STRUCTURAL AND FUNCTIONAL STUDIES OF TRANSLIN-LIKE PROTEINS

Submitted by

Gagan Deep Gupta

(Enrollment no. LIFE01200604008)

Bhabha Atomic Research Centre, Mumbai-400085

Under the guidance of

Dr. Vinay Kumar

Professor, Homi Bhabha National Institute

A thesis submitted to

Board of Studies in Life Sciences

In partial fulfillment for the degree of

Doctor of Philosophy

of

Homi Bhabha National Institute

Mumbai, India

February 2012

Homi Bhabha National Institute

Recommendations of the Viva Voce Board

As members of the Viva Voce Board, we certify that we have read the dissertation prepared by Gagan Deep Gupta entitled "Structural and Functional studies of Translin-like proteins" and recommend that it may be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

Chairman, Dr S. F. D'Souza	Date
Convener, Dr Vinay Kumar	Date
Examiner, Dr B. Gopal	Date
Member, Dr M. V. Hosur	Date
Member, Dr S. Zingde	Date

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to HBNI.

CERTIFICATE

I hereby certify that I have read this dissertation prepared under my direction and recommend that it may be accepted as fulfilling the dissertation requirement.

Date: 10/08/2012

Place: Mumbai

(Dr Vinay Kumar)
Supervisor

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at Homi Bhabha National Institute (HBNI) and is deposited in the Library to be made available to borrowers under rules of the HBNI.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgement of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the Competent Authority of HBNI when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

(Gagan Deep Gupta)

DECLARATION

I, hereby declare that the investigation presented in the thesis has been carried out by me. The work is original and has not been submitted earlier as a whole or in part for a degree/diploma at this or any other Institution/University.

(Gagan Deep Gupta)

Dedicated

to

my family

Acknowledgements

First and foremost I offer my sincerest gratitude to my mentor and research supervisor Dr Vinay Kumar who introduced me to the fascinating field of structural biology. This thesis would not have been possible without his encouragement and efforts. He supported me throughout my thesis with his patience and knowledge. One simply could not wish for a better or friendlier supervisor.

The work presented in this thesis was carried out at Bhabha Atomic Research Centre (BARC). I sincerely acknowledge the research facility, infrastructure and funding provided by my institute. I express my regards to Dr S. K. Sikka and Dr S. K. Mahajan for their constant support and advice that helped me to achieve this land mark of my scientific career.

I am grateful to Prof B.J. Rao (TIFR, Mumbai) who introduced me with translin proteins and provided me all the help including the clones of drosophila and human translin, and human trax. I sincerely thank the honorable members of my doctoral committee, Drs J. Duttagupta, S.F. D'souza, M.V. Hosur, Mrs S. Zingde and M. Seshadri. Their critical evaluation and suggestions helped me to refine the experimental strategies.

I offer my sincere regards to Dr S.M. Sharma, Head, High Pressure & Synchrotron Radiation Physics Division for providing me all the necessary facilities and help, to complete my research work. I am also thankful to Dr S.K. Apte, Dean-Academics (Life sciences), HBNI to give me the opportunity to pursue my doctoral degree in this prestigious institute.

A special thanks to Dr Ravindra Makde, my colleague and a very good friend, who have been a constant source of help and support. He made my transition from the world of particle-physics to molecular biology quite easy.

I sincerely thank to Dr T.P. Singh (AIIMS, New Delhi), Drs B. Gopal and M.R.N. Murthy (IISc, Bangalore) and Dr M.V. Hosur (BARC) for their help in diffraction intensity data acquisition. I also thank the Department of Biotechnology (DBT, India) for providing the data acquisition facility at European Synchrotron Radiation Facility (ESRF, France). I am grateful to Dr Dulal Panda (IIT, Bombay) for his help in MALDI-TOF and CD data acquisition, and Dr S. Zingde (ACTREC, Navi Mumbai) for her help in MALDI-TOF and DNA sequencing.

I am thankful to Dr Rita Mukhopadhyay for her help in radiolabeled experiments and DNA sequencing, Dr Sahayog Jamdar for his help in gel-filtration experiments and, Dr Ramesh Hire and Archana Rai for helping with many chemicals. I am also thankful to C.L. Prajapat for his help in UV-crosslinking experiments and, Vishal Parashar and Amit Das for their help in crystallization using robotics. I extend my gratitude to all my friends and colleagues who directly or indirectly helped to bring this work to the present form.

I am also indebted to the many countless contributors of crystallographic and bioinformatics community for providing the numerous tools and software, I have used to produce both my results and this thesis.

I express my love and regards to my parents who carved my persona with their unconditional love and support, and provided me strength during this journey. It is difficult to express my feelings for my better-half, Richa Gupta, who handled all my share of responsibilities towards my kids and my family with a smile. All I can say to her is "I love you". Last but not least, I offer my love to my kids, Utkarsh and Tamanna, for their patience and love.

Gagan Deep Gupta

Table of contents

Synopsis

Translin is a ssDNA/RNA binding protein and is very well conserved in eukaryotes [1,2]. Human translin, having 228 residues, shares 99% and 50% sequence identity with mouse and *drosophila* orthologs, respectively. Translin associated factor-X (trax), a 290 amino acid residue protein, is homolog of translin, and its interaction with translin is conserved in evolution [3]. Both the proteins play vital roles in cellular processes such as DNA recombination, spatial and temporal expression of mRNA, dendritic trafficking in neurons, and in siRNA processing [4-7]. The translin and trax proteins have also been reported to interact with several other important proteins [4,5]. Initial characterization of translin and trax revealed that translin, but not trax, forms a homomeric complex capable of binding ssDNA or RNA [8,9]. Analysis of the human, mouse or yeast translin demonstrated that these proteins form octamer, while chicken ortholog was observed to be a decameric complex [8,10]. The octameric structure was also observed in the crystal structure of human and mouse translin [11,12]. Also two basic motifs, B1 and B2, had been identified on translin sequence. The nucleic acid binding domain of translin results from a combination of basic motifs in multimeric structure [8]. Loss of multimeric structure or substitution in the B2 motif abrogates its nucleic acid binding ability [8]. A mutant of *Drosophila melanogaster* (drosophila) translin (P168S) was observed to be DNA binding incompetent although the substitution was not in the known DNA binding motifs [13].

The present thesis aims to understand structure-function of translin proteins using structural, bioinformatics, mutational and biochemical analysis. The brief outlines of the studies conducted are:

- Three-dimensional structure of DNA binding incompetent drosophila mutant translin was elucidated using single crystal X-ray diffraction methods.

- Role of C-terminal residues in translin oligomeric status was established, using chimera (fusion) proteins synthesized by genetic engineering methods.
- Translin-trax complexes were characterized with the aim to understand the DNA binding properties of translin and trax proteins, and to understand the mechanism of formation of the complex.
- Evolutionary linkage between translin and trax was established using bioinformatics methods.

Chapter 1 of the thesis includes introduction to translin proteins and crystallographic techniques. A brief review of literature on translin family is given. Also, recent developments in protein crystallography are discussed here.

The details of 'Materials and Methods' used in the present investigations are described in **Chapter 2**. The methods and protocols are grouped in three sections namely, molecular biology, protein chemistry and structural biology. In molecular biology section, protocols for cloning, expression and mutagenesis are elaborated. Protein chemistry section includes description of protocols for protein purification, and protein estimation and detection. Biophysical and biochemical methods such as molecular mass determination, circular dichroism (CD) and DNA-binding assays, used to characterize proteins and protein-DNA interactions are also included in this section. The methods employed in this study for protein crystallization and structure analysis are discussed in the last section of this chapter.

Chapter 3 includes the elucidation of three-dimensional crystal structure of drosophila P168S mutant translin. This mutant translin lacks the nucleic acid binding activity. The mutant protein with N-terminal 6XHis tag (6XHis-translin) was crystallized in space-group $P222_1$. Initial crystallographic phases were obtained

with molecular replacement method using partial atomic model of human translin (1-170 residues). The structure was refined using diffraction data upto 3.4Å (R_{work}/R_{free} 0.203/0.269, PDB ID 2QRX). The drosophila mutant translin (lacking affinity tag) was purified with conventional chromatographic techniques and could be crystallized in $P6_222$ space-group. The crystal structure in the hexagonal space-group was refined against 3.6Å resolution diffraction data (R_{work}/R_{free} 0.215/0.249, PDB ID 2QVA). The structures of translin mutant in both the crystal forms were identical. The monomeric structure of translin mutant matches up to 187 amino acids with the corresponding residues of human and mouse orthologs with an RMSD of 1.1Å, but differs at its C-terminal region. The oligomeric structure of drosophila mutant translin also differs dramatically. The drosophila translin mutant forms an open doughnut shaped tetramer in contrast to closed octameric structure of human and mouse proteins. This altered quaternary structure of drosophila mutant translin was taken to be responsible for the abrogation of its nucleic acid binding ability. The structural analysis also revealed that up-down dimer is energetically stable and evolutionary conserved unit in translin structures. It was shown that different oligomeric status observed in translin orthologs can be achieved by repetition of a number of up-down dimers [14].

The biochemical and biophysical characterization of wild-type drosophila translin, a chimeric translin and translin-trax complex is described in **Chapter 4**. Wild-type drosophila translin was constructed by restoring proline at position 168 in the mutant translin. The protein was found to be DNA binding competent. Its molecular weight was estimated to be ~295 kDa using gel-filtration technique. Interestingly, GTP binding to drosophila translin was not detected in dot blot assay. In contrast, GTP binds to human translin and modulates its nucleic acid binding

properties [15]. Despite extensive trials, the best obtained crystals of wild-type drosophila translin diffracted to limited resolution of 6Å only. A chimeric protein, comprising of 186 N-terminal amino acids of drosophila translin (with P168S mutation) and 43 C-terminal residues of human translin was synthesized using primer overlap extension method (GenBank accession no.: EU155117). The chimeric translin protein attained octameric state and was DNA binding competent. The migration of its gel-shifted complex was similar to that of human translin. These results emphasized the critical role of C-terminal residues in the formation of DNA binding competent oligomer [14].

Translin associated factor-X (trax) alone is not stable and is removed by ubiquitination pathways in translin null mutants [3]. However, it forms a very stable and nucleic acid binding competent complex with translin [16]. The 6XHis-trax (TRAX) and translin were co-expressed in *E. coli* BL21 (DE3) cells. The translin-TRAX complex was purified to homogeneity and, was observed to have 1:1 stoichiometry of translin and TRAX. Molecular mass of the complex was estimated to be ~300 kDa that corresponds to octamer/decamer formed with four or five copies each of translin and trax proteins. The CD spectrum of translin-TRAX complex was nearly identical to that of homomeric translin suggesting very similar secondary structure of the two proteins. Based on multiple sequence alignment of trax and translin orthologs, the partial structural model of human trax (residues 38-264 of trax sequence) could be constructed using human translin structure as a template. The constructed structural model matches very closely with the recent crystal structures of translin-trax complexes [17,18]. Various truncated forms of trax were synthesized, based on modeled structure, with the expectations to stabilize trax. However, the truncated mutants were observed in inclusion bodies only. Attempts to

refold the protein from inclusion bodies were unsuccessful. Further to understand mechanism of formation of translin-trax complex, heterologous complex of drosophila P168S mutant translin with human trax was synthesized. Surprisingly, the heterologous complex was observed to be structurally analogous to human translin-trax complex, despite dramatic differences in quaternary structure of human translin and drosophila mutant translin. Also, the heterologous complex was DNA-binding competent. The data thus suggested that translin-trax proteins co-oligomerize *in situ*.

