BTCH90009

Assignment 1: Analysing protein structure and function

Name: Jiayu Wang

Student ID: 1039580

Word Count: 1497

Introduction

Protein identification and analysis are basic processes for protein study. This is important for species identification, taxonomy and evolutionary relationship study, especially when there is only peptide sequence but no DNA sequence available.

Previous ZEPO spy obtained two protein samples in Tube A and Tube B. After mass spectrometry analysis and Edman degradation, it is discovered that Tube A mixed four different protein components. Tube B was found in a special lab full of unusual sea creatures.

In this area, NCBI-BLAST is an essential tool to perform sequence alignment to support the downstream studies [1]. It applies local alignment algorithm instead of the commonly used Smith-Waterman algorithm to reduce time complexity and hence significantly reduce aligning time. NCBI-BLAST supports the accessibility to the most comprehensive databases to provide a more reliable matching result. Using BLAST, this research identified protein components in the two tubes and exhibited the potential purpose of these proteins to help with the national security work.

Methods

Protein Samples

Samples for Tube A are four isolated and degraded peptides, with only partial sequences provided. Sample for Tube B is a small peptide with sequence LWAELQRGCCSDPRCNYDHPEICN.

Tools

NCBI-BLAST and accessed databases

We used Protein BLAST (blastp) to align the four isolated peptide sequences in Tube A by the access to non-redundant (NR) database [2, 3]. We used Position-Specific Iterative BLAST (PSI-BLAST) in NR database for Tube B peptide to perform an iteration alignment to find all the sequences in the database with a particular conserved sequence motif.

Protein Data Bank (PDB)

We accessed to PDB to study the structure of searched proteins [4].

ScanPROSITE

We applied ScanPROSITE to assist the protein study by matching proteins against PROSITE motif collections [5]. By comparing the differently sourced information we gained more confidence about the conclusion.

Results

Tube A

(i) The nature of the four proteins

Protein 1: Glycoprotein 1 [*Machupo mammarenavirus*]

Protein 2: Tetanus toxin [*Clostridium tetani*]

Protein 3: Matrix protein VP40 [*Ebola virus*]

Protein 4: DNA topoisomerase 1 [*Variola virus*]

(ii) Further analysis of Protein 4

The BLASTP result of Protein 4 is shown in **Table 1.**

Table 1. The top 5 sequences of the protein 4 BLAST results

Sequence ID	Name	Identity	Similarity	Score	E value
2H7F_X	Structure of variola topoisomerase covalently bound to DNA [*Variola* virus]	49/49 (100%)	49/49 (100%)	108	8e-24
2H7G_X	Structure of variola topoisomerase non-covalently bound to DNA [*Variola* virus]	49/49 (100%)	49/49 (100%)	108	8e-24
NP_042133.1	DNA topoisomerase type I [*Variola* virus]	49/49 (100%)	49/49 (100%)	107	9e-24
APR62827.1	DNA topoisomerase type I [*Variola* virus]	48/49 (98%)	48/49 (97%)	106	4e-23
ABF26069.1	topoisomerase type IB [*Variola* virus]	48/49 (98%)	48/49 (97%)	100	6e-23

Scores represent the matching percentage to show how close the searched sequence matches with the subject sequence. E value represents the number of sequence alignments that are expected by chance. For the identified DNA topoisomerase 1 [*Variola virus*], the similarity and identity are both 100%, the score is high up to 108, the E value is close to 0 at 8e-24. Besides, the top five records are all *Variola virus* DNA topoisomerase 1 with high scores, high identities, and low E values. Thus, we can confidently identify protein 4 as DNA topoisomerase 1 [*Variola virus*].

The characteristics of type IB topoisomerase from *Variola virus*

Variola virus belongs to *Poxviridae virus* family and causes smallpox, an infectious disease with 20% - 30% mortality [6]. Type IB topoisomerase (TopIB) is highly conserved and essential to transcribe virus DNA [7]. TopIB is one of the known smallest topoisomerases, with a size of only 34 kDa [8]. It exhibits a strong preference for specific DNA sequences that contain the pentamer 5′-(T/C)CCTT-3′. As shown in **Fig.1a** and **Fig.1b,** TopIB introduces a cleavage on one strand of DNA and subsequently allows the flanking duplexes to rotate.

The chemical process of this TopIB reaction cycle can be further explained in detail. The initial state of the TopIB is wrapping about the substrate DNA duplex [9]. When the pentamer is identified, the active

site Tyr274 on TopIB attacks the phosphodiester on the flanking 3′ side of the identified pentamer to form a covalent linkage. This linkage permits the rotation of the supercoiled duplex and leads to DNA relaxation [10].

