Molecular modeling of NK-CT1, from Indian monocellate cobra (*Naja kaouthia*) and its docking interaction with human DNA topoisomerase II alpha

Pathikrit Bandopadhyay¹, ², Soma Halder¹, Mrinmoy Sarkar¹, Sujay Kumar Bhunia¹, Sananda Dey¹, Antony Gomes³, Biplab Giri*¹

¹Experimental Medicine and Stem Cell Research Laboratory, Department of Physiology, West Bengal State University, Barasat, Kolkata 700 126, West bengal, India; ²University of Kalyani, Kalyani 741 235, West Bengal, India; ³Department of Physiology, University of Calcutta, 92 APC Road, Kolkata 700 009, India; Biplab Giri – E mail: bgiri.emscrl@gmail.com; Phone: +91-9432904884; *Corresponding author

Received February 07, 2016; Revised April 7, 2016; Accepted April 17, 2016; Published June 15, 2016

Abstract:
A 6.76 kDa molecular weight cardio and cytotoxic protein of 60 amino acids in length called NK-CT1, was purified from the venom of Indian monocellate cobra (*Naja kaouthia*) by ion-exchange chromatography and HPLC as described in our earlier report. Therefore it is of interest to utilize the sequence of NK-CT1 for further functional inference using molecular modeling and docking. Thus homology model of NK-CT1 is described in this report. The anti-proliferative activity of the protein, binding with human DNA topoisomerase-II alpha was demonstrated using docking data with AUTODOCK and AUTODOCK MGL tools. Data shows that M26, V27 and S28 of NK-CT1 is in close contact with the nucleotides of the oligonucleotide, bound with topoisomerase-II alpha complex.

Keywords: NK-CT1, *Naja kaouthia*, human DNA topoisomerase II alpha, docking

Background:
Snake venom consists of a large number of biologically active peptides, proteins, non-proteins, macromolecules as well as several metallic elements. It is an abundant source of many toxins including cardiotoxins, neurotoxins, myotoxins, dendrotoxins, haemorrhagins, fibrinolytic enzymes and PLA2s as some major components. Cytotoxins or cardiotoxins are a group of polypeptides (60–70 amino acid residues) present in elapidae family of snake having diversified pharmacological actions such as haemolysis, depolarization of muscles [1], muscle fusion, selective killing of certain type of tumor cells, inhibition of protein kinase C and muscle contraction [2]. Cytotoxins have the ability to damage a wide variety of cells including the cancerous types by targeting lysosome [3]. Calmette et al. (1933) had first reported the anticancer property of snake venom [4]. Since then a number of research studies have been carried out regarding the anticancer activities of snake venoms. Earlier we reported the cardiotoxic and cytotoxic effects of the crude venom of Indian monocellate cobra, *Naja kaouthia* [5]. Later, it was demonstrated that the cardiotoxic and cytotoxic activities were due to the presence of the protein NK-CT1 in the venom [6]. Research study by Bhowmik et al. (2013) [7] had shown that PEGylated gold nanoparticle conjugated with NKCT1, GNP-NKCT1, possesses significant and selective anticancer activity, likely by inducing programmed cell death through mitochondrial and lysosomal pathways.

Topoisomerase II alpha plays a key role in DNA replication and hence, is a target for several chemotherapeutic agents. Its primary function is to alter DNA topology from its super-coiled form to a more exposed (partially uncoiled) form by inducing single strand DNA breaks and simultaneously passing another intact double helix through the gap. Etoposide is an important
A chemotherapeutic agent that is used to treat wide spectrum of human cancers. It has been used clinically for more than two decades and remains one of the most highly prescribed anticancer drugs in the world. The primary target for etoposide is topoisomerase II alpha [8]. In breast cancer, topoisomerase II alpha expression has been linked to cell proliferation and HER2/neu protein over-expression [9]. Therefore, we are interested to investigate the interaction between NK-CT1 and human topoisomerase II alpha to understand whether the enzyme is linked with the anti-proliferative activity of the toxin. Moreover, determination of X-ray crystallographic structure as well as functional characterization of biological active site of NK-CT1 has not been reported till date. Therefore, in the present study our attempt is to derive the 3-dimensional (3-D) structure of this protein from its primary amino acid sequence and its docking interactions with human topoisomerase II alpha for the identification of the specific amino acid residues, which are responsible for its functions.