Earlier reports described that trax itself can not bind DNA but modulates the nucleic acid binding properties of translin. Based on structural model and bioinformatics analysis, we hypothesized that trax contributes towards DNA binding activity of translin-trax complex. **Chapter 5** covers the characterization of trax for its affinity towards ssDNA. The DNA binding affinity of trax was assayed using translin-TRAX complex (complex of 6XHis-trax with translin) as trax alone is not stable. To detect binding of DNA with trax, the radiolabeled DNA was crosslinked with translin-TRAX complex using UV-laser. The irradiated complex was disrupted into subunit proteins with 8M urea and the TRAX-DNA covalent complex, purified under denaturing conditions, was detected on SDS-PAGE by autoradiogram on X-ray film. Purification of TRAX alone was also confirmed by MALDI-TOF analysis. Further two DNA binding motifs, designated B2 and B3, were identified on trax sequence. The mutants of these motifs, (traxB2- ^{115}QFHRA^{119} to ^{115}LFNAA^{119}) and (traxB3- ^{241}KKLY^{244} to ^{241}NTLN^{244}), were synthesized using QuickChange site-directed mutagenesis. Various heteromeric translin-TRAX complexes were synthesized with combinations of translin and trax mutants, and mutation in traxB3 motif was observed to adversely influence ssDNA binding ability of the complex.

The CD-spectra and gel-filtration elution profiles for mutant heteromeric complexes were observed to be similar to that of wild-type complex. Taken together, the data suggested that loss of activity was solely due to substitution in residues responsible for DNA binding activity. **Chapter 5** also includes identification of new DNA binding site, designated B3, on translin sequence. Substitution in B3 motifs, (translinB3, ^{192}RKRY^{195} to ^{192}TNSN^{195}), abrogated DNA binding activity of homomeric translin. To structurally characterize interactions of DNA, human translin was co-crystallized with a 24-mer ssDNA. The translin-DNA complex crystals diffracted to 2.5Å resolution using synchrotron radiation source at ESRF, France. However, clear electron density to fit DNA molecule could not be detected in the Fourier maps.

Chapter 6 includes evolutionary analysis of translin/trax proteins in view of recently available structure-function details along with the concluding remarks on the work presented in the thesis. It was found that both translin and trax were conserved in all eukaryote kingdoms including Amoebozoa. Based on multiple sequence alignment and Logo-plots, specific sequence motifs were identified for each of translin and trax protein family. Interestingly, only single copy of translin-like protein could be detected in prokaryotes. The prokaryote proteins possess the sequence motifs specific to trax proteins. The acidic amino acid triad (two glutamate and one aspartate), expected to be the Mg^{2+} coordinating site, crucial for RNase activity of trax were conserved in the prokaryotic translin-like protein. However, translin specific sequence motifs were missing in the prokaryotic proteins. Based on the conservation of biologically important residues, it is being proposed that trax is the direct descendant of prokaryote translin-like protein (ortholog). The phylogenetic analysis suggests that translin (paralog of trax) evolved in eukaryotes,

probably with gene duplication event. This chapter also includes the concluding remarks on the work presented in the thesis.

Followings are the conclusions drawn from the presented work.

- Up-down dimer is the minimum structural unit, energetically stable, in translin proteins and different oligomeric states observed in translin orthologs can be achieved by repetition of these up-down dimers.
- The quaternary structure of translin proteins play important role in DNA binding activity and change in quaternary structure leads to loss of activity.
- The C-terminal residues are critical for oligomeric status of the translin proteins.
- Translin-associated factor-X, the trax protein, is DNA binding competent. Two DNA binding motifs (B2 and B3) have been identified on trax sequence using mutational studies.
- The B3 domains of both translin and trax contribute dominantly to DNA binding activity of biologically important heteromeric translin-trax complex.
- The ancestor of eukarotic translin and trax is closer to trax and may possess the RNase activity observed for trax proteins.

Future prospectives

Identification of nucleic acid binding motifs on trax may accelerate research in understanding the mechanism of RNase activity observed for translin-trax complex. Also, many of the interacting partners of translin/trax proteins are involved in DNA metabolic pathways, the work presented here may direct future studies on role of translin/TRAX in DNA metabolism. Functional studies of prokaryote translin-like protein may provide important clues on translin-independent function of trax proteins.

REFRENCES

1. Aoki K, Suzuki K, Sugano T, Tasaka T, Nakahara K et al. (1995) A novel gene, Translin, encodes a recombination hotspot binding protein associated with chromosomal translocations. Nat Genet 10: 167-74.
2. Wu XQ, Gu W, Meng X, Hecht NB (1997) The RNA-binding protein, TB-RBP, is the mouse homologue of Translin, a recombination protein associated with chromosomal translocations. Proc Natl Acad Sci USA 94:5640-5.
3. Yang S, Cho YS, Chennathukuzhi VM, Underkoffler LA, Loomes K et al. (2004) TRAX is post-transcriptionally regulated by its partner protein TB-RBP and both are essential for normal cell proliferation. J Biol Chem 279:12605-14.
4. Jaendling A, McFarlane RJ (2010) Biological role of translin and translin-associated factor-X: RNA metabolism comes to the fore. Biochem J 429:225-34.
5. Li Z, Wu Y, Baraban JM (2008) The Translin/TRAX RNA binding complex: Clues to function in the nervous system. Biochim Biophys Acta 1779:479-85.
6. Liu Y, Ye X, Jiang F, Liang C, Chen D et al. (2009) C3PO, an endoribonuclease that promotes RNAi by facilitating RISC activation. Science 325:750-3.
7. Lluis M, Hoe W, Schleit J, Robertus J (2010) Analysis of nucleic acid binding by a recombinant translin-TRAX complex. Biochem Biophys Res Commun 396: 709-13.
8. Aoki K, Suzuki K, Ishida R, Kasai M (1999) The DNA binding activity of Translin is mediated by a basic region in the ring-shaped structure conserved in evolution. FEBS Lett 443:363-6.
9. Chennathukuzhi VM, Kurihara Y, Bray JD, Hecht NB (2001) TRAX (Translin-associated factor X), a primarily cytoplasmic protein, inhibits the binding of TB-RBP (Translin) to RNA. J Biol Chem 276:13256-63.

10. Eliahoo E, Ben YR, Pérez-Cano L, Fernández-Recio J, Glaser F et al. (2010) Mapping of interaction sites of the Schizosaccharomyces pombe protein Translin with nucleic acids and proteins: a combined molecular genetics and bioinformatics study. Nucleic Acids Res 38:2975-89.
11. Pascal JM, Hart PJ, Hecht NB, Robertus JD (2002) Crystal Structure of TB-RBP, a novel RNA-binding and regulating protein. J Mol Biol 319:1049-57.
12. Sugiura I, Sasaki C, Hasegawa T, Kohno T, Sugio S, Moriyama H, Kasai M, Matsuzaki T (2004) Structure of human translin at 2.2 Å resolution. Acta Cystallogr D60:674-79.
13. Sengupta K, Kamdar RP, D'Souza JS, Mustafi SM & Rao BJ (2006) GTP-induced conformational changes in translin: a comparison between human and Drosophila proteins. Biochemistry 45:861-70.
14. Gupta GD, Makde RD, Roa BJ, Kumar V (2008) Crystal structures of *drosophila* mutant translin and characterization of translin variants reveal the structural plasticity of translin proteins. FEBS J 275: 4235-49.
15. Chennathukuzhi VM, Kurihara Y, Bray JD, Yang J & Hecht NB (2001) Altering the GTP binding site of the DNA/RNA-binding protein, Translin/TB-RBP, decreases RNA binding and may create a dominant negative phenotype. Nucleic Acids Re*s* 29:4433-40.
16. Gupta GD, Makde RD, Kamdar RP, D'Souza JS, Kulkarni MG et al. (2005) Co-expressed recombinant human Translin–TRAX complex binds DNA. FEBS Lett 579:3141-6.
17. Ye X, Huang N, Liu Y, Paroo Z, Huerta C et al. (2011) Structure of C3PO and mechanism of human RISC activation. Nat Struct Mol Biol 18:650-7.

18. Tian Y, Simanshu DK, Ascano M, Diaz-Avalos R, Park AY et al. (2011) Multimeric assembly and biochemical characterization of the TRAX-translin endonuclease complex. Nat Struct Mol Biol 18: 658-664.

Abbreviations

Abbreviation used	Full name
ATP	Adenosine triphosphate
APS	Ammonium per sulphate
bp	Base pair
BSA	Bovine serum albumin
DW	Distilled Water
dNTP	Deoxyribonucleoside triphosphate
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic Acid
EMSA	Electrophresis mobility shift assay
GTP	Guanosine triphosphate
IPTG	Isopropyl β-thiogalactopyranoside
kDa	Kilo-Dalton
LB	Luria Bertani broth
Mr	Molecular weight
MWM	Molecular weight marker
MPEG	Polyethylene glycol monomethyl ether
Ni-IDA	Ni(II)-iminodiacetic acid
O/N	Over night
OD	Optical Density
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase Chain Reactions
PEG	Polyethylene glycol
RISC	RNA induced Silencing complex
RMSD	Root mean square deviation
RCF	Relative centrifugal force
rpm	Rotation per minutes
RT	Room Temperature
ssDNA	Single stranded DNA
TAE	Tris-Acetate-EDTA
TBE	Tris-Borate-EDTA
TE	Tris-EDTA
TRAX	6XHis-trax
UTR	Untranslated region

Other than these, standard SI units were used throughout the text. For protein and DNA sequences, single letter codes were used. The residues positions were described as three letter code followed by position in the sequence (Cys-225; Cysteine at posion 225). The point mutations were depicted with single letter code (P168S stands for Pro-168 is substituted to Ser).