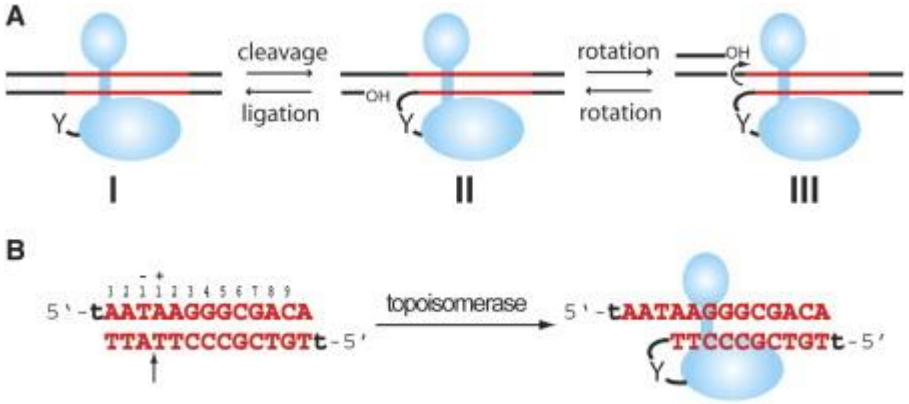

Fig.1 TopIB reaction cycle. [11] **a.** TopIB performs cleavage and leads to supercoil relaxation. **b.** The active site of TopIB

Information on TopIB structure

From searched information from Protein Data Bank (PDB), we can conclude that TopIB from *Variola virus is* monomer protein. It consists two domains composing a C-shaped clamp about the DNA duplex as shown in **Fig.2a** and **Fig.2b** [11]. Both domains contain α helices and β sheets. Besides, activated TopIB forms a covalent linkage between Tyr274 and +1 as shown in **Fig.2c** and **Fig.2d**.

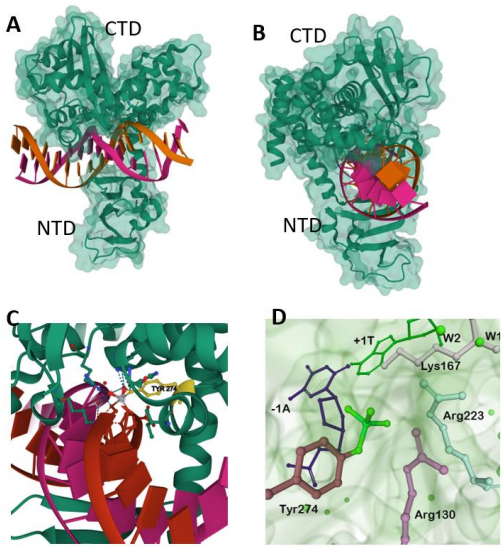

Fig.2 Structure of *Variola* virus TopIB-DNA complex. **a.** Overview from the front side. NTD: N-terminal domain. CTD: C-terminal domain. **b.** Overview from another angle. **c.** Covalent reaction between Tyr274 and +1 phosphate **d.** The same covalent reaction with **c** in a ball-stick representation style

Tube B

(i) Peptide identification

The top eight sequence records from iteration 1 are listed in **Table 2**.

Table 2. The top eight sequence records for Tube B peptide

Sequence ID	Name	Identity	Similarity	Score/bits	E value
2H8S_A	Solution structure of alpha-conotoxin Vc1.1 [synthetic construct]	16/16 (100%)	16/16 (100%)	60.4	2e-10
4TTL_A	Racemic structure of cyclic Vc1.1 (cVc1.1-1) [*Conus victoriae*]	16/16 (100%)	16/16 (100%)	60.4	4e-10
P69747.1	Alpha-conotoxin Vc1a [*Conus victoriae*]	17/17 (100%)	17/17 (100%)	63.8	4e-10
2MG6_A	[3,16]-trans dicarba Vc1.1 [*Conus victoriae*]	14/15 (93%)	14/15 (93%)	51.5	8e-07
2MFX_A	[2,8]-cis dicarba Vc1.1 [*Conus victoriae*]	14/16 (88%)	14/16 (87%)	51.1	1e-06
6CGX_A	Backbone cyclised conotoxin Vc1.1 mutant - D11A, E14A [*Conus victoriae*]	14/16 (88%)	14/16 (87%)	51.5	1e-06
L8BU87.1	Alpha-conotoxin LvIA [*Conus lividus*]	14/17 (82%)	14/17 (82%)	45.2	0.001
2N07_X	Design of a Highly Stable Disulfide-Deleted Mutant of Analgesic Cyclic alpha-Conotoxin Vc1.1 [*Conus victoriae*]	13/14 (93%)	13/14 (92%)	43.9	0.001
2MDQ_A	A Novel 4/7-Conotoxin LvIA from *Conus lividus* that Selectively Blocks 3 2 vs. 6/3 2 3 Nicotinic Acetylcholine Receptors [*Conus lividus*]	13/16 (81%)	13/16 (81%)	41.8	0.005

The top eight sequences are different than those in iteration 2, with the 7th and 8th replaced by two new records as shown in **Fig.3**.

