Chapter 5

Purification of Nk-3FTx, a three finger toxin from Naja kaouthia venom

5.1 Introduction

In chapter 4, the partial composition of *N. kaouthia* venom has been revealed using proteomic approach. Based on this, it was found that 3FTxs (3FTx) is the major venom protein family of northeast *N. kaouthia* constituting 69.9% of the total protein. Functionally they are known for neurotoxicity, cytotoxicity, channel blocking effect, anticoagulation, analgesic property and others [150, 162, 203, 297–299]. Structurally they are classified into short chain, long chain, and non-conventional 3FTx based on the presence or absence of the fifth disulphide bridge [85, 86]. Literature survey shows that *N. kaouthia* venom from northeast India is unexplored venom and till date not a single toxin from this venom has been purified and characterized.

In this chapter we describe the purification and characterization of Nk-3FTx, a 3FTx from the venom of *N. kaouthia* from northeast India. The primary structure determination reveals that the purified protein is non-conventional 3FTx.

5.2 Results

5.2.1 Fractionation of crude venom

For isolation of Nk-3FTx, the crude *N. kaouthia* venom was fractionated as described in Section 4.2.1. Briefly, crude venom (2mg) was loaded into C18 RP HPLC column and fractionated on HPLC system at a flow rate of 1ml/min. Fractionation was carried out using buffer B (0.1% TFA in 80% ACN) which separated into 22 protein peaks and numbered (1 to 22). Based on the ESI-MS profile of the crude venom (Table 4.2), peak 10 was selected for purification of the Nk-3FTx (Figure 5.1). Peak 10 was subjected to SDS-PAGE (non-reduced) to check its homogeneity. Under non-reduced condition it showed a band approximately at 10kDa (Figure 5.2). The appearance of thick band could be presence of isoforms, hence peak 10 was further fractionated on RP-HPLC for purification of Nk-3FTx.

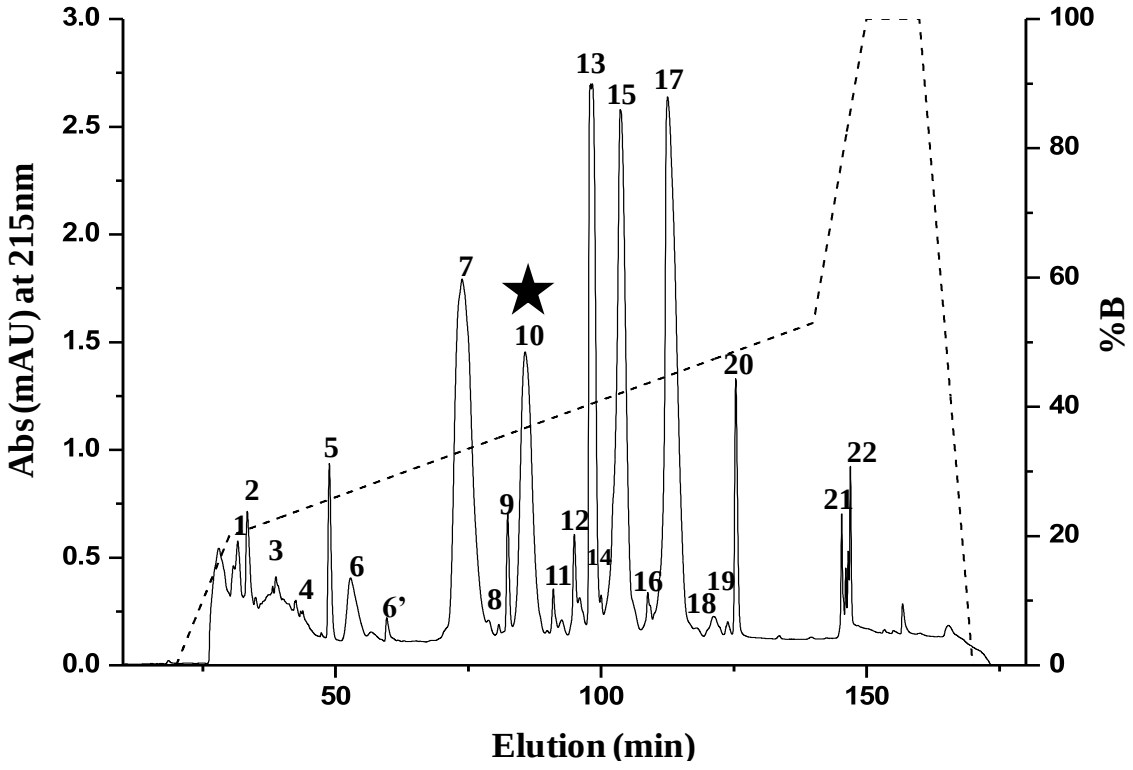

Figure 5.1: Fractionation of crude *N. kaouthia* venom for isolation of 3FTx. The peak indicated with star was selected for purification of 3FTx.