Amino acid	**Three letter code**	**One letter code**
alanine	Ala	A
arginine	Arg	R
asparagine	Asn	N
aspartate	Asp	D
cysteine	Cys	C
glutamate	Glu	E
glutamine	Gln	Q
glycine	Gly	G
histidine	His	H
isoleucine	Ile	I
leucine	Leu	L
lysine	Lys	K
methionine	Met	M
phenylalanine	Phe	F
proline	Pro	P
serine	Ser	S
threonine	Thr	T
tryptophan	Trp	W
tyrosine	Tyr	Y
valine	Val	V

List of figures

List of Tables

Chapter 1

General Introduction

INTRODUCTION

"There are two fundamental conditions of life. First, the living entities must be able to self-replicate and another is organisms must be able to perform all cellular processes efficiently and selectively" (Nelson & Cox, 2004). The first condition requires the protection of genome from the external and internal challenges and second condition necessitates the regulation of gene expression. These tasks are achieved by a special class of proteins, nucleic acid binding proteins. Genome structure and stability, DNA replication, DNA recombination and repair, transcription, post-transcriptional regulation, mRNA transport and storage, translation and many more tasks are performed by nucleic acid binding proteins. Translin is a protein that binds with both DNA and RNA and is associated with many of these processes. As a DNA binding protein, it is associated with chromosomal translocations, DNA recombination, and DNA repair mechanism (Kasai et al., 1997). Translin as an RNA binding protein regulates the mRNA transport and translations, and plays a key role in germ line development during spermatogenesis, and in neuronal function and development (Wu et al., 1999b; Jaendling et al., 2010). A heteromeric complex of translin and translin-associated factor-X (trax) is involved in mRNA trafficking in neuronal dendrites (Li et al., 2008). The heteromeric translin-trax complex has also been reported as an activator of RNA induced silencing complex (RISC) (Liu et al., 2009). Translin and trax proteins share sequence homology and are very well conserved in eukaryotes, including single cellular eukaryote yeast through to humans. Considering that these proteins are involved in a variety of biological functions and are highly conserved in evolution, it is of immense importance to understand the mechanism how these proteins function? During the work presented in this thesis, I have studied the functional aspect of translin and trax proteins in view of their three dimensional

structure. In this chapter, I have provided an introduction to translin and trax proteins and a brief overview of the published work. Also, I have discussed the recent developments in the field of protein X-ray crystallography.

AN OVERVIEW OF TRANSLIN AND TRAX PROTEINS

Identification of translin and trax protein

Translin is a single-stranded DNA (ssDNA) and RNA binding protein that was identified by different groups working on altogether different systems. Aoki et al. (1995) identified a DNA binding protein that exhibits strong affinity towards the consensus sequences found at the breakpoint junctions of chromosomal translocations in human lymphoid neoplasms and thus named it translin (**transl**oca**t**io**n**). Translin was also identified in brain and testis extracts of mouse that bound to highly conserved cis-acting sequences in the 3'-untranslated regions (UTRs) of a number of testicular and brain mRNAs and it was named TB-RBP (**T**estis-**B**rain **R**NA **B**inding **P**rotein; Han et al., 1995a & 1995b). The TB-RBP was later recognised as mouse orthologue of human translin (Wu et al., 1997). Taira et al. (1998) and Aaharoni et al. (1993) identified translin independently as a protein that specifically binds with GC-rich sequences in brains. These studies thus suggested that translin is a multi-functional protein involved in a variety of cellular processes.

Human translin sequence was observed to possess nuclear export signal sequence (NES) but not the nuclear localization signal (NLS). To find the interacting partner that facilitates the localization of translin to nucleus, Aoki et al. (1997) performed yeast two hybrid assays and identified a 33 kDa protein **tr**anslin-**a**ssociated factor-**X** (trax). Trax was observed to share significant sequence identity with translin and possessed a bipartite nuclear localization signal sequences. Co-precipitation of trax and translin in immunoprecipitation studies, performed with translin specific

antibodies, suggested a close association of trax with translin (Wu et al., 1999a). Both translin and trax proteins were also identified as the components of a GS-1 strand specific DNA binding complex (Taira et al., 1998) and a ribonucleoprotein complex observed in neurons (Murumatsu et al., 1998; Kobayashi et al., 1998). Recently, translin-trax complex was identified as a key promoter of RNA induced silencing complex (RISC) in *Drosophila melanogaster* (drosophila) and human (Liu et al., 2009; Ye et al., 2011), and was named as C3PO (component 3 promoter of RISC). These observations taken together suggested a close interaction between translin and trax proteins. Both translin and trax are very well conserved in the evolution and are observed in chicken, *Xenopus*, drosophila, *Arabidopsis thaliana* and fission yeast *Schizosaccharomyces pombe* (S. pombe) sharing high sequences identities with their human orthologs (Figure 1.1).

Translin-null mutants of mice were observed to lack both translin and trax proteins, without detectable alterations in the mRNA level of trax and, trax was removed via the ubiquitin pathways (Chennathukuzhi et al., 2003a; Yang et al., 2004). The results suggested that stabilization of trax is dependent on translin and trax is post-transcriptionally regulated by the translin (Yang et al., 2004). The post-transcriptional regulation of trax was also observed in drosophila and S. pombe (Claussen et al., 2006; Jaendling et al., 2008). However, trax null mutants of drosophila and S. pombe exhibited normal level of translin.

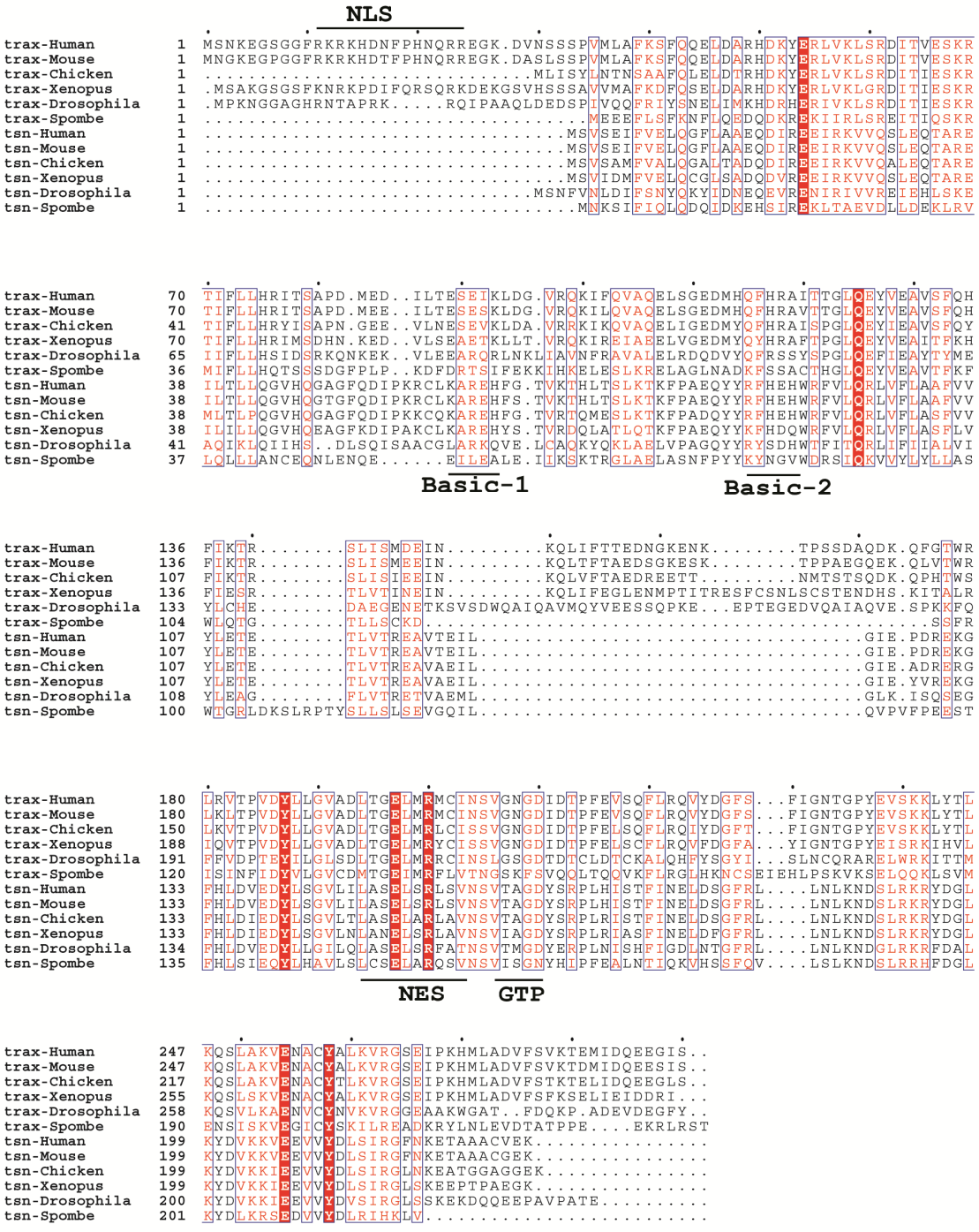


Figure 1.1 Alignment of trax and translin sequences. The amino acid sequences (UniProtKB/Swiss-Prot entries) of the trax and translin proteins of Human (**Q99598 and Q15631)**, Mouse (**Q9QZE7** and **Q62348)**, Chicken (**E1BQ94** and **P79769)**, Xenopus (**Q7SZ15** and **Q9IAM5)**, Drosophila (**Q8INE1 and Q7JVK6)** and Spombe (**O74955** and **Q9P7V3)** were aligned by ClustalW (Thompson et al., 1994). The identical amino acid residues are shaded and similar residues are boxed. The nuclear localization signal (NLS) on the human trax sequence is identified. Also marked are two basic regions that have been implicated in nucleic acid binding (Basic-1 and Basic-2), GTP binding region and the nuclear export signal (NES) on human and mouse translin sequences.

Role of translin/trax proteins in DNA repair and recombination

Translin exhibited strong affinity towards the consensus sequences found at the breakpoint junction of chromosomal translocation in human lymphoid neoplasms (Aoki et al., 1995; Sengupta et al., 2002; Jacob et al., 2004). Nucleic acid consensus motifs recognised by translin were later observed in many malignant conditions those were associated with chromosomal translocation, like synovial sarcoma, myeloid leukaemia and, myxoid and round cell liposarcomas and many more (Chalk et al., 1997; Jeffs et al., 1998; Atlas et al., 1998; Kanoe et al., 1999; Hosaka et al., 2000; Wei et al., 2003; Xiang et al., 2008).

Translin, primarily a cytoplasmic protein, was observed in nucleus in lymphoid cell lines (Aoki et al., 1995). The localization from cytoplasm to nucleus was also induced in response to DNA damaging reagents like Mitomycin C (Kasai et al., 1997). In addition, translin was observed to bind both *in vivo* and *in vitro* with GADD34, a DNA damage-inducible protein (Hasegawa & Isobe, 1999). These findings hinted at the likely role of translin in DNA repair. The hypothesis was further supported as the translin-deficient haemopoietic cells in mouse displayed a slow growth following irradiation with ionizing radiation (Fukuda et al., 2008).