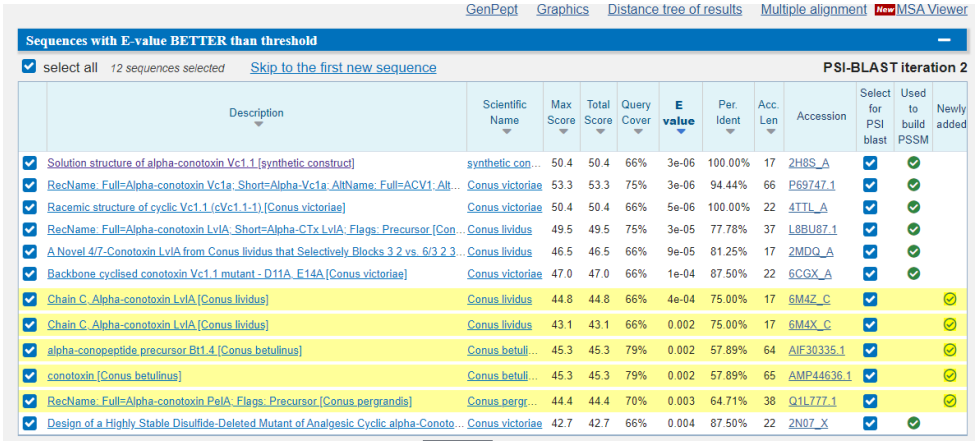

Fig.3 Screenshot of the second PSI-BLAST iteration

There is no new sequence record found at the 16th PSI-BLAST iteration. The top five sequence records are listed below in **Table 3**.

Table 3. The top five sequences at the end of PSI-BLAST iterations

Sequence ID	Name	Identity	Similarity	Score	E value
AFD18482.1	alpha-conotoxin, partial [*Conus lividus*]	12/24 (50%)	13/24 (54%)	65.2	2e-11
AFD18451.1	alpha-conotoxin, partial [*Conus diadema*]	12/23 (52%)	13/23 (56%)	63.9	6e-11
AFD18531.1	alpha-conotoxin, partial [*Conus quercinus*]	12/23 (52%)	14/23 (60%)	63.5	9e-11
AFD18539.1	alpha-conotoxin, partial [*Conus sanguinolentus*]	12/23 (52%)	13/23 (56%)	63.1	2e-10
AIC77102.1	conotoxin Lt1.1 precursor, partial [*Conus litteratus*]	12/23 (52%)	14/23 (60%)	62.7	4e-10

From PSI-BLAST results, we can conclude the polypeptide in Tube B belongs to alpha-conotoxin sourced from the genus *Conus*. The most matching species is *Conus lividus* with a relatively low E value 2e-11, though this conclusion is not confident since the identity is only 50%.

(ii) characteristics and function of alpha-conotoxin

Conotoxins are 17-residue small peptides that contain disulfide bonds as shown in **Fig.4**. As for the hallmark, alpha-conotoxins contain a conserved cysteine arrangement [CC-C-C] [12]. The disulfide bond targets G-protein coupled receptors (GPCRs). Alpha-conotoxin is a competitive-nicotinic-acetylcholine-receptor antagonist and can be used as a neurotoxin. Nicotinic acetylcholine receptors are found in muscle, nervous systems and other tissues. Thus, alpha-conotoxin can cause numbness on many tissues.

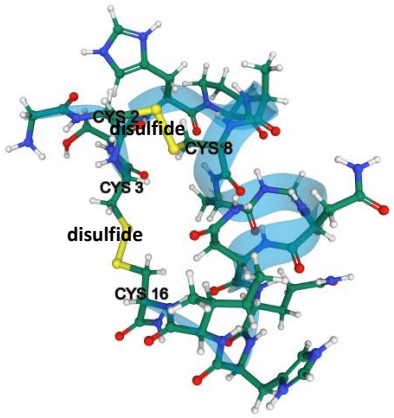

Fig.4 Structure overview of alpha-conotoxin

Discussion

The nature of tube A is a mixture of four proteins from viruses that are commonly used as biochemical weapons. ISCOMs are immune-stimulating complexes used as immune stimulators to produce a stronger immune response and hence are widely applied in vaccines [13]. Thus, a possible purpose for tube A content is to develop a comprehensive vaccine against the four viruses, namely, *Machupo mammarenavirus*, *Clostridium tetani, Ebola* virus, *Variola* virus.