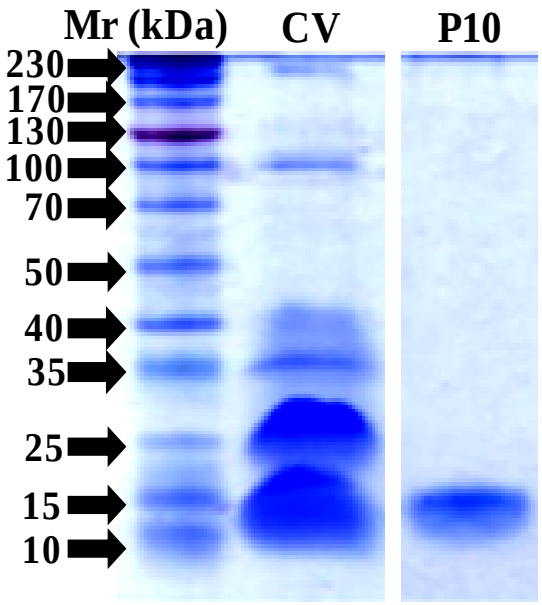

Figure 5.2: SDS-PAGE of protein peak 10 (non-reduced). CV- crude *N. kaouthia* venom (10µg). P10- Peak 10 (10µg) . Mr. - standard molecular weight protein marker (kDa)

5.2.2 Re-chromatography of peak 10

Peak 10 (100µg) was fractionated on an Aeris WIDEPORE (XB-C18, 3.6µ, 150x 2.10 mm, 200Å) column using buffer B (0.1% TFA in 80% ACN) at a flow rate was 0.3ml/min. Various gradient of buffer B (80% acetonitrile) was used to fractionate peak 10 for purification of 3FTx. Fractionation was carried with buffer B in a gradient from 38% to 43%, over 70mins which revealed the presence one major peak and several minor peaks (Figure 5.3A). The major peak was also found to have a shoulder peak hence gradient was further modified. The percentage of buffer B was changed to 40-45%, over 45mins which resulted into one major peak and two minor peaks (Figure 5.3B). Further when the buffer B gradient was changed to 40-42% over 45mins a major sharp symmetrical peak with two minor peaks was obtained (Figure 5.3C). This optimized condition was used for purification for further purification of Nk-3FTx. This purified protein was estimated, lyophilized and used for the studies.