The translin-associated factor X, trax, has also been shown to interact with DNA-binding proteins. Trax binds with C1D protein following exposure to γ-irradiations (Erdemir et al., 2002a). C1D is an activator of DNA-PK (DNA-dependent protein kinase) that is essential for the repair of DNA double strand breaks and V(D)J recombination, and the interaction of trax with C1D and translin is mutually exclusive (Erdemir et al., 2002a & 2002b). Lluis et al. (2010) have proposed a model that both trax and translin co-localize to the nucleus in meiotic pachytene spermatocytes and interact with DNA, presumably functioning in DNA crossing-over. Thereafter, trax

dissociates, exchanging translin for C1D, and freeing translin to interact with mRNAs marked for export from the nucleus via nuclear export signal present on translin sequence.

Despite the evidence discussed above, translin knock-out mice did not display any abnormalities in chromosomal recombination or higher sensitivity towards DNA damage (Chennathukuzhi et al., 2003a; Stein et al., 2006). Similar observations were reported for translin and/or trax null mutants of drosophila (Claussen et al., 2006). Taken together, a critical role of translin/trax proteins in DNA recombination or DNA repair was argued against. However, a systematic study is required to find the role of these proteins in redundant DNA repair/recombination pathways.

Role of translin/trax proteins in mRNA regulation and transport

During spermatogenesis sperm mother cell undergoes through various stages of development. These stages are highly organised temporally and spatially, and are controlled by regulated transcription and protein expression (Lewin, 2004). However, transcription ceases mid-way through spermiogenesis and stored mRNAs are translationally suppressed until the coded protein is required. Mouse translin was found to bind with specific y- and h- element sequences found at 3′-UTR of many mRNAs found in male germ cells and suppressed their translation (Han et al., 1995a & 1995b) suggesting its role in translation regulation. Also translin, localized in the nucleus during meiosis, was observed in cytoplasmic fractions in later developmental stages suggesting the role of translin in temporal regulation of mRNAs (Morales et al., 1998). Spermatids are haploid cells and contain either X or Y sex chromosome. To maintain the genetic equivalence between haploid cells, genetic information is shared by inter-cellular exchange of mRNAs. Translin interacts with KIF17b protein,

a motor protein required for driving mRNAs to microtubules (Chennathukuzhi et al., 2003b). Moreover, translin was observed to link the translationally delayed mRNAs to microtubules (Wu & Hecht, 2000) and thus proposed to be involved in inter- and intra- cellular transport of specific mRNAs in male mouse germ cells. Trax has also been shown to interact with Tsnaxip1, Golga3, Sun1 and Akap9 protein, expressed during spermatogenesis, *in vivo* and *in vitro* (Bray et al., 2002) and thus hinted a role during spermatogenesis.

Translin and trax were observed in high levels in brain extracts and have been proposed to function in RNA trafficking in neurons (Finkenstadt et al., 2000; Kobayashi et al., 1998). Chiaruttini et al. (2009) demonstrated the role of translin-trax complex in intracellular targeting of *BDNF* (brain-derived neurotrophic factor) mRNA. Various splice variants were observed for *BDNF* mRNA with variation in their 5′-UTR and thus specificity to target dendrites. However, translin-trax recognition sequence was observed in the common open reading frame (*orf*) of all splice variants. A point mutation (G to A base change) in the translin-trax recognition site on *BDNF* mRNA inhibited their dendritic localization. The mutation was associated with human neurological and psychological disorders. Moreover, siRNA mediated silencing of translin also blocked the dendritic trafficking of *BDNF* (Chiaruttini et al., 2009, Wu et al., 2011). Translin and trax were also observed to stimulate transcription of many mRNAs mediated through their interaction with steroidogenic factor-1 protein (Mellon et al., 2006).

In addition, Trax has been proposed to interact with adenosine A2A receptor (A2AR) and be involved in neuronal development (Zezula et al., 2008; Keuerleber et al., 2011; Aisiku et al., 2011). The impaired neuritogenesis caused by p53 blockade was rescued by activation of the A2AR in a trax-dependent manner (Sun et al., 2010).

Also *trax* gene, located in the vicinity of *DISC1* (**d**isrupted **i**n **sc**hizophrenia) gene, has been genetically linked to psychological disorders such that an intergenic splicing of *trax* and *DISC1* genes were observed to associate with schizophrenia (Cannon et al., 2005; Hennah et al., 2005; Thomson et al., 2005; Palo et al., 2007; Okuda et al., 2010).

Role of translin/trax in enhancing the activity of RNA-induced silencing complex

RNA mediated gene silencing is a phenomenon of regulation of gene expression by small RNAs, small interfering RNA (siRNA) and micro RNA (miRNA), derived from the cleavage of double stranded RNA (Bagasra & Prilliman, 2004). In post-transcriptional gene silencing, targeted mRNAs are degraded by RNA induced silencing complex (RISC). Small interfering RNAs guide the RISC to target mRNA based on sequence complementarity. In an attempt to reconstitute an active RISC complex of drosophila *in vitro*, Liu et al. (2009) purified RISC activity enhancer complex C3PO (component 3 promoter of RISC). The analysis revealed that C3PO is constituted by translin and trax proteins, and addition of heteromeric translin-trax complex, not translin alone, could enhance the RISC activity *in vitro*. Also, translin null mutants displayed very minimal RISC activity compared to wild-type. The C3PO complex was shown to degrade the passenger strand (Liu et al. 2009). Translin-trax complex was observed to exhibit the RNase activity in a Mg^{2+} dependent manner. Furthermore, three conserved acidic residues were identified on trax sequence such that substitution in any one of these resulted in the loss of RNase activity of translin-trax complex and its ability as RISC activity enhancer. Similar role of translin-trax complex as RISC activity enhancer was observed in humans (Ye et al., 2011).

Biochemical and structural analysis of translin and trax proteins

Human translin is a 228 residue protein that forms an octamer of 220 kDa molecular mass under native conditions (Aoki et al., 1995). Glycerol gradient ultra-centrifugation studies revealed that nucleic acid binding state of translin is an octamer (Lee et al., 2001). Initial characterization of human translin observed a leucine zipper like motif at the C-terminus of translin (184–212) and a Cys-225 involved in disulfide linkage (Wu et al., 1998). However, in crystal structure (discussed below) the residues proposed to from a leucine zipper were involved in intra-monomer contacts thus ruling out the leucine zipper in dimerization. Cys-225 could not be observed in crystal structure. The mouse, Xenopus and S. pombe translin were also observed to form an octameric complex while chicken translin was expected to form a decamer (Castro et al., 2000; Laufman et al., 2005; Aoki et al., 1999). The oligomeric status of translin was thought to be critical for DNA-binding activity. Several point mutations were scored those disturbed the oligomeric status and dramatically reduced DNA-binding activity of these proteins. For instance, substitution of Leu-184 and Leu-191 resulted in loss of octameric state and concomitantly, nucleic acid binding ability (Aoki et al., 1999). Two basic regions (basic-1 and basic-2) were identified on human and mouse translin sequences implicated in nucleic acid binding activity (Aoki et al., 1999). Mutations in basic-2 region (^{86}RFHEH^{90} to ^{86}TFNEN^{90}) abrogated the DNA and RNA binding activity of translin while the substitution in basic-1 region (^{60}KAREH^{64} to ^{60}NAQEN^{64}) caused the loss of RNA binding activity alone, without affecting the oligomeric state of the translin (Aoki et al., 1999). Basic-1 and basic-2 motifs, similar to human translin, were identified in S. pombe translin; point mutation in basic-2 region residues were observed to reduce the DNA/RNA binding activity, while basic-1 region residues did not influence the nucleic acid binding activity of S. pombe

translin (Eliahoo et al., 2010). Eliahoo et al. (2010) also identified Arg-211 and Arg-212 of S. pombe translin such that the substitution of theses residues resulted in significant loss of DNA/RNA binding activity. A putative GTP binding motif was identified in human and mouse translin sequences. GTP binding to mouse translin was observed to decrease the RNA binding affinity and the mutation in GTP binding motif (VTAGD to VTNSD) resulted in loss of GTP binding as well as RNA binding activity of translin (Chennathukuzhi et al., 2001b). In contrast, DNA binding affinity of translin was not influenced by either GTP binding or mutation in GTP-binding motifs (Chennathukuzhi et al., 2001b).

Trax protein alone is not stable by itself and requires translin for its stable fold. The observation was further strengthen by the report that trax, expressed alone in *E. Coli*, aggregated to inclusion bodies; however, a heteromeric complex of translin and trax could be purified by co-expressing translin and trax proteins in the bacterial co-expression system (Gupta et al., 2005). Trax protein has been proposed to modulate the nucleic acid binding properties of translin such that RNA binding activity of translin-trax complex reduced by more than 50%, while DNA binding activity increased as compared to translin alone (Chennathukuzhi et al., 2001a; Lluis et al., 2010). Also, addition of trax to the basic-2 region mutant of translin restored the nucleic acid binding activity (Chennathukuzhi et al., 2001a; Finkenstadt et al., 2002). However, trax alone solubilized with thioredoxin tag did not exhibit any affinity towards DNA or RNA (Chennathukuzhi et al., 2001a; Ye et al., 2011).

The Electron microscopy (EM) studies performed with native and DNA/RNA bound state of human translin at ~20 Å resolution revealed that translin formed an octameric ring shaped structure, with C8 symmetry, and DNA was bound to the central aperture of the ring (VanLoock et al., 2001). The diameter of central aperture

in octameric ring structure was 30 Å at one end and 50 Å at another end. However, the crystal structure of human and mouse translin proteins, which share more than 98% sequence identity, differ from the image constructed by EM studies. In contrast to the EM picture, the crystal structures showed close octameric barrel of the translin oligomers. The crystal structure analysis revealed that translin monomer has a highly helical structure that consists of seven α-helices (PDB ID 1J1J & 1KEY; Sugiura et al., 2004; Pascal et al., 2002). Monomer structure of human and mouse translin are very similar with r.m.s. difference (RMSD) of less that 1Å in C^{α} atoms. In human translin, four translin monomers are related to each other by pseudo four-fold symmetry to form a tetramer (Figure 1.2A) and two such tetramers form a closed octamer (ellipsoidal structure) with two fold symmetry (Figure 1.2B). A similar octameric structure was observed in mouse translin. The DNA/RNA binding residues (basic-2 region residues) are located in the inner surface (concave surface) of the ellipsoidal octamer; however, opening of ellipsoidal structure is only 4.4 Å, too narrow for entry of nucleic acid inside the cavity.

Recently, crystal structure of human and drosophila translin-trax complex have been reported (PDB ID, 3RIU & 3PJA; Ye et al., 2011; Tiara et al., 2011). The monomeric structure of trax was very similar to translin with RMSD of ~1.6 Å in superposed C^{α} atoms. The heteromeric human translin-trax was also observed to form an ellipsoidal shaped octamer analogous to homomeric translin, with six translin and two trax subunits (Figure 1.3). The acidic residues, essential for RNase activity of trax, were located in the internal cavity of octamer. These findings suggest that nucleic acid binding site is formed inside the octameric structure. However, how DNA/RNA enters into the central cavity is not yet clear, owing to narrow opening observed in crystal structure.