Methodology:
Template Selection
The primary amino acid sequence (60 amino acids) of NK-CT1, as already reported in our earlier study [6], was used for protein blast search (BLASTP 2.2.29) [10] as query against PDB database using BLOSUM80 matrix having gap costs as existence = 10 and extension=1 and using compositional score matrix alignment [11] to scale all substitution scores by an analytically determined constant, while leaving the gap scores fixed. This procedure was adopted to get more accurate E-value rather than standard un-scaled score. Based on the obtained results, the templates were selected for homology modeling.
considering the lowest E-value with the query sequence and >30% identity. Uniprotd Blast was performed with the primary amino acid sequence of NK-CT1 as query, using BLOSUM62 matrix, having threshold value of 10 and allowing gap and filtering the low complexity region to find out the family of NK-CT1.

Conserved Domain Analysis by CDD
Sequence of NK-CT1 was analysed in CDD database (Conserved Domain Database) of NCBI to find out the span and composition of conserved protein sequence region present in it. CDD database is a resource for the annotation of protein sequences with the location of footprints of conserved domain, and functional sites which are inferred from these footprints. Proteins similar to the query are grouped and scored by architecture [12]. The CD-Search service uses RPS-BLAST, which stands for "Reverse Position-Specific BLAST" that uses the query sequence to search a database of pre-calculated ‘Position Specific Scoring Matrices’ (PSSMs), and report significant hits in a single pass.

Homology Modeling using SWISS MODEL
The protein structure of NK-CT1 was modeled using SWISS MODEL server [13, 14, 15]. The resultant model was selected based on the lowest QMEAN 4 score. The selected model was checked for its stereo-chemical quality using QMEAN4&6 and was assessed by different tools present in SWISS MODEL server. At the initial run the database returns 64 template hits. Template selection was done by Blast with lowest E-Value cut off $2e^{-10}$ to up to 4 $e^{-34}$ [10, 11] and HHBlits [16] was performed against the SWISS-MODEL template library. A total of 54 templates were found with there corresponding QMEAN4 score [17, 18] which is a Z- score, composite score for prediction of model quality based on energy profile, C-beta, torsion, solvation and all atom. Models were built based on the target-template alignment using Promod-II [19]. Finally, using a force field regularized the geometry of the resulting model. After completion of the modeling the database returns the modeled structures [20] with there corresponding QMEAN4 score [17, 18]. Prediction of the model quality using QMEAN4 score showed that the predicted model falls within the most favorable range for composite Z-score of a model (QMEAN4 score) that lies between -1 to +1 and has the QMEAN4 score 0.00 in comparison to the QMEAN4 scores of other models that were generated (based on other template) by the SWISS MODEL server (Data not shown).

ASA versus residue Number plot
Accessible surface area or solvent accessibility of amino acids in a protein helps for localization of active sites. A characteristic two-dimensional spiral plot of solvent accessibility provides a convenient graphical view of residues in terms of their exposed surface areas. In addition sequential plots of bar charts are also provided by the tool for each amino acid residues with the color coding corresponding to there location i.e. either in the surface or in the core. ASA plot of NK-CT1 was done by ASA-View a database and tools for the solvent accessibility representation in proteins [21].