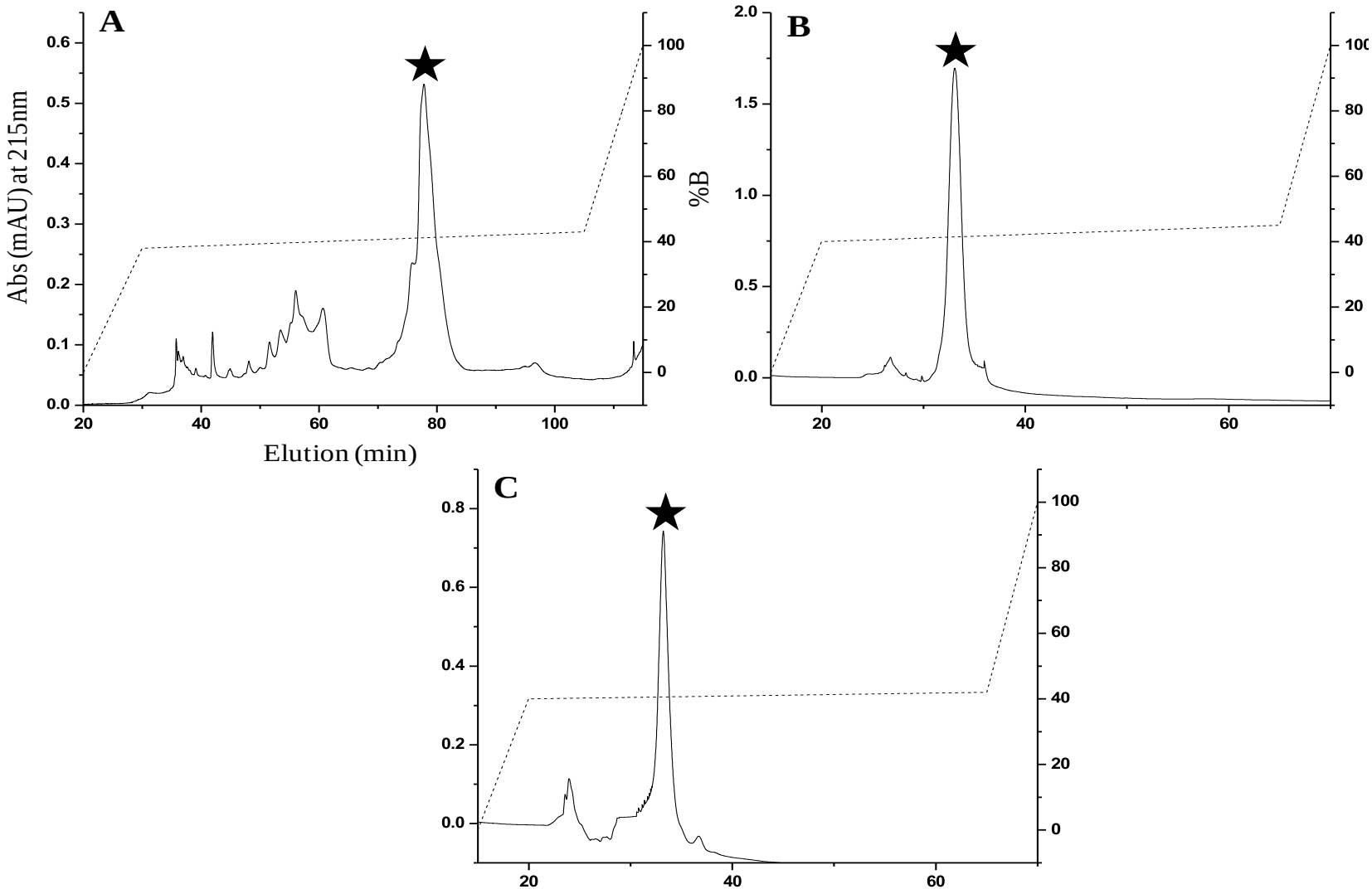

Figure 5.3: Rechromatography of peak 10. 100µg of peak 10 protein was loaded for rechromatography. **A.** The protein was rechromatographed at a gradient of 38-43% buffer B. **B.** Rechromatography of peak 10 at a gradient of 40-45% buffer B. **C.** Rechromatography of peak 10 at a gradient of 40-42% buffer. Flow rate for elution was maintained at 0.3ml/min. The starred symmetrical peak was considered as purified protein and was further characterized.

5.2.3 Molecular weight determination of Nk-3FTx

Molecular mass of the purified peak 10 was determined by electrospray ionization mass spectrometry (ESI/MS). Purified protein (0.01mg) was directly injected into the ESI-MS. The ESI-MS profile was analyzed using Promas for Xcaliber to and decipher the mass from the raw mass data. The spectrum shows a series of multiple charged ions, corresponding to a single, homogenous protein with a molecular mass of 7579.5 ± 0.591 Da (Figure 5.4A&B). This mass was found to be in the range of 3FTx family.

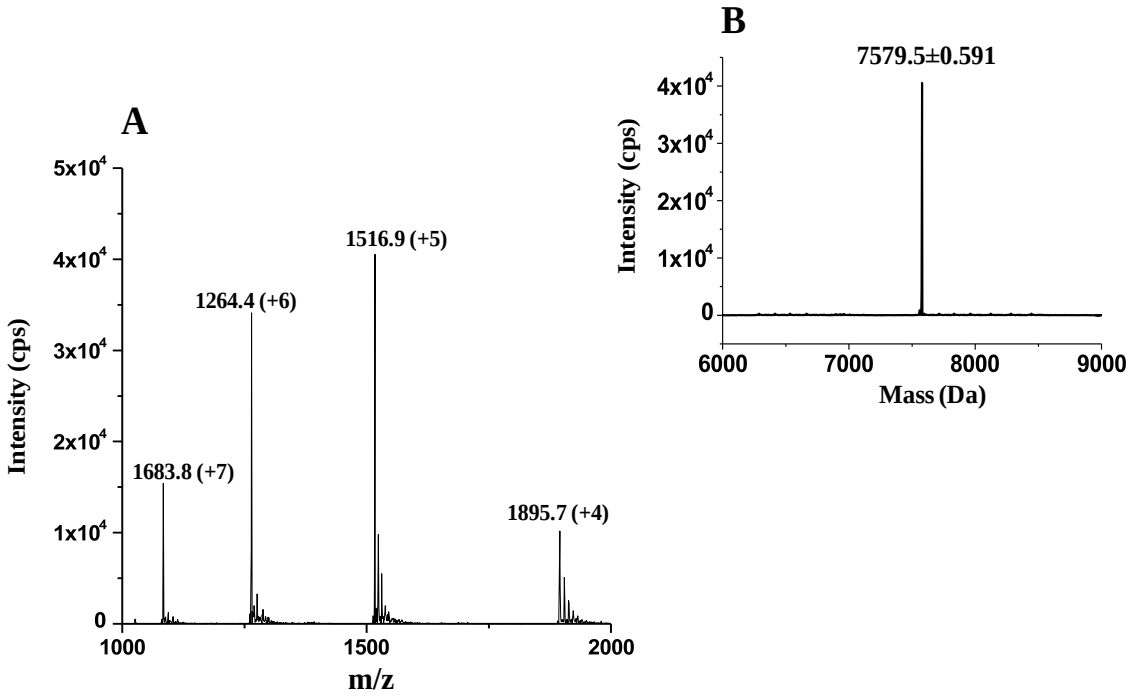

Figure 5.4: Molecular mass determination of purified peak 10. A: The spectrum shows a series of multiply charged ions, corresponding to a single, homogenous peptide with a molecular weight of 7579.5 ±0.591 Da. **B:** Reconstructed mass spectrum of the protein, CPS = counts/s; Da = dalton. The protein was named as Nk-3FTx.