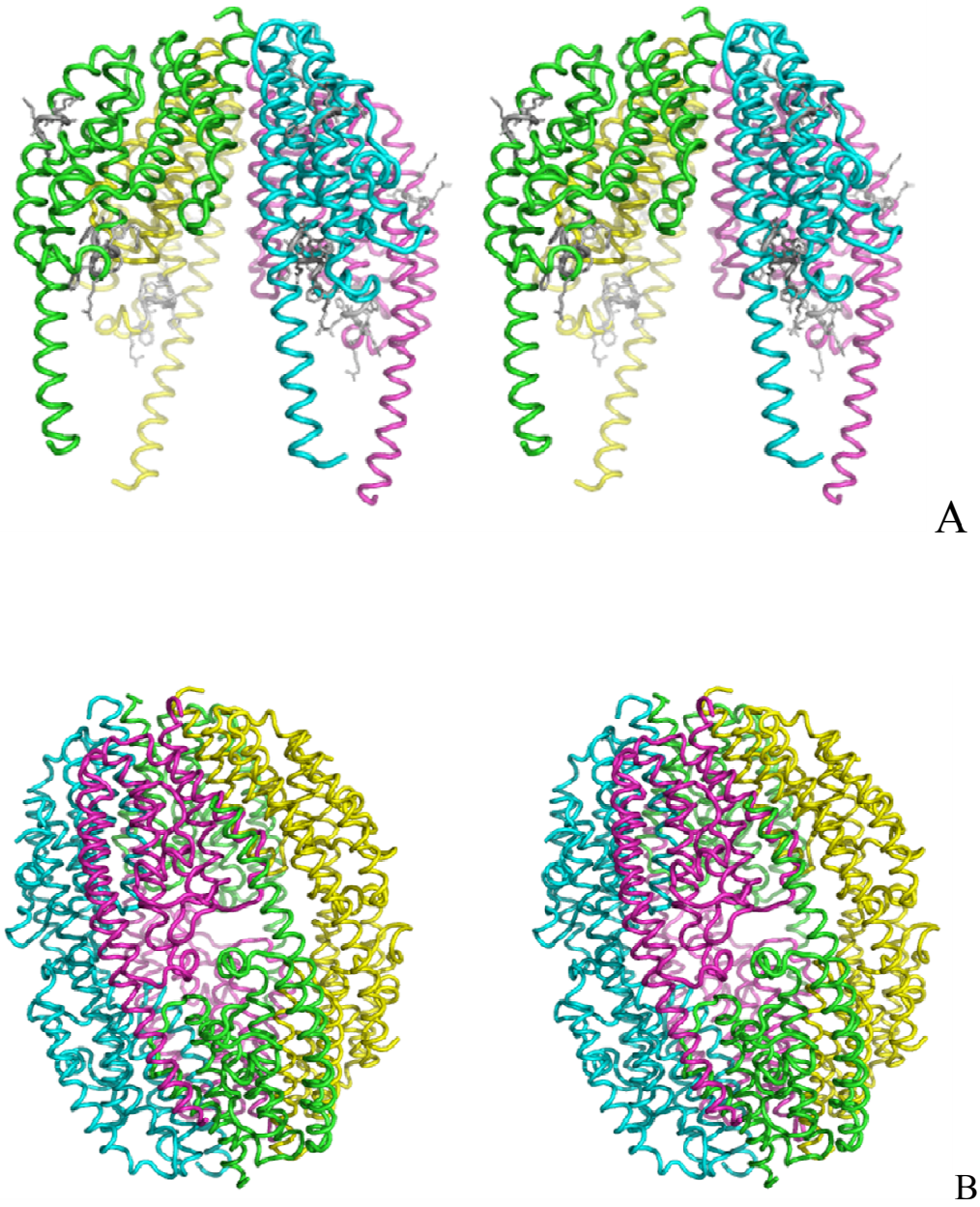


Figure 1.2 Cartoon representation of human translin structure (PDB ID, 1J1J, Sugiura et al., 2004). A) Tetramer of human translin observed in asymmetric unit. Side chains of basic-1 and basic-2 region residues are displayed. B) Two such tetramers form a closed octamer with two fold symmetry. The basic-2 region residues are disposed on the inner surface of the octamer. Figure was prepared with Pymol suite (Delano, 2002).

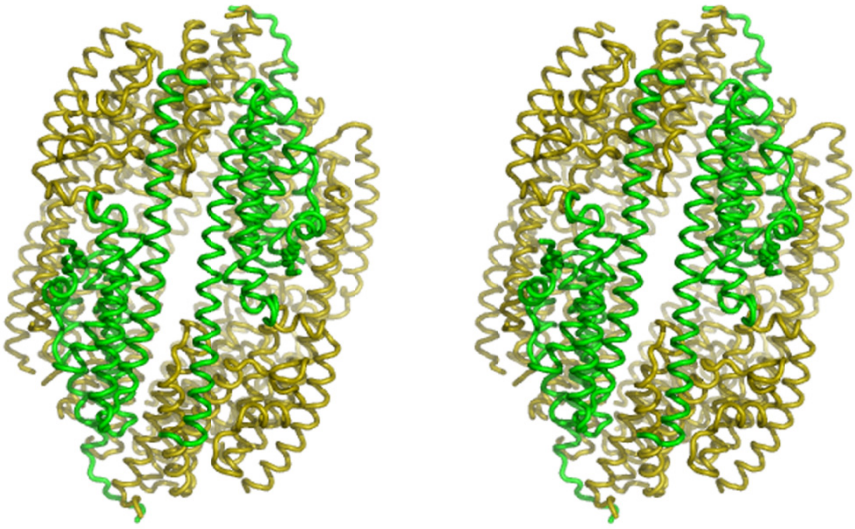


Figure 1.3 Cartoon representation of heteromeric human translin-trax complex structure (PDB ID, 3PJA; Ye et al., 2011). The heteromeric translin-trax complex is composed with 6 translin monomers (olive-orange colour) and two trax monomers (green colour) and forms an octamer similar to human translin. Figure was prepared with Pymol suite (Delano, 2002).

SECTION-II

Overview of protein X-ray crystallography

Knowing the three-dimensional structure of a protein is the important aspect of understanding how the protein functions, even inside the living cell. In a globular protein, different segments of polypeptide chain fold to generate a compact and unique structure. The function of a protein depends on this three dimensional arrangement of polypeptide chain, called the three-dimensional structure. Thus understanding of three dimensional structure of a protein helps to unravel the functional mechanism of a protein. Single crystal X-ray crystallography, Nuclear magnetic resonance (NMR) and Electron microscopy (EM) techniques are the most used methods to elucidate the three dimensional structure of biological macromolecules. However, X-ray crystallography of biological molecules (protein crystallography) has provided the maximum number of available structures (Figure 1.4). Here I will briefly introduce the theory and experimental techniques used in protein crystallography and discuss the recent developments.

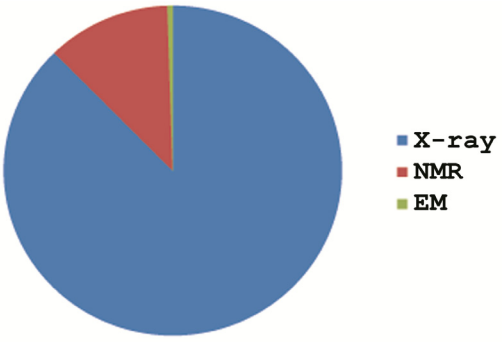


Figure 1.4 Content of Protein Data Bank (PDB) by experimental method**.**

Theory of X-ray diffraction

When X-rays are bombarded on the matter, it causes the electrons to oscillate at their position due to its oscillating electric field. These oscillating charge particles behave like scatterers and emit radiations of same frequency in all the directions, and the phenomenon is called scattering. X-rays scattered from different electrons travel different distances and thus differ in their phases. The total scattering in a given

direction is the vector sum of the waves coming from all electrons and the amplitude of the resulting wave depends on their relative phases and can be written as

$$\mathbf{F(s)} = \int_{space} \rho(\mathbf{r}) \exp(2\pi i\, \mathbf{s.r})\, d\mathbf{r}$$

Where $\rho(\mathbf{r})$ is the electron density at position **r**, exponential term is their relative phase and **s** is called the scattering vector. In a crystal, huge numbers of molecules are arranged in the same orientation and smallest repetition unit is called unit cell (with **a, b, c** axis vectors). Unit cells are related by a translation in three dimensions. The repeating arrangement causes the scattered waves to add up in phase in some directions and to cancel out in a lot other directions. The conditions for diffraction, also known as Laue's conditions, are:

$$\mathbf{a.s} = h,\ \mathbf{b.s} = k \text{ and } \mathbf{c.s} = l$$

Here, (*h k l*) are whole numbers, called miller indices. The Laue's conditions are similar to Bragg's condition and can be used to derive the **Bragg's equation** [$2d \sin\theta = n\lambda$; here d is interplaner distance and equals to $1/|\mathbf{s}|$].

Because of this, we see discrete spots in the diffraction image and each spot or reflection can be defined with unique indices (*h k l*). The structure factor can now be defined over a single unit cell instead of entire space.

$$\mathbf{F(h)} = \int_{cell} \rho(\mathbf{x}) \exp(2\pi i\, \mathbf{h.x})\, d\mathbf{x}$$

The expression interprets that structure factors are the Fourier transform of electron density, and inverse Fourier transform of structure factors can be used to find the electron density, $\rho(\mathbf{x})$, in a unit cell, with the complete knowledge of **F(h)**

$$\rho(\mathbf{x}) = 1/V \sum_{h} \mathbf{F(h)} \exp(-2\pi i\, \mathbf{h.x})$$

However, we measure the intensity of each reflection during the diffraction experiments, that is proportional to the square of the amplitude of structure factor, F(h). The phase information, required to solve the electron density equation, is thus not retrieved in a typical X-ray diffraction experiment. This is classically known as '**phase problem**'.

SOLUTIONS OF PHASE PROBLEM

The phases of the scattered waves, not recovered in the typical diffraction experiments, are generally obtained by using relationships between the amplitudes (direct methods), or specifically in protein crystallography, by the use of additional experiments or known information.

Isomorphous Replacement method

In isomorphous replacement method, phase information is deduced from the differences in intensity between corresponding reflections from native crystal and its isomorphous heavy atom derivative crystals (Perutz, 1956; Kendrew et al., 1958). Introduction of heavy atom perturbs the structure factors significantly due to its large atomic number and the observed differences in diffraction intensities can be used to compute the position of heavy atom using Patterson maps and thus their contribution to the structure factors ($\mathbf{F}_H$). This allows us to deduce some possible values for the protein phase angles, as follows. The structure factor for the derivative crystal

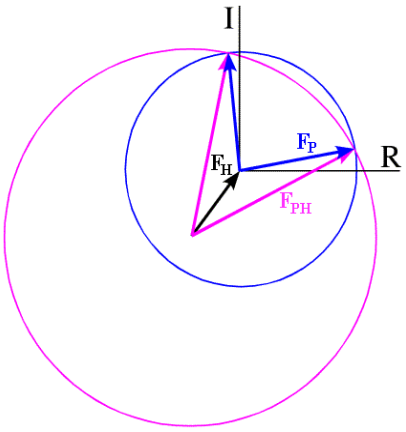


Figure 1.5 Harker construction for SIR

($\mathbf{F}_{PH}$) is equal to the vector sum of the protein structure factor ($\mathbf{F}_P$) and the heavy atom structure factor ($\mathbf{F}_H$), [$\mathbf{F}_{PH} = \mathbf{F}_P + \mathbf{F}_H$].