Phyre² Analysis for Prediction of Mutational Sensitivity, Prediction of Receptor Binding Site
Phyre² (Protein Homology/analogY Recognition Engine V 2.0) [22] was used for the prediction of receptor binding site, mutational sensitivity of the protein NK-CT1. Phyre² is a web based service that is used for protein structure prediction with template identity score and confidence score indicating the quality of the model.
Figure 3: Phyre Investigator analysis for prediction of mutational sensitivity, receptor binding site, rotamers and clashes of the modeled protein NK-CT1, residues are color coded in the model according to that of specific features. (A) Result obtained for mutational sensitivity of each amino acid of NK-CT1. Red color indicates that substitution of this amino acid likely to affect function, blue indicates very low sensitivity, it is to be noted that in NK-CT1, C at position 3, 14, 21, 38, 42, 53 and 54, L at 20, P at 33, R 36 and Y 51 are marked as red indicates that mutation in this amino acid will affect function of this protein. (B) Prediction of Receptor binding site of NK-CT1 by Phyre Investigator residues are color coded as red indicates the receptor binding site of NK-CT1. It is to be noted that C3, F10, 25 and M 26, T31, V32, P33, V34, K35 are involved in receptor binding. (C) Electrostatic surface distribution of the modeled surface (by UCSF Chimera) the surface were color coded as per the standard protocol of UCSF Chimera, each amino acids were marked with standard code (red for negative potential, to white near neutral, to blue for positive potential).

Figure 4: Docked structure of NK-CT1 with DNA Topoisomerase II alpha. Figure was generated in the GUI of AUTODOCK MGL Tools. Docked structure showing the protein NK-CT1 (in ball and stick model) with DNA Topoisomerase II alpha (marked as blue arrow) complexes with oligo-nucleotide (PDB ID: 4FM9) shown in Panel A (marked as white circle). Interaction of NK-CT1 (Panel B marked with white circle) with the DNA Topoisomerase II alpha showed that some specific residues are interacting with the oligo-nucleotide bound to the protein and therefore can be predicted that NK-CT1 interacts with the DNA binding domain of Topoisomerase II alpha.

Electrostatic distribution of the modeled surface
The electrostatic distribution of the modeled 3D structure of NK-CT1 was analysed by UCSF Chimera [23], a highly extensible programme for analysis of molecular structure. It uses C++ code for color calculations. Electrostatic surface mapping of NK-CT1 was performed for analysis of charge distribution and charge related properties of molecules and the surface of NK-CT1 was color coded as per the Coulomb’s law.
Motif Prediction by ProFunc

ProFunc server [24] was used for the prediction of motif for the protein NK-CT1 and to know which residues are involved in the motif formation. PROMOTIF [25] analysis was performed (a tool nested under ProFunc database) for the identification of biochemical functions of NK-CT1 from its three dimensional structure using both sequence and structure based methods for prediction of motifs present in NK-CT1, which includes supersecondary structures i.e. beta turns, 3_{10} helices, beta hairpins.

Table 2: Intermolecular interaction of docked complexes of NK-CT1 with DNA Topoisomerase II alpha, DG, DT, DA= Deoxy Guanine, Thymine and Adenine respectively of the polynucleotide complexed with DNA Topoisomerase II alpha.

| No | Bond   | Ligand Residue | Receptor Residue | Interaction Constituents |
|----|--------|----------------|------------------|-------------------------|
| 1  | Hydrogen | Lys 12 | Asp 1013 | LYS12-NZ-ASP1013:OD2 |
| 2  | Close contact | Val 27 | DA2 | VAL27-DA2 |
| 3  | Close contact | Met 26 | DC11 | MET26-DC11 |
| 4  | Close contact | Lys 35 | Lys 1011 | LYS35-LYS1011 |
| 5  | Close contact | Phe 10 | Glu 663 | PHE10-GLU663 |
| 6  | Close contact | Pro 8 | Thr 666 | PRO8-THR666 |
| 7  | Close contact | Thr 31 | Lys 656 | THR31-LYS656 |
| 8  | Close contact | Val 7 | Lys 1011 | VAL7-LYS1011 |
| 9  | Close contact | Ser 28 | C:DG4, D: DT13, C: DA5, SER28-DG4, SER28-DT13, SER28-DA5. | |
| 10 | Close contact | Met 24 | Asn 998 | MET24-ASN998 |