5.2.4 N-terminal sequencing of Nk-3FTx

N-terminal sequencing of Nk-3FTx (Edman degradation) identified the first 33 amino acid residues. The residues obtained were: LTxLNxPEMFxGKFQIxRNGEKIxFKKLHQRRP. The cysteine residues were not modified and could not be detected in the chromatogram hence "x" represents cysteine residues. The first residue was Leucine and the cysteine residues are found at 3rd, 6th,

11th, 17th and at 24th positions in the 33 amino acid N-terminal sequence of Nk-3FTx. BLAST analysis of the 33 residues from N-terminal of Nk-3FTx shows that it is identical to weak toxin CM-9a, a 3FTx reported from the *N. kaouthia* (Monocled cobra) [191] (Figure 5.5)

```
This study      LTXLNXPEMFXGKFQIXRNGEKIXFKKLHQRRP--------------------------------
P25679          LTCLNCPEMFCGKFQICRNGEKICFKKLHQRRPLS-RYIRGCADTCPVGYPKEMIECCSTDKCNR
P60814          LTCLNCPEMFCGKFQICRNGEKICFKKLHQRRPFSLRYIRGCAATCPGTKPRDMVECCSTDRCNR
Q9YGI4          LTCLNCPEMFCGKFQICRNGEKICFKKLHQRRPFSLRYIRGCAATCPETKPRDMVECCSTDRCNR
P82935          LTCLNCPEMFCGKFQICRNGEKICFKKLHQRRPLSWRYIRGCADTCPVGKPYEMIECCSTDKCNR
O42255          LTCLNCPEMFCGKFQTCRNGEKICFKKLQQRRPFSLRYIRGCAATCPGTKPRDMVECCSTDRCNR
O42256          LTCLNCPEMFCGKFQTCRDGEKICFKKLQQRRPFSLRYIRGCAATCPGTKPRDMVECCSTDRCNR
```

Figure 5.5: Homology search of N-terminal sequence of Nk-3FTx using BLASTP. "x" represents the cysteine residues in the sequence.

5.2.5 ESI-LC MS/MS of Nk-3FTx

To determine the complete sequence of Nk-3FTx, the peptides were generated by trypsinization and analyzed by ESI LC-MS/MS. Sequence of seven peptides were obtained using Proteome Discoverer 3.1. The MS/MS sequence, similarity to the protein in that data base, protein group accession numbers, modification, charge, molecular mass and delta mass (in ppm) are shown in table 5.1. The carbamidomethyl C and Oxidized M are shown in lowercase letters. The MS/MS sequences and the N-terminal sequence were assembled to determine the overlapping sequences. Based on the overlapping and comparison with *Naja naja kaouthia* CM-9a (P25679.2) sequence, the sequence of Nk-3FTx was obtained (Figure 5.6). We identified 55 amino acid residues of Nk-3FTx by a combination of N-terminal sequencing and ESI LC-MS/MS. However, sequence of three fragments could not be obtained as they were short peptides containing either R or K at the C-terminal end.

Table 5.1: Tryptic digest peptides sequences of Nk-3FTx obtained by LC-MS/MS

Sl no	MS/MS sequences	Protein group accessions	Modifications	Charge	MH+ [Da]	ΔM [ppm]
1	LTcLNCPE MFcGK	14195693	C3(Carbamidomethyl) C11(Carbamidomethyl)	2	1573.01	219.7
3	LTcLNcPE MFcGKFQI cR	14195693	C3(Carbamidomethyl) (Carbamidomethyl); M9 C11(Carbamidomethyl)	2	1646.23	328.3
4	LTcLNcPE MFcGK	14195693	C3(Carbamidomethyl) C6(Carbamidomethyl) C11(Carbamidomethyl) C17(Carbamidomethyl)	2	2334.52	210.3
5	FQIcR	14195693	C3(Carbamidomethyl) C6(Carbamidomethyl) C11(Carbamidomethyl)	2	1629.67	-13.05
6	LTcLNCPE MFCGK	14195693	C4(Carbamidomethyl)	1	723.48	173.5
7	IcFKK	14195693	C3(Carbamidomethyl)	2	1515.39	-171.1
8	LTcLNcPEm FCGK	14195693	C2(Carbamidomethyl)	1	695.36	-43.1
9	IcFK	14195693	C3(Carbamidomethyl) C6(Carbamidomethyl) M9(Oxidation)	3	1588.06	-377.6
10	RPSLR	14195693	C2(Carbamidomethyl)	1	567.36	107.8
11	GcAAcPK	290560289	M2(Oxidation) C12(Carbamidomethyl)	1	14.25	0
12	DMVEccST DR	290560289	C2(Carbamidomethyl) C6(Carbamidomethyl)	1	750	60.4
			C5(Carbamidomethyl) C6(Carbamidomethyl)	2	1272.98	404.9