Since these viruses are no longer threaten human being and not widely existing in population, one of the most potential intentions of Borduria's army is to develop this specific vaccine and inject to their soldiers and citizens. Hence, when Borduria applies these virus biochemical weapons to its army, their own soldiers and citizens can have antibodies against these weapons. However, the proteins in Tube A themselves might not harm the inhabitants in Khemed otherwise the existence of ISCOM cannot be reasonably explained. Thus, we cannot confidently confirm the harming intention.

As for the Tube B protein, it is alpha-conotoxin, a competitive-nicotinic-acetylcholine-receptor antagonist as a neurotoxin that caused the numbness and death of the spy. If this protein exposes, it can cause numbness and death of people and hence might also be used as a biochemical weapon. Besides, the Tube B alpha-conotoxin is probably a synthetic peptide according to the 100% identity matching record in the 1st PSI-BLAST iteration and the incomplete alignment to the natural alpha-conotoxin.

Conclusion

The contents from Tube A are Glycoprotein 1 [*Machupo mammarenavirus*], Tetanus toxin [*Clostridium tetani*], Matrix protein VP40 [*Ebola virus*], DNA topoisomerase 1 [*Variola virus*] and ISCOM, which can be used to develop a comprehensive vaccine. The Tube B peptide is an alpha-conotoxin that is dangerous for humans. Thus, we can conclude that the invading intention is possible.

Reference

1. Korf I, Yandell M, Bedell J: *Blast.* " O'Reilly Media, Inc."; 2003.
2. **Database resources of the National Center for Biotechnology Information.** *Nucleic Acids Res* 2018, **46:**D8-d13.
3. Pruitt KD, Tatusova T, Maglott DR: **NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins.** *Nucleic acids research* 2007, **35:**D61-D65.
4. Berman H, Henrick K, Nakamura H: **Announcing the worldwide Protein Data Bank.** *Nature Structural & Molecular Biology* 2003, **10:**980-980.
5. de Castro E, Sigrist CJ, Gattiker A, Bulliard V, Langendijk-Genevaux PS, Gasteiger E, Bairoch A, Hulo N: **ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins.** *Nucleic Acids Res* 2006, **34:**W362-365.
6. Harrison SC, Alberts B, Ehrenfeld E, Enquist L, Fineberg H, McKnight SL, Moss B, O'Donnell M, Ploegh H, Schmid SL: **Discovery of antivirals against smallpox.** *Proceedings of the National Academy of Sciences* 2004, **101:**11178-11192.
7. Da Fonseca F, Moss B: **Poxvirus DNA topoisomerase knockout mutant exhibits decreased infectivity associated with reduced early transcription.** *Proceedings of the National Academy of Sciences* 2003, **100:**11291-11296.
8. Anderson BG, Stivers JT: **Variola type IB DNA topoisomerase: DNA binding and supercoil unwinding using engineered DNA minicircles.** *Biochemistry* 2014, **53:**4302-4315.
9. Koster DA, Croquette V, Dekker C, Shuman S, Dekker NH: **Friction and torque govern the relaxation of DNA supercoils by eukaryotic topoisomerase IB.** *Nature* 2005, **434:**671-674.
10. Vologodskii AV, Cozzarelli NR: **Conformational and thermodynamic properties of supercoiled DNA.** *Annual review of biophysics and biomolecular structure* 1994, **23:**609-643.
11. Perry K, Hwang Y, Bushman FD, Van Duyne GD: **Structural Basis for Specificity in the Poxvirus Topoisomerase.** *Molecular Cell* 2006, **23:**343-354.
12. Johnson DS, Martinez J, Elgoyhen AB, Heinemann SF, McIntosh JM: **alpha-Conotoxin ImI exhibits subtype-specific nicotinic acetylcholine receptor blockade: preferential inhibition of homomeric alpha 7 and alpha 9 receptors.** *Molecular Pharmacology* 1995, **48:**194-199.
13. Sanders MT, Brown LE, Deliyannis G, Pearse MJ: **ISCOM-based vaccines: the second decade.** *Immunol Cell Biol* 2005, **83:**119-128.