Docking of NK-CT1 by AUTODOCK

NK-CT1 was docked with human topoisomerase-II alpha by AUTODOCK 4.2.6 [26] using Lamarckian genetic algorithm [27] implemented in the programme suite to identify appropriate binding modes and conformations of the ligand molecules. In this docking study NK-CT1 molecule was treated as ligand and topoisomerase molecule as receptor. Gasteiger charges were added and the rotatable bonds were set by the AutoDock tools. Polar hydrogen atoms were added and Kollaman charges were assigned to the protein using AutoDock tools (ADT). The grid map was constructed using Autogrid (a package nested in AUTODOCK) centered on the macromolecule with a gridbox size 90X90X90 Å³ with grid spacing 0.398 Å. The Lamarckian genetic algorithm, the pseudo-Solis and wets methods were applied for minimization using default parameters. Binding energy, torsional energy, intermolecular energy, number of H-bonds and RMS value were recorded in each ligand bound conformations. The docked conformations that were generated after docking were selected based on their binding energy and corresponding figure was developed from Graphics User Interface of AUTODOCK MGL Tools.

Results & Discussion:

NK-CT1 protein was isolated from Naja kaouthia venom which belongs to Elapidae family of snakes and has cardiotoxic and cytotoxic activity which was reported earlier by our research group [6]. It was observed that crude venom from Indian Monocellate cobra has prominent anticancer effect on carcinoma, sarcoma and leukemia both in vitro and in vivo models [5]. The protein (6.7Kda) was isolated by HPLC from snake venom and the amino acid sequence was determined as: VCPKNSLVLKYVCRCN. But no such research studies have done the homology modelling and subsequent functional characterization and prediction of structure of bungarotoxin from B. fasciatus snake. Ruggieri et al. (2011) [29] had done the homology modelling and subsequent functional characterization and structure/function comparison of phytoxic protein PcF from Plantophthora cactorum. Therefore in this study, an attempt has been made for homology modelling of NK-CT1 from its primary amino acid sequence and also to identify the residues that are responsible for maintaining the different biological activities of this protein and also to unveil its antiproliferative activity whether linked with human topoisomerase II alpha by docking interaction.