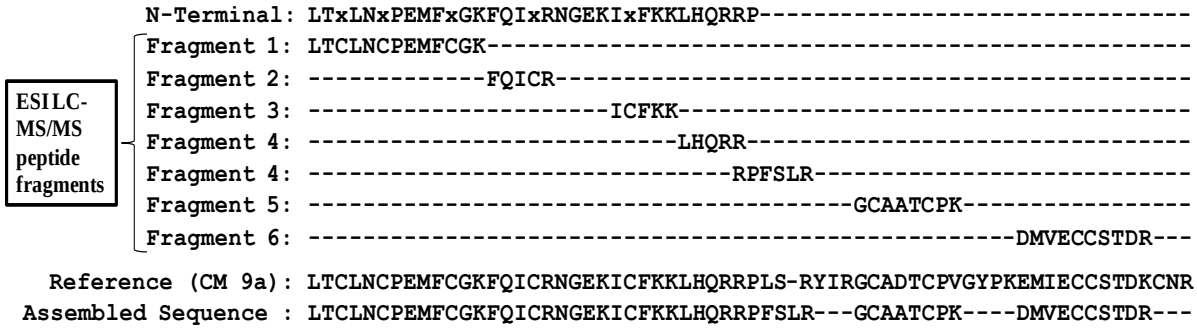

Figure 5.6: Assembly of amino acid sequence of Nk-3FTx: Sequences were obtained from N- terminal sequencing and ESI LC-MS/MS of tryptic digested fragments of Nk-3FTx. The Cys residues are represented by "x" in N-terminal sequence. A non-conventional 3FTx, CM 9a (P25679.2) from *Naja naja kaouthia* is used as a reference to determine the overlapping sequences.

5.2.6 Sequence analysis

The assembled sequence of Nk-3FTx was subjected to BLAST (BLASTP 2.2.32) using NCBI database (National Centre for Biotechnology Information) search to confirm the identity of Nk-3FTx (Figure 5.8). The purified protein was found to belong to snake toxin superfamily. The BLAST search confirms Nk-3FTx to contain conserved snake toxin domain. Nk-3FTx was found to be similar with 11 snake toxins with a similarity score of 80-200, 28 sequences with a similarity score of 50-80 and 17 sequences with 40-50 similarity score.

Nk-3FTx was found to be homologous with weak neurotoxin NNAM2 (Q9YGI4) from *Naja atra* and one unnamed protein (CAA04578.1) from *N. kaouthia* with E-score value ($5E^{-27}$) and similarity score of 80-200. The identity of both the sequences to Nk-3FTx was 87% and total score was found to be 103. Also, there was a substitution of amino acid residue where Glu44 in both NNAM2 and unnamed toxin was replaced by Lys44 in Nk-3FTx (Figure 5.8). Another toxin reported from from *Naja atra* also showed 87% identity to Nk-3FTx, however, the E-score value was found to be $7E^{-27}$.

The query sequence was found to hit putative domains present in short and long neurotoxins, cytotoxins and short toxins and also some other miscellaneous peptides. The domain was reported to contain 60-75 amino acids which are fixed by 4-5 disulfide

bridges (Figure 5.7). It contains mostly beta sheets and may exist either as monomers or dimers. The conserved domain hit length was 64 amino acids with an E-score value of $4.11e^{-11}$ (Figure 5.7).

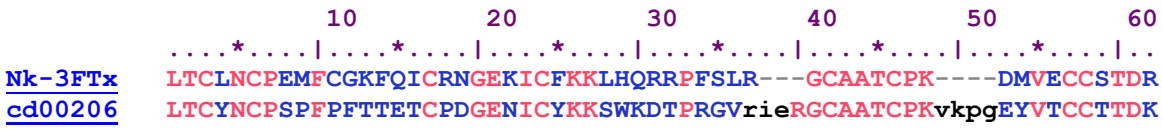

Figure 5.7: Domain hit of Nk-3FTx. Snake toxin (cd00206). Conserved domain length of cd00206- 64 amino acids, Bit score 52.41, E-score value- $4.11e^{-11}$. The upper case amino acids are aligned to generate the PSSM (position specific scoring matrix). The amino acid highlighted in red are the most frequently observed residues. The red and blue color represents degree of conservation, the red color depicts highly conserved.