This equation gives two possible phases for $\mathbf{F_P}$ as illustrated using Harker construction in Figure 1.5. However, due to errors in measurement of structure factor amplitudes, it is difficult to get such clear picture of Harker construction. To get better estimate of phase angle, phase probability for 0^o to 360^o degree (say, in 10^o interval) is calculated and centroid of the curve is taken as best phase, α_{best} (Blow & Crick, 1959; Hendrickson & Lattman, 1970). The electron density is calculated with figure of merit (m, cosine of the phase error) weighted electron density maps, calculated with obtained best phases α_{best} and experimentally estimated amplitudes. If the phases were calculated with single derivative, the method is called Single Isomorphous Replacement (**SIR**). The phase ambiguity can be removed by the use of more than one heavy atom derivatives. The overall phase probability is calculated as the multiplication of individual phase probabilities and the method is known as Multiple Isomorphous Replacement (**MIR)**.

Anomalous scattering methods

In the calculation of atomic scattering factor, the scattering electrons are assumed as free electrons. However, when energy of the bombarded X-rays matches with the absorption edge of an atom (with the natural frequency of electron), electrons behave similar to driven oscillator at resonance and scatter anomalously. The contribution of imaginary part in the structure factor becomes significant. As a consequence, the $\mathbf{F_H}$ *(h k l)* and $\mathbf{F_H}$ *(-h -k -l)*, called Friedel pair, will have different phase angles. This leads to anomalous difference in the amplitudes of Friedel pairs of F_{PH} [F_{PH} *(h k l)* and F_{PH} *(-h -k -l)*] and this difference can be used to derive the phase probability. This

phase information combined with single or multiple isomorphous replacement experiments, called **SIRAS** or **MIRAS** respectively, is useful to break the phase ambiguity.

If a protein has anomalous scatterer in its molecule, multi-wavelength anomalous diffraction (**MAD**) can be used to get the phase information (Hendrickson et al., 1990). In MAD method, the data are collected at several wavelengths; at the absorption, f'', peak (λ_1), at the dispersive term, f', peak (λ_2), and at remote wavelength (λ_3 and/or λ_4). Usually, methionine residues are replaced with Selenomethionine during the cell-culture growth using Met^- strain of *E coli*. These Se-derivative crystals are used for MAD experiments. In **SAD** experiments, diffraction intensity data are collected at a single wavelength usually at the absorption edge (λ_1) (Dodson, 2003).

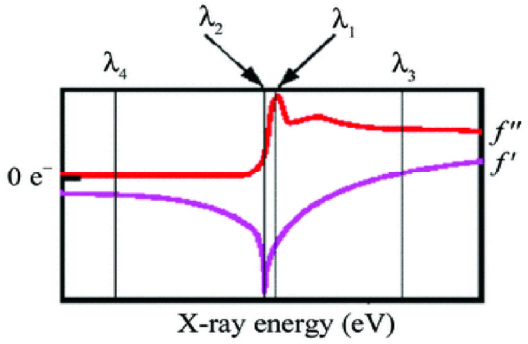


Figure 1.6 Typical absorption curve for an anomalous scatterer. Figure produced from Taylor (2003)

Molecular Replacement method

Molecular replacement method can be used to get initial phases, if structure of a homologous protein is available (Rossmann & Blow, 1962). The molecular replacement is carried out in two steps; first the model is reoriented in the new unit cell, called rotation function, and then the correctly oriented model is translated relative to the origin of the new unit cell, called translation function. Since there are a limited number of folds found in naturally occurring proteins, this method will have wide role to resolve new structures.

Phase improvement by density modification

The experimental phases obtained from isomorphous replacement or anomalous scattering methods, are improved by density modification methods. These methods work by suppressing the noise in electron density maps. In solvent flattening method, solvent region is identified by defining molecular surfaces (based on sigma cut-off in electron density maps) and a mean value of electron density is assigned to solvent region. In another approach, electron density is reassigned within protein volume to match with the observed pattern (in well resolved structures) called 'Histogram matching'. Use of non-crystallographic symmetry (NCS), can also improve the initial phases dramatically if more than one identical subunits are present within asymmetric unit. In this method, the electron density for all identical molecules is forced to be equal to an average value calculated from the electron density of all related molecules. Improved phases are estimated from these modified maps and are used to calculate electron density by combining with experimental amplitudes.

Direct methods

Small molecule structures can easily be solved by the *ab initio* methods, called direct methods, without a prior information of phases. These methods are based on dual space iteration methods and deduce the phase information with bootstrapping procedure from a random set of atoms. The method has even been used for small proteins provided native dataset is available to atomic resolution (better to 1.2 Å) (Sheldrick, 1990; Morris & Bricogne, 2003).

Direct methods are used more frequently to find the heavy atom substructures like Selenium atoms in MAD experiments using *Shake-and-Bake* (Miller et al., 1994)

and *SHELXD* (Schneider & Sheldrick, 2002). These methods have been employed to search upto 70 Selenium atoms of a protein structure, where Patterson maps are very noisy and un-interpretable (Uson & Sheldrick, 1999).

MODEL BUILDING AND REFINEMENT

Refinement is the optimization of a function of a set of observations by changing the parameters of a model (Tronrud, 2004). The observations include the experimentally determined structure factor amplitudes and stereochemical restraints like bond length, bond angles etc. The model parameters include the position of each atom (*x, y, z*), a 'B' factor (a measure of atomic displacement) and occupancy. However, if resolution of data is not very high, overall B-value of model is often refined. The refinement is achieved by either least-square minimization or maximum likelihood or energy minimization based methods. Once the initial phases are refined, electron density maps can be calculated. The process of defining of atomic positions, including tracing the polypeptide chain or 'rotamer' selection etc, in the electron density map is called the model fitting.

Automated model building in protein crystallography

In this era of automation, interpretation of electron density maps is also becoming automated. The automation is based on pattern recognition in the electron density maps using sophisticated computational tools like artificial neural network and maximum likelihood (Morris et al., 2004). The modern model building programs like *FFFEAR* (Cowtan, 1998) and *DADI* (Diller et al., 1999) employ the patterns search from a subset of secondary structure elements and frequently observed domains while programs like *ARP/wARP* (Perrakis et al., 1999) use the syntactic approach of pattern

recognition and attempt to identify the atoms. All these programs work well with the atomic resolution data sets, however, require lot more sophistication for fully automation of model building.

Quality of experimental data

In a diffraction experiment, we measure the integrated intensities spread over few pixels, for each recorded reflection. All reflections, usually measured multiple times (multiplicity or redundancy), are scaled and merged to get a unique set of reflection intensity. The quality of the experimental data is determined by multiple parameters, like R_{merge} (a measure of mismatch between equivalent reflections), I/σ_I (signal to noise ratio), completeness of the data set, multiplicity and, importantly, the resolution of diffraction data.

Quality of Structural model

Cross-validation

At every step of structure refinement, we validate our model to keep a check on errors. Because of unfavourable observation to parameter ratio, any refinement program can overfit the model into the data thus lowering the R-factor, disagreement between model and observed data. In cross-validation, we mark a subset of diffraction data (usually 5–10%) as test data set. This test data set is used to compute R_{free} in the same conventional way, but without being used in the refinement procedure (Brunger, 1997). For a reasonable model, difference between R_{work} and R_{free} should not be more than 5–6%. A cross-validation in real space, can be done with OMIT map method, where phases are calculated with partial model. For a correct model, an unambiguous electron density would appear for the omitted part of the model (Bhat, 1988).

Stereochemistry

The stereochemical parameters of the model like bond lengths, bond angles, dihedral angles, planarity, chiral volumes and non-covalent interactions (like Vander Waals or hydrogen bond interactions) are compared with the standard set of parameters derived from well-resolved protein structures and small molecules. A large deviation from the standard values indicates errors in the model. These parameters can be checked by programs like MolProbity (Chen et al., 2010), Procheck (Laskowski et al., 1993) and WhatCheck (Hooft et al., 1996).

Another validation tool known as 3D-1D profile method is also very useful, especially in cases where sequence is out of register due to error in tracing the main chain (Wilmanns & Eisenberg, 1993). The method is based on assumption that every residue has environment preferences, like hydrophobic side chains tend to be buried in protein structure and surrounded by other hydrophobic residues. Similarly, hydrophilic residues tend to be surrounded by polar residues.

Ramachandran plot

Main chain torsion angles (Φ and Ψ) are strictly constrained due to steric hindrance. The stereochemistry of main chain torsion angles can be validated by the Ramachandran plot in which dihedral angles are plotted for each residue in a square matrix [Φ, -180 to 180; Ψ, -180 to 180]. Favourable, allowed and disallowed regions are enclosed in boundaries. A well refined structure should have torsion angle in the favourable region for most of the residues (usually >90%). Residues in the disallowed region are energetically unfavourable and should be rechecked for errors.

RECENT ADVANCES IN PROTEIN CRYSTALLOGRAPHY

Technological advances

Last decade has observed an enormous growth in the field of protein crystallography (Figure 1.7). A significant contribution to this trend has been due to advances in technology. Use of synchrotron radiation sources with high intensity X-rays and precise wavelength tunability (a requirement of MAD phasing) has contributed significantly towards growth in resolution of protein structures. High brilliance beam using insertion devices (wigglers and undulators) in synchrotron radiation sources made it possible to reduce the beam size to micron dimensions, thus facilitating diffraction experiments with crystals as small as 10–20 μm (termed as protein micro-crystallography). The use of high resolution CCD detectors, with readout time of less than 1 sec and pixel size in the range of 10–50 μm, coupled with high intensity sources have reduced the data acquisition time to a few minutes with superior spatial resolution. The diffraction data acquisition under cryo-conditions helps to collect the complete diffraction data with a single crystal. Also cryo-preservation and cryo-shipment of crystals, crystal handling with robotics and remote data collection facilities are useful technical advancements leading to the rapid growth of this field. Availability of high computational power machines, near-to-automation phase problem solution and model building tools has encouraged more

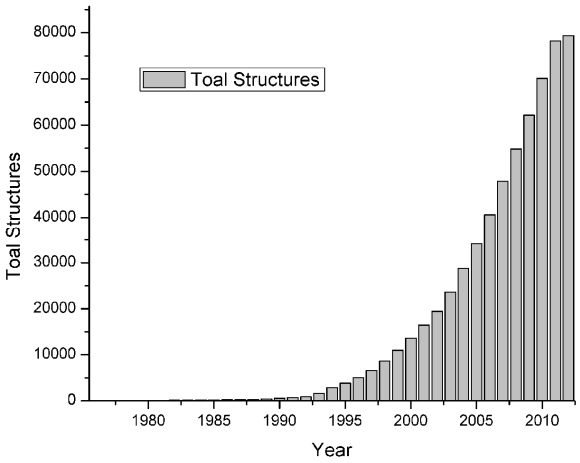


Figure 1.7 Year wise growth of Protein Data Bank (PDB)

researchers to enter the field of crystallography with more challenging problem that seemed intractable earlier.