The conserved domain of NK-CT1 was identified using CDD of NCBI and it was observed that the domain of this protein have
specific features i.e. receptor binding site comprise of some specific conserved residues i.e. from 3rd residue to 10th residue, 25th, 26th residues and 31st to 35th residues which are involved in receptor binding that was mapped from the conserved domain annotations to the query sequence (NK-CT1). The conserved domain model of NK-CT1 was based on multiple sequence alignments of related proteins spanning a variety of organisms to reveal sequence regions containing the same, or similar, patterns of amino acids. It showed the similarity of its conserved residue with fasciculin (a powerful inhibitor of acetylcholinesterase), bungarotxin and other three finger toxins (Figure 1; Panel A). It was also observed that the receptor binding is mainly achieved through 2 fingers which are present in the toxin (Figure 1; Panel-B). Uniprot Blast analysis had revealed that NK-CT1 belongs to snake three finger toxin families, Type IA cytototoxin subfamily. Mutational sensitivity analysis by Phyre2 investigator had shown that C at position 3, 14, 21, 38, 42, 53 and 54, L 20, P 33, R 36 and Y 51 are very much susceptible to mutation (colour coded as red, Figure 3; Panel A). There are 4 disulphide bonds present in NK-CT1 and mostly all are located inside the core of the protein as observed from the ASA plot (Figure 2; Panel-B), so mutation of cysteine leads to loss of its structural stability. We also observed that C 3, F 10, 25 and 26th M and 31st residue (T) to 35th residue (K) constitute the receptor binding site (Figure 3; Panel B). It was interesting to observe that in the receptor binding site K 35, R 36 are present of which R 36 is susceptible to mutation and the residues comprising receptor binding site are the part of the conserved domain (Refer above to conserved domain analysis by CDD). Surface analysis of the NK-CT1 (Figure 2; Panel B) by ASA-View showed that 31st T, 35th K and 36th R are located outside suggesting the fact that these residues may be involved in receptor binding which supports our previous analysis of receptor binding site in Phyre2. Electrostatic distribution by UCSF Chimera [23] showed that NK-CT1 has more positive charge residues (indicated as blue, Figure 3; Panel C) on the outer surface suggesting the fact that the residues are the part of the conserved domain which have specific functionalities i.e. receptor binding. Type II topoisomerase is required for chromosome segregation by introducing negative supercoiling of DNA by double strand break in DNA and therefore serves as a prime targets for different cancers [8]. In breast cancer, topoisomerase II alpha expression has been linked to cell proliferation and HER2/neu protein overexpression [9]. NK-CT1 molecule was docked with human topoisomerase-II alpha (bound to oligonucleotide) and ten possible interaction sites were found where NK-CT1 can bind with this protein. Out of the ten possible interaction sites, the specific site was selected having the least binding energy (Table 1). The total intermolecular energy of the docked complex was: -13.78Kcal/mol and lowest RMSD value from the reference structure is 6.7Å in compare to the other conformations (Table 1) and it was observed that NK-CT1 interacts with the core of DNA binding sites of topoisomerase - II alpha. The interactions analysis of AutoDock Suite reveals that this protein interacts with the oligonucleotide molecule bound with the topoisomerase II alpha. It was observed that V 27, M 26, K35, F 10, P 8, T 31, V 7 and S 28 are in close contact with the receptors and also M 26, V 27 and S 28 are in close contact with the nitrogenous base of the oligonucleotide (Table 2) which is complexed with DNA topoisomerase-II alpha (Figure 4). It supports our hypothesis that NK-CT1 shows its antiproliferative action by interacting with the core DNA binding region of human DNA topoisomerase-II alpha and correlates with our present analysis as all these residues (except 28th S) are the components of conserved domain which is involved in receptor binding.

The 3D structure of NK-CT1 was derived by in Silico method using SWISS Model server [13-15] which generates a composite Z-score i.e. Qmean4 score towards quality of the structure prediction using the template (chain B of ITGX, gamma cardio-toxin isolated from Taiwan cobra venom). The modeled structure has the highest positive QMEAN4 score amongst all other structures that were generated as a resultant of SWISS Model server. To further investigate we performed QMEAN6 scoring function [18] which is a linear combination of six structural descriptors using statistical potentials: The local geometry was analysed by a torsion angle potential over three consecutive amino acids. Predicted structure was checked by ProCheck which showed maximum atoms (86.5%) lie within the core and allowed region of the Ramachandran Plot. It was observed that the modeled structure of NK-CT1 has QMEAN6 score of 0.595 which falls within the favourable range when it was compared with the structures of the PDB database (data not shown).

The modeled structure (Figure 2; Panel-A) was superimposed with the template 1tgx1.B, it was observed that RMSD value on superposition of the modeled structure of NK-CT1 with the template structure was 0.89 and all the atoms were found to lie within <5.0Å, the TM score [30] for the superposition was 0.90 which means that the modeled structure of NK-CT1 and template structure was of same fold (Table 1). The secondary structure of the protein was predicted along with motif analysis by PROMOTIF v3.0 [25] which revealed that NK-CT1 has 2 beta sheets, 2 beta hairpins, 1 beta bulge, 5 strands, 1 helix, 5 beta turns and 4 disulphide bonds. K (2nd), 35th and 49th residue), N (4th residue), Y (11th residue), T (13th residue), L (20th and 49th residue), M (24th and 26th residue), I (39th) residue) were involved in motif formation. From PROMOTIF analysis it was observed that C 3 is involved in disulphide bond, V 32 and F 10 are involved in beta turn, M 26 is involved in beta hairpin formation, recalling motif is the functional part of the proteins; it can be predicted that these beta turn, beta hairpin and strand of sheet B are involved in receptor binding (data not shown). So it can be predicted that this residues, which constitute a motif, achieve receptor binding and anti-proliferative activity of NK-CT1.