Nk-3FTx was found to contain 8 conserved cysteine residues (4 disulfide bonds) with additional 2 residues which might be forming the fifth disulfide bond in its first loop (Figure 5.8). Hence, considering the presence of fifth disulfide bridge, Nk-3FTx is categorized as non-conventional 3FTx.

The assembled sequence of Nk-3FTx aligned with other reported 3FTx from *N. kaouthia* venom from different geographical locations is shown in Figure 5.9. The sequence, which showed similarities with Nk-3FTx are all reported from *N. kaouthia* from Thailand (Figure 5.9). The sequence showed highest similarity with CM-9a, a non conventional 3FTx (P25679.2) of *N. kaouthia* from Thailand origin [191]. Although, CM-9a (7438kDa) and Nk-3FTx (7579.5kDa) are comparatively similar in their molecular masses, the percentage of similarity between the two sequences could not be determined as the full sequence of Nk-3FTx was not obtained. Alignment with the reported sequences of 3FTx from *N. kaouthia* venom from different geographical origin showed substitutions of amino acids at various positions of the sequence (Figure 5.9). Comparative analysis between CM-9a and Nk-3FTx showed substitutions of 6 amino acids in Nk-3FTx. Amino acids Leu34, Asp43, Val48, Glu53, Ile54 and Lys62 in CM-9a were replaced by Phe34, Ala43, Lys48, Asp53, Val54 and Arg62 in Nk-3FTx (Figure 5.9). Also, when compared to the second highest similar 3FTx, a tryptophan containing 3FTx (P82935.2), it was observed that Trp28 was replaced by Leu28 in Nk-3FTx.

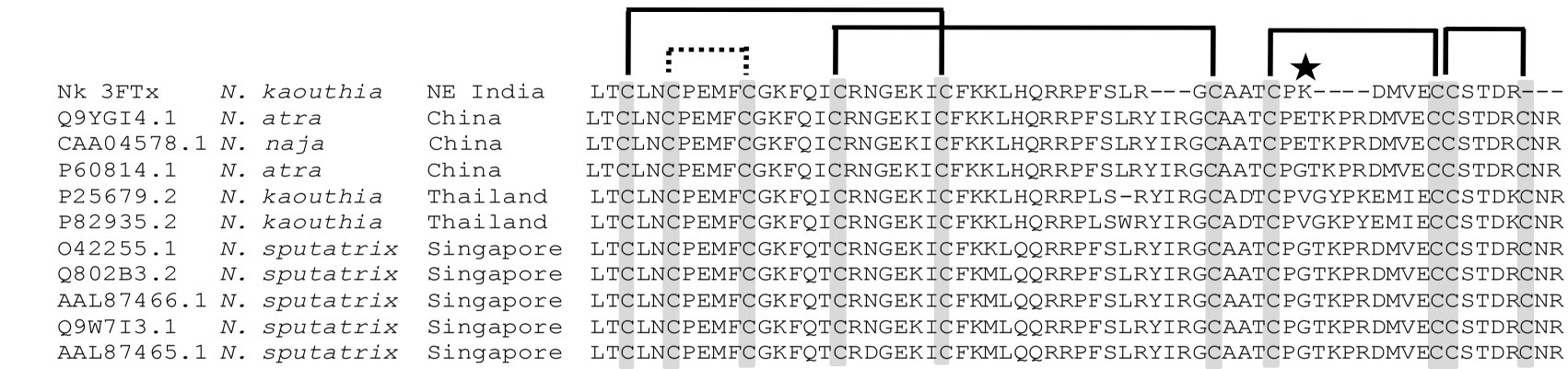

Figure 5.8: Multiple sequence alignment of amino acid sequence of Nk-3FTx. The conserved cysteine residues are highlighted with grey. The solid line represents the four disulfide bridges present in the core. The dotted line represents the fifth disulfide bond between Cys2 and Cys3 present in the first loop. Presence of fifth disulfide bridge in the first loop confirms it as a non-conventional 3FTx. Substitution of amino acid with other closely related 3FTx is marked with a star. The sequences were retrieved from the database which showed maximum similarity during BLASTp.