Structure solution using theoretical models

It is usually assumed that if a homologous structure with better than 30% sequence identity is available, a molecular replacement method can be employed to get the initial phases. Intriguingly, researchers recently have successfully used Rosetta based fold-prediction methods, despite very low sequence identity, to get an initial model to be used for molecular replacement methods (Das & Baker, 2009). The advancement in fold-prediction algorithms may thus reduce in future dependence on experimental phasing methods.

Structure of large macromolecular assemblies

Most of the cellular processes are performed by large macromolecular assemblies. The experimental structural of these supramolecular assemblies will provide a mechanistic understanding of these cellular processes. To reproduce such biological assemblies *in vitro*, is a difficult task. Also, crystals of such large assemblies often diffract weakly (resolution worse than 4 Å). New approaches to work with low resolution data are required. Schröder et al (2010) have shown that a structure of low resolution data even can have quality similar to high resolution structure if high resolution structures of individual subunits are available. A combined approach using electron microscopy image reconstruction methods coupled with available high resolution subunit structure can also be helpful in understanding the cellular process at atomic level (Trabuco et al., 2008). The protein purification and crystallization are still the bottlenecks for structure solution of large macromolecular assemblies, though

advances in molecular biology, high throughput protein purification, robotics in crystallization with drop volumes as small as few nanolitres have contributed significantly in resolving structures of supramolecular assemblies.

Aim of the thesis

The present thesis aims to understand structure-function of translin proteins using structural, bioinformatics, mutational and biochemical analyses. The brief outlines of the studies conducted are:

- Three-dimensional structure of DNA binding incompetent drosophila mutant translin was elucidated using single crystal X-ray diffraction methods.
- Role of C-terminal residues in translin oligomeric status was established using chimera (fusion) proteins synthesized by genetic engineering methods.
- Translin-trax complexes were characterized with the aim to understand the DNA binding properties of translin and trax proteins, and to understand the mechanism of formation of the complex.
- Evolutionary linkage between translin and trax was established using bioinformatics methods.

Chapter 2

Materials and Methods

The details of 'Materials and Methods' used in the present investigations are described in this chapter. The manufacturers and suppliers of enzymes, chemicals and other materials used in the present study are given here. The methods and protocols are grouped in two sections namely, molecular biology, and protein chemistry and crystallization. In molecular biology section, protocols for cloning, expression and mutagenesis are elaborated. Protein chemistry and crystallization section includes description of protocols for protein expression, purification, and protein estimation and detection. A small note on protein crystallization has also been given in this section.

MATERIALS

Materials- dNTPs, T4 DNA ligase, rapid ligation kit and protease inhibitor cocktail tablets were obtained from Roche (Germany), and restriction enzymes and T4 polynucleotide kinase were from New England BioLabs (Hitchin, UK). QuickChange site-directed mutagenesis kit, *PfuTurbo* DNA polymerase and *E. coli* XL1-Blue cells (*E. coli recA1 endA1 gyrA96 relA1 thi-1 supE44 hsdR17 lac*$^-$ F' [*proAB lacI*^{q} *lacZ ΔM15* Tn*10 (*Tet^r)]) were obtained from Stratagene (La Jolla, USA). Plasmids *p*QE9 and *p*QE30, and *E. coli* DH5α cells [*E. coli recA1 endA1 gyrA96 relA1 thi-1 supE44 hsdR17 ΔlacU169* (*Φ80 lacZΔM15*)] were obtained from Qiagen (USA), *E. coli* BL21(DE3) cells [*E. coli* B F$^-$ *dcmompThsdS* (r_B^-, m_B^-) *galλ*(DE3)], and plasmids *p*ET21a and *p*ET28a were obtained from Novagen (Madison, USA). *Taq* DNA Polymerase, DNA molecular weight standards and protein molecular weight markers were obtained from Bangalore Genei (India). Chromatography media Ni-IDA matrix, Superdex^{TM}200, phenyl sepharose and thrombin protease were from GE-healthcare (USA) and pre-packed econo-columns like Q-sapharose and S-sapharase were obtained from Bio-Rad

(USA). Oligonucleotides for cloning, ssDNA sequence (Bcl-CL1), [α-^{32}P]GTP and [γ-^{32}P]ATP were synthesized at BRIT (Navi Mumbai, India). Other fine chemicals were procured from SRL (Mumbai, India).

Composition of reagents

All reagents were prepared in double-distilled water or Milli-Q water as required. The buffers and reagents were prepared as described in the Molecular Cloning: A Laboratory-Manual (Sambrook & Russell, 2001). The reagents were either autoclaved or prepared in autoclaved distilled water. All pH buffers were prepared as 1 M stock solution in autoclaved DW and their pH was adjusted as described in Molecular Cloning: A Laboratory Manual (Sambrook & Russell, 2001). The buffers were sterilized with 0.22 μm syringe filters.

Reagents for plasmid preparation

TEG or Solution-I (for alkaline lysis):

TEG was prepared by mixing the required constituents in DW at a final concentration of 25 mM Tris-HCl pH 8.0, 10 mM EDTA pH 8.0 and 50 mM glucose. It was autoclaved and stored at 4^{o}C.

Solution-II (for alkaline lysis)

This solution contained 0.2 N NaOH and 1% SDS (prepared fresh before use).

Solution III (for alkaline lysis)

To 28.5 mL of DW, 60 mL of 5 M potassium acetate and 11.5 mL of glacial acetic acid (GAA) were added.

TE (Tris-HCl EDTA buffer)

TE was prepared by mixing the required solutions at a final concentration of 10 mM Tris-HCl pH 8.0, 1 mM EDTA. The solution was autoclaved and stored at RT.

50X TAE (Tris-acetate-EDTA)

Tris base, 242 g; glacial acetic acid, 57.1 mL; and 0.5 M EDTA pH 8.0, 100 mL were added to DW to final volume of 1.0 L. The sock solution was sterilized by autoclaving and stored at RT.

6X gel-loading dye

Bromophenol blue, 0.25%; xylene cyanol FF, 0.25%; ficoll 400, 15% were added to DW to prepare concentrated 6X gel-loading dye. The dye solution was stored at 4^{o}C.

5X TBE (Tris-borate-EDTA)

Tris base, 54 g; boric acid, 27.5 g and 0.5 M EDTA pH 8.0, 20 mL were added to DW to final volume of 1.0 L. The sock solution was sterilized by autoclaving and stored at RT.

Reagents for SDS-PAGE

30% acrylamide (29.2% acrylamide and 0.8% bis-acrylamide) solution, 1.5 M Tris-HCl pH 8.8, 1.0 M Tris-HCl pH 6.8, 10X Tris-glycine buffer (tank buffer) were prepared as per the standard procedure given in the Molecular Cloning: A Laboratory Manual (Sambrook & Russell, 2001).

Separating gel 12%

4.0 mL of 30% acrylamide solution, 2.5 mL of 1.5 M Tris-HCl pH 8.8, 100 µL of 10% SDS and 100 µL of APS were added to 3.3 mL of DW. The contents were mixed well

and 8 μL of TEMED was added.

Stacking gel 5%

1.67 mL of 30% acrylamide solution, 1.25 mL of 1.0 M Tris-HCl pH 6.8, 100 μL of 10% SDS and 100 μL of 10% APS were added to 6.8 mL of DW. The contents were mixed well and 20 μL of TEMED was added.

2X SDS gel-loading buffer

2X SDS gel-loading buffers was prepared by mixing the required constituents at a final concentration of 125 mM Tris-HCl pH 6.8, 4% SDS, 5% β-mercaptoethanol, 0.02% bromophenol blue and 10% glycerol, and stored at 4^{o}C.

Fixing and destaining solution

40% methanol, 10% glacial acetic acid.

Coomassie Blue Staining solution

40% methanol, 10% glacial acetic acid, 0.25% coomassie blue R-250.

Culture media

To prepare the Luria-Bertani (LB) media, 10.0 g of sodium chloride, 5.0 g of bacto-yeast extract and 10.0 g of bacto-tryptone were dissolved in DW to final volume of one litre and pH was adjusted to 7.4 with NaOH. All media were sterilized by autoclaving at 15 lb/inch^2 for 15 min and stored at RT. To prepare solid media (LB-agar), 1.5% agar-agar was added to LB media.

SECTION-I

In the present study, molecular biology techniques were used extensively. *Translin* and *trax* cDNA were cloned into bacterial high expression systems using recombinant DNA technology for large scale purification of translin/trax proteins or their complexes. Mutants of functionally important residues of translin and trax were also synthesized using site directed mutagenesis techniques. Fusion genes were synthesized to express chimera proteins and truncated proteins with deleted loop regions. Here, I briefly present molecular biology methods used in the present work.

Plasmid preparation by alkaline lysis method

During the present studies plasmids were used as the expression vectors and were procured from Novagen (*p*ET21a and *p*ET28a) and Qiagen (*p*QE9 and *p*QE30). Plasmids are small, circular, extra-chromosomal elements present in bacteria that are genetically engineered and widely used as cloning and expression vehicles. We used alkaline lysis method to purify plasmids from the bacterial host cells (usually XL1-Blue or DH5α strain of *E. coli*). The protocol is described here briefly:

1.5 mL of saturated culture (grown O/N in LB media containing an appropriate antibiotic) was spun in a micro-centrifuge tube. The harvested cell pellet was resuspended in 100 µL of solution I (TEG) by vortexing and was left on ice for 5 min. 200 µL of freshly prepared solution II (0.2 N NaOH, 1% SDS) was added, mixed gently by inverting the tube several times and left for 5 min at RT. Subsequently, 150 µL of chilled solution III was added to the mixture that was mixed gently by inverting the tube several times and was left on ice for 15-20 min. After centrifugation at 12000 rpm for 10 min, the supernatant was transferred to a fresh tube. The supernatant was extracted

once with equal volume of phenol (saturated with Tris-HCl pH 8.0) and chloroform [chloroform:isoamylalcohol mixture (24:1)] mixture of 1:1 ratio, followed by extraction with chloroform. To precipitate the plasmid DNA, 2 volumes of distilled ethanol was added to the aqueous phase, mixed by inverting the tube several times and was left at 4^{o}C for 30 min. The plasmid DNA was recovered by centrifugation for 15 min at 12000 rpm. The pellet was washed with 70% ethanol, dried at 37^{o}C for 30 min, resuspended in 50 µL of TE containing 20 mg/mL of RNaseA, incubated at 37^{o}C for 1 hr and stored at -20^{o}C.