Conclusion: The structure of NK-CT1 from the venom of Indian cobra is unknown. Hence we describe a homology model of the protein using its structural features and electrostatic properties. We also describe its binding with the human DNA topoisomerase – II alpha using docking illustrating its potential anti proliferative activity.

Acknowledgement: The authors acknowledge West Bengal State University for providing laboratory infrastructure.
References:
[1] Dufton MJ & Hider RC, Pharmacol Ther. 1988 36: 1 [PMID: 3277206]
[2] Kumar TK et al. Adv Exp Med Biol. 1996 391: 115 [PMID: 8726052]
[3] Feofanov et al. Biochem J 2005 390: 1 [PMID: 15847607]
[4] Calmette A et al. C R Acad Sci 1933 197: 205
[5] Debnath A et al. Ethnopharmacol 2007 111: 681 [PMID: 17258913]
[6] Debnath A et al. Toxicon. 2010 56: 569 [PMID: 20595038]
[7] Bhowmik T et al. Cancer Nanotechnol. 2013 4: 39 [PMID: 26069500]
[8] Baldwin EL et al. Curr Med Chem Anticancer Agent. 2005 5: 363 [PMID: 16101488]
[9] Depowski PL et al. Mod Pathol. 2000 13: 542 [PMID: 10824926]
[10] Altschul FS et al. Nucleic Acids Res 1997 25: 3389 [PMID: 9254694]
[11] Altschul FS et al. FEBS J 2005 272: 5101 [PMID: 16218944]
[12] Marchler-Bauer A et al. Nucleic Acids Res 2013 41: D348 [PMID: 23197659]
[13] Arnold K et al. Bioinformatics 2006 22: 195 [PMID: 16301204]
[14] Biasini M et al. Nucleic Acids Res 2014 42: W252 [PMID: 24782522]

[15] Bordoli L et al. Nature Prot 2009 4: 1 [PMID: 19131951]
[16] Remmert M et al. Nat Methods 2012 9: 173 [PMID: 22198341]
[17] Benkert P et al. Nucleic Acids Res 2009 37: W510 [PMID: 19429685]
[18] Benkert P et al. Bioinformatics 2011 27: 343 [PMID: 21134891]
[19] Guex N et al. Electrophoresis 1997 18: 2714 [PMID: 9504803]
[20] Sali A et al. J Mol Biol 1993 234: 779 [PMID: 8254673]
[21] Ahmad S et al. BMC Bioinformatics 2004 5: 51 [PMID: PMC420234]
[22] Kelley AL et al. Nature Prot 2009 4: 363 [PMID: 19247286]
[23] Pettersen EF et al. J Comput Chem 2004 25: 1605 [PMID: 15264254]
[24] Laskowski AR et al. J Mol Biol 2005 351: 614 [PMID: 16019027]
[25] Hutchinson GE et al. Protein Sci 1996 5: 212 [PMID: 8745398]
[26] Morris MG et al. J Comput Chem 2009 30: 2785 [PMID: 19399730]
[27] Morris MG et al. J Comput Chem 1998 19: 1639
[28] Roly ZY et al. Bioinformatics 2014 10: 617 [PMID: 25489170]
[29] Roly ZY et al. Protein Sci 2014 10: 617 [PMID: 25489170]
[30] Zhang Y et al. Proteins 2004 57: 702 [PMID: 15476259]

License statement: This is an Open Access article which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. This is distributed under the terms of the Creative Commons Attribution License

Edited by P Kanguuane

Citation: Bandopadhyay et al. Bioinformation 12(3): 105-111 (2016)

ISSN 0973-2063 (online) 0973-8894 (print)