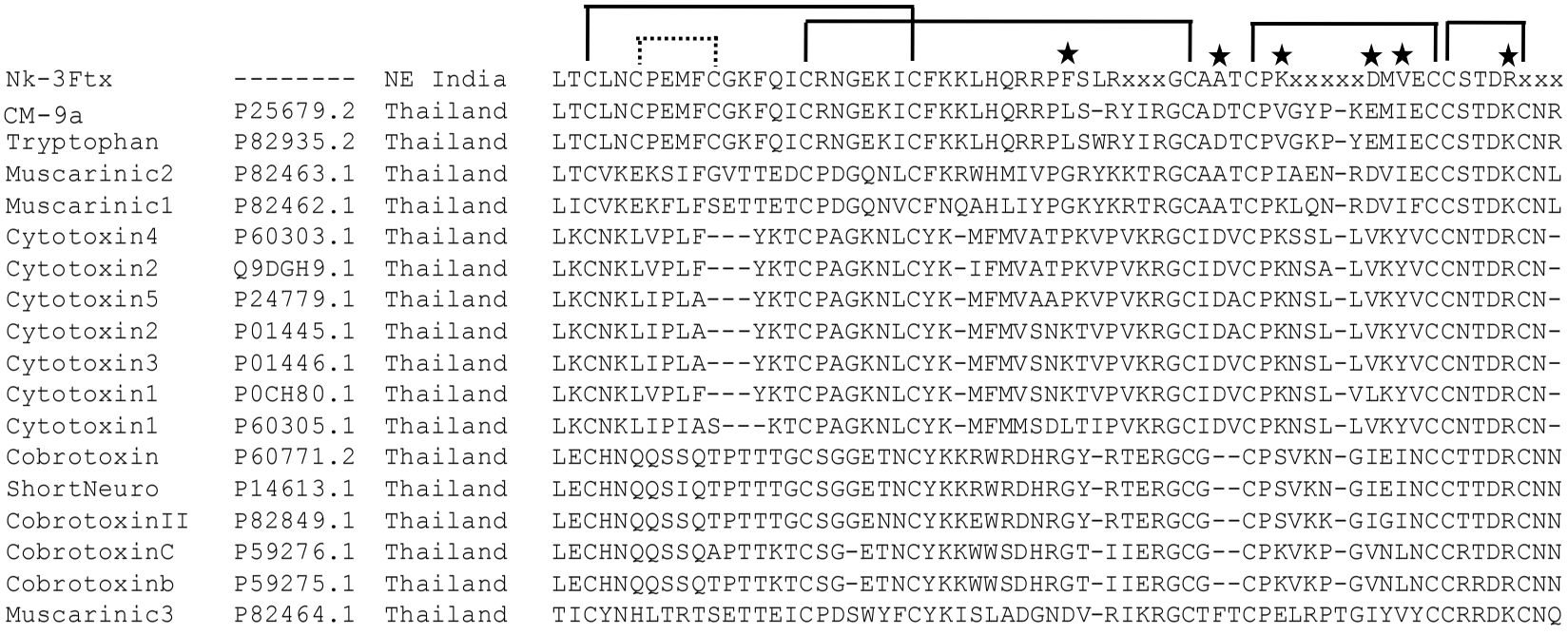

Figure 5.9: Multiple sequence alignment of Nk-3FTx with *N. kaouthia* 3FTxs from different geographical locations. The linkage of Cys residues are shown in solid line and the fifth disulphide bridge is shown in dotted lines. Unidentified amino acid residues of Nk-3FTx are denoted by "x" and the gaps are shown by dashes. The substituted amino acids in Nk-3FTx as compared to CM-9a are marked in stars.

5.3 Discussion

Snake venom is a mixture of proteins, peptides and other compounds. Despite being highly toxic, snake venom is a natural biological reservoir of potential therapeutic molecules. With advanced proteomics and genomics, studies of snake venom have helped tremendously in isolating and purifying components which aids in understanding its therapeutic properties. Deciphering the therapeutic property of a particular toxin helps in future applications in both basic research and pharmaceutical studies.

The partial compositional analysis of crude venom of *N. kaouthia* from Northeast India by LC-MS analysis reveals 3FTx as the major family followed by PLA_2, Ohanin/vespryn, thaicobrin and C-type lectins based on their molecular mass. Similarly the venom of Thailand origin is also reported to contain 3FTx as the major protein family [68]. The fractionated protein peaks eluted together with higher molecular weight and also low molecular weight proteins which needed further investigation. This is because snake venom components do exists in multimeric states or forms complexes which might elute in the same fraction [87, 138, 300]. However, there might be variation in expression level of 3FTx and presence of unique toxins in these venoms which cannot be denied. 3FTxs are structurally similar but functionally diverse group of proteins [131, 301] and binds to their targets through specific amino acid residues [302].

The molecular mass of Nk-3FTx (7579.5 ±0.591Da) was found to be similar with 3FTxs molecular mass range. Identification of 55 amino acid residues of Nk-3FTx by a combination of N-terminal sequencing and ESI LC-MS/MS and BLAST search confirmed that it belongs to 3FTx family. The 3FTx superfamily consists of cytotoxins/cardiotoxins [303], short and long chain neurotoxins [298] and non-conventional toxins [86] and others. The number of amino acid residues varies in each class. The cytotoxin/cardiotoxins and short chain neurotoxins [303] are consists of 60-62 amino acids, long chain neurotoxins contains 66 to 74 amino acid residues [298], non-conventional toxins are monomers of 62-68 amino acids [86].

