111 1111111111 1111111111 1111111111 1111111111 1111111111
101        111        121        131        141        151        161        171        181        191
RAAQSKAESP EKTDDNDKLL SGAGKGFLIG GAPGAVIGGI VGHQQDKAEQ EKAHQRMVQV VAKLAEGYDF SAYGRIVVPD PPETELPGHT SNGDPTLQNG
1111111111 1111111111 1111111111 1111111111 1111111111 1111111111 1111111111 1110000000 0000000000 0000000000
201        211        221        231        241        251        261        271        281        291
PSVKTPSSGP SLGSFGPGAN ATATTSGVHH TAPTGGTPGE GTPGAGTPGG QPGAGAPGSV PTSGTVDPGG TSLAGAAPLT STVGGPTVGG GPGFGTGGAG
0000000000 0000000000 0000000000 0000000000 0000000000 0000000000 0000000000 0000000000 0000000000 0000000000
301        311        321        331        341        351        361        371        381        391
```

Section II. Summary Prediction Results

The predicted model for the whole sequence. Left click for an image of higher quality; right click to save.

Summary (help)

- The input predicted as 1 domain(s)
- Best template: 2g38B, p-value 1.57e-03
- Overall uGDT (GDT): 90 (21)
- 173(40%) residues are modeled
- 265(61%) positions predicted as disordered
- Secondary struct: 28%H, 0%E, 71%C
- Solvent access: 78%E, 11%M, 9%B

Download

Download the predicted model as PDB file.

Download detailed prediction results.

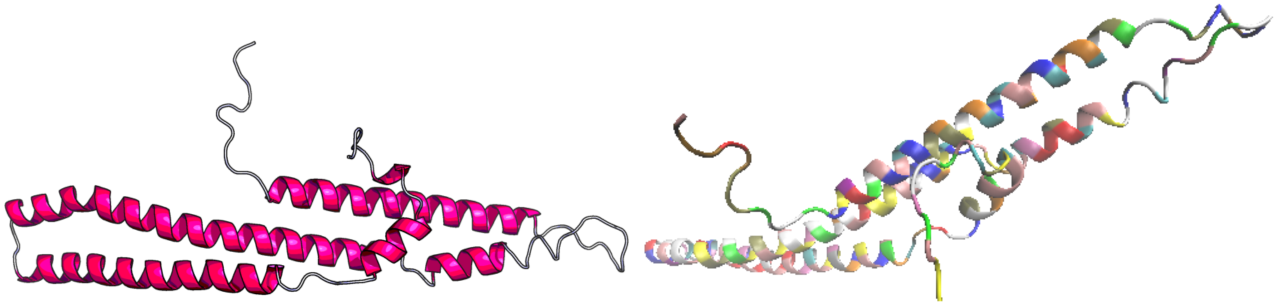
To open .zip files, you may use 7-zip for **Windows** or unzip for **Linux/Unix/MacOS**.

Status

Current status:	*Complete*
Submitted on:	*2020-12-02 17:57*
Scheduled on:	*2020-12-02 18:25*
Finished on:	*2020-12-02 19:34*

Figure 9: Threading result for the **Uncharacterized protein YukE** - *Micromonospora sp. CNZ285*.

The resulted model from the threading for the **Uncharacterized protein YukE** can be seen in Figure 10, where its corresponding PDB file is visualized with the same RaptorX web page [4] and with the VMD software [5].



(a) Threading Model in RaptorX [4].

(b) Threading Model in VMD [5].

Figure 10: PDB file for the **Uncharacterized protein YukE** - *Micromonospora sp. CNZ285* by the Threading process.

4 Conclusions and recommendations

Protein structure prediction is one of the most challenging problems to solve. Fortunately there are several prediction methods that make use of the computational power provided by the computers hardware. The most accurate prediction process performed in this practice was threading; however, the most accurate one of all is the de novo structure prediction, because it's starting to use relative new technologies, such as artificial neuronal networks (deep learning).

5 References

[1] Wikipedia, "CASP," https://en.wikipedia.org/wiki/CASP, [Online; last access December 7, 2020].

[2] T. U. Consortium, "UniProt: a worldwide hub of protein knowledge," https://www.uniprot.org/, [Online; last access November 21, 2020].

[3] B. U. of Bazel, "SWISS-MODEL Interactive Workspace," https://swissmodel.expasy.org/interactive, [Online; last access December 4, 2020].

[4] U. of Chicago, "RaptorX: a protein structure and function prediction server," http://raptorx.uchicago.edu/StructurePrediction/predict/, [Online; last access December 4, 2020].

[5] "Download VMD," https://www.ks.uiuc.edu/Development/Download/download.cgi?PackageName=VMD, [Online; last access October 13, 2020].

[6] T. U. Consortium, "CLT75_5425 - Uncharacterized protein YukE - Micromonospora sp. CNZ285 - CLT75_5425 gene protein," https://www.uniprot.org/uniprot/A0A4Y9VKH6, [Online; last access December 4, 2020].

[7] ——, "BLAST," https://www.uniprot.org/blast/, [Online; last access November 21, 2020].