Purification of DNA using phenol:chloroform extraction method

Tris-EDTA (TE) buffer was added to the DNA solution to be purified (usually 100 µL of the PCR or restriction digestion product) to a final volume of 400 µL. Equal volume of phenol (saturated with Tris-HCl pH 8.0) and chloroform mixture was added and vortexed for 1 min before centrifuging at 12000 rpm for 5 min. Extraction was repeated with equal volume of chloroform. The aqueous phase was transferred to a fresh tube and the DNA was precipitated with 2 volumes of ethanol in the presence of 0.3 M sodium acetate pH 5.2. The pellet was washed with 70% ethanol and dried at RT, and was resuspended in small volume of TE buffer.

Qualitative and quantitative detection of DNA

Purification, amplification and restriction digestion of plasmid DNA and DNA fragments, during cloning experiments, were monitored using agarose gel electrophoresis. Typically 1-1.5% agarose gels were prepared in 1X TAE buffer with 0.5 µg/mL concentration of ethidium bromide. Electrophoresis was carried out in

horizontal electrophoresis unit at 5-10 V/cm. Gels were visualized on an ultraviolet (UV) transilluminator. The quantity of the DNA was estimated approximately from the band intensity.

DNA purification using agarose gel electrophoresis method

DNA purification using agarose gel involved electrophoresis of DNA on a low melting point (LMP) agarose gel and excision of the gel slice containing the DNA fragment of interest. To extract DNA from gel slice, the gel slice was melted in a sterile micro-centrifuge tube at 68^{o}C for 10 minutes with 4 volumes of TE buffer. It was allowed to cool to 42^{o}C and then digested with 1.0 unit of β-agarase per 100 μL of molten agarose at 42^{o}C for 2 - 3 hrs (or till complete digestion). The DNA was then extracted from the solution using phenol:chloroform extraction method.

Cloning of *translin* and *trax* genes in high expression vector systems

Primer design

Primers are short ssDNA sequences that are required to amplify a target sequence from the template DNA. Primer sequences were designed from the target nucleotide sequences and were validated for their secondary structure and melting temperature (T_m) using PRIMER program.

Polymerase Chan Reaction (PCR)

This technique involves highly specific enzymatic amplification of a DNA sequence *in vitro* defined by two flanking primers. The amplification occurs in three major steps. During *initial* ***denaturation*** the double stranded DNA template is separated into single strands by heating at 95^{o}C. In the next step of ***annealing*** the primers anneal to the

template and then ***primer extension*** takes place. All the three steps are carried out at three different temperatures, optimum for that particular reaction, in the presence of a thermostable DNA polymerase and dNTPs, resulting in the amplification of the target product.

During the present work, typically, for a 100 µL PCR reaction mix, 10 µL of 10X PCR reaction buffer, 2 µL of 10 mM dNTP mix (200 µM), 2 µL (200 nM) of each primer, 1 µL (10-20 ng/~10^8 copies) of template DNA and 1 µL (2.5 units) of DNA polymerase were added to sterile Milli-Q water to a final volume of 100 µL. The DNA polymerase was added in the last. The following thermal cycle was used for amplification of required DNA template.

		Temperature ^oC	Time	Number of Cycle/s
1.	Initial Denaturation	94^{o}C	2 minutes	1
2.	Denaturation	94^{o}C	50 seconds	35 Cycles
3.	Primer Annealing	58^{o}C*	1 minutes	
4.	Primer Extension	72^{o}C	1 minutes#	
5.	Final Extension	72^{o}C	10 minutes	1

*Annealing temperature was varied from 54^{o}C-60^{o}C depending upon primer-template complementarity and GC content

extension time also varied based on the length of amplified product. Typically, 1 min extension time per 1000 bases was used in the reaction.

The amplified PCR products were analyzed on 1.5% agarose gel and purified with phenol:chloroform extraction followed by alcohol precipitation.

Restriction digestion

A restriction enzyme cuts double-stranded DNA at very specific short sequences, known as restriction sites. To clone a gene fragment into the desired plasmids, both plasmid

DNA and gene insert were digested with the restriction enzymes. The buffer conditions used for all restriction digestions were essentially those that were recommended by the manufacturers. Typically, 20-30 units of restriction enzyme were added to the 50 µL reaction mixture (500-1000 ng of DNA) followed by three to five hour incubation at 37^{o}C. For double digestion of DNA template with *NdeI* and *BamHI*, both the enzymes were added simultaneously in the *BamHI* enzyme specific buffer. However, in case of double digestion of DNA template with *NheI* and *BamHI*, both the reactions were performed in sequential manner. The DNA was purified by phenol:chloroform extraction methods followed by alcohol precipitation.

DNA ligation

T4 DNA Ligase catalyzes the formation of a phosphodiester bond between juxtaposed 5'-phosphate and 3'-hydroxyl termini in duplex DNA or RNA. In recombinant DNA technology, ligase enzyme is used to join DNA fragments from different sources to prepare recombinant DNA. Rapid ligation kit (Roche) was used for ligation reaction. Typically, vector (4000-5000 bases) and insert (~1000 bases) DNA fragments were added in 1:5 molar ratios (100 ng of insert and 100 ng of plasmid) in a total 20 µL reaction volume in the buffer conditions recommended by the manufacturer. The ligation reaction mixture was maintained at 25^{o}C for 5-10 min. 2-5 µL of ligation mixture was used for transformation in cloning host cells.

Transformation

Transformation is a process in which a recipient cell acquires free DNA molecules from surrounding medium. Some cells like *Salmonella typhimurium* have natural competence to acquire DNA molecules while many cells become competent under stress conditions.

During the course of this study, we used $CaCl_2$ method to prepare the competent cells of *E. coli*. The competent cells were prepared freshly before use. An isolated single colony of the host strain was inoculated in 25 mL of LB broth and incubated on shaker at 37^oC till OD_{600} reached to 0.1. The cells were harvested with centrifugation at 4^oC and resuspended in 10 mL of chilled 50 mM $CaCl_2$ solution. The suspended cells were maintained on ice for 30 min. The cells were again centrifuged at 8000 rpm for 10 min at 4^oC, and harvested cells were resuspended in 1 mL of chilled 50 mM $CaCl_2$ solution. The suspended cells were maintained on ice for 2 hrs.

100 µL of the competent cells were taken in a pre-chilled eppendorf and approximately 100 ng of plasmid/ligated DNA were added to the cells and maintained on ice for 10 min. The cells were incubated at 42^oC for 3 min (heat shock method) and immediately transferred on ice. After five min, 1 mL of LB media was added to the cells and incubated at 37^oC for 45 min. Subsequently, cells were harvested and plated out on LB-agar plates having appropriate antibiotics. The plates were incubated at 37^oC overnight. The transformed bacterial cell colonies were selected and streaked on fresh LB-agar plate.

Selection of positive clones

Colony PCR was performed to screen the bacterial colonies that contain the recombinant plasmids. The O/N incubated transformants were transferred to fresh LB-agar plates. Using sterile 0.5-10 µL pipette tips a single colony was touched (~1/10th of colony) and suspended into the reaction mixture. The preparation of PCR reaction mixture (30 µL) and thermal cycling conditions were as described for PCR protocol except for the addition of template.

SITE DIRECTED MUTAGENESIS AND GENE FUSION

The site-directed mutagenesis experiments were carried out using PCR overlap extension method or 'Quick-Change site-directed mutagenesis kit' (Stratagene). Chimeric and truncated genes were synthesized using PCR overlap extension method.

Quick-change site-directed mutagenesis

This is a rapid four step-protocol to generate site specific mutations as explained in Figure 2.1. This method was used to construct the wild-type drosophila translin (using drosophila translin with P168S mutation as template) and mutants of trax having substitution in the putative DNA binding motifs. The template plasmids were purified strictly from DH5α strain of *E. coli.*

Sample reaction was prepared with 10X reaction buffer (5 µL), DNA template (2 µL, 5-50 ng), 2 µL (125 ng) of both the primers, dNTP mix (1 µL) and *PfuTurbo* DNA Polymerase (1 µL). Autoclaved DW was added to make final volume to 50 µL. The following thermal reaction protocol was followed:

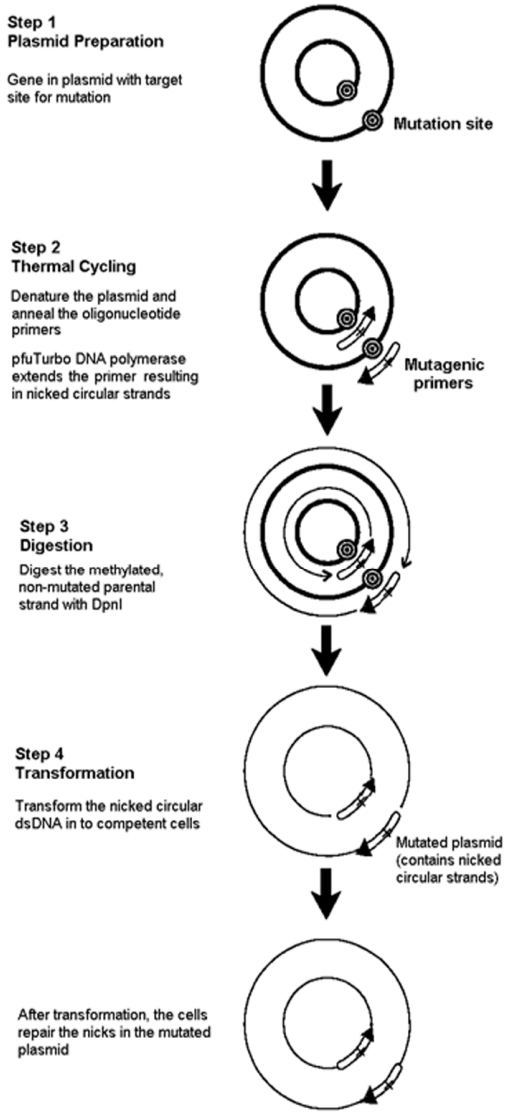


Figure 2.1 Overview of the Quick-change site-directed mutagenesis method *(adapted from Stratagene manual)*

Number of Cycles		Temp ^{O}C	Time
1	Initial Denaturation	95	40 s
25	Denaturation	95	30 s
	Primer Annealing	55	1 min
	Primer Extension	68	6 min 20 s

After completion of thermal cycles, 1 µL (10 U/µL) of *Dpn*I enzyme, an endonuclease (target sequence: 5′-Gm^{6}ATC-3′) specific for methylated and hemimethylated DNA, was added to the reaction tube and incubated at 37°C for 2 hr to digest the methylated parental plasmid. 2-5 µL of the reaction mixture was transformed into competent cells of *E. coli* XL1-Blue strain. The positive clones were screened