BLAST search of Nk-3FTx revealed several 3FTxs out of which NNAM2 (Q9YGI4) and one unnamed protein (CAA04578.1) showed similar E-score value. The

sequence Q9YGI4 from *Naja atra* was obtained from the venom gland cDNA [304] and sequence CAA04578.1 from *N. kaouthia* venom were not characterized for functional studies. The sequences are reported to have higher similarity with long chain neurotoxin homologs than with α-neurotoxins, κ-neurotoxins and cardiotoxins [304]. However, presence of a fifth disulfide bond in the first loop show the two toxins as well as Nk-3FTx belong to the non-conventional group of 3FTx. 3FTxs may vary in its function due to single amino acid substitution; this is the reason why 3FTxs exerts multiple toxicity despite being structurally similar [85]. Substitution of Glu44 in Q9YGI4 and CAA04578.1 to Lys44 in Nk-3FTx might attribute different pharmacological activity. The isoforms of 3FTx are reported to undergo ASSET in their evolution [153]. Due to such exchange of amino acid segments, functionally important residue alteration might affect their functional properties. 3FTxs are found to be functionally evolved through ASSET and such changes brings evolution of novel functions [146, 154]. The presence of fifth disulfide bridge in non-conventional toxin is proposed to be due to exchange of segments [153]. However, the substitutions of amino acids in Nk-3FTx at six different positions as compared with CM-9a, which might be due to accelerated rate of point mutation. Both ASSET and point mutation in 3FTxs can contribute to the functional diversity [153]. Accelerated point mutations within a toxin family is reported to be a key factor for binding to similar receptors due to minor changes in their surface charge and topology [153].

Multiple sequence alignment of Nk-3FTx with other reported 3FTxs from *N. kaouthia* venom reveals Nk-3FTx as the first report of a 3FTx from Northeast Indian origin. The NCBI database reveals all the reported 3FTx sequences are from Thailand. Nk-3FTx was found to be most similar to CM-9a of *Naja naja kaouthia* of Thailand origin among the reported 3FTxs from *N. kaouthia* species [191], however substitutions of amino acid were observed at six positions with CM-9a. The partial amino acid sequence of Nk-3FTx was found to carry 11 basic amino acids (5 Lys residues, 5 Arg residues and 1 His residue) and 5 acidic amino acids (3 Glu acid residues and 2 Asp acid residues). Mostly, 3FTxs are reported to have a conserved Tyr residue at 25th or homologue aromatic amino acid residue Phe at 27th [170, 305] positions, however in case of Nk-3FTx such amino acid were not observed. Instead Lys and Leu were found to be present at 25th and 27th position. Some monomeric charged 3FTxs such as Arg39 in erabutoxin-a and Asp60 in α-cobratoxin are

reported to stabilize the native conformation of the protein by forming a salt link with C or N terminus of the toxin which has two residues in its N-and C-terminal before and after the first and the last cysteine residues [306]. In Nk-3FTx the amino acid present in 39th and 60th position could not be detected due to small peptide fragments of three amino acid residues which might carry the similar amino acids, however presence of two short residues in its N-terminal end before the first cysteine residue in Nk-3FTx and presence of last two residues after C-terminal cysteine in the reference sequence CM-9a, makes Nk-3FTx likely to be a monomeric protein. Considering the snake venom α-neurotoxin, Tremeau and co-workers performed site directed mutagenesis and identified the functional site of erabutoxin-a, a short chain neurotoxin [169]. The important residues at tip of the loop I were Gln7, Gln10 and Ser8, which were found to be functionally essential for binding to acetylcholine receptor. In the stretch between loop I to loop II, the mutational analysis showed no important residue having affinity to bind to acetylcholine receptor, however mutation at Ser18 showed substantial structural change. Loop II functional residues were already determined by Pillet and co-workers as Lys27, Trp29, Asn31 and Arg33 and some lesser affinity residues Phe32, Gln34 Glu38 and no other residues having significant effect was determined [307]. In loop III, the concave face oriented Lys47 was found to be functionally important [169]. However, in Nk-3FTx except Lys27, none of the amino acid residues were observed, which shows it is not a short chain neurotoxin. Also, in case of long neurotoxin it was been reported that generally they have additional 2-9 residues in their C-terminal end which was not observed in the reference sequence CM-9a and also might not be present in Nk-3FTx [85].

In the present study we have purified and characterization a previously uncharacterized non-conventional 3FTx from the venom of *N. kaouthia* of northeast India origin. Although, reports of CM-9a and other approximately similar proteins are available, Nk-3FTx differs in its amino acid composition. Also, most of the 3FTx reported were found to be cytotoxins, cobrotoxins and muscarinic toxins from the venom of *N. kaouthia* of Thailand origin, however functional characterization is required to determine the biochemical and biological property of Nk-3FTx [68, 162, 164, 191, 220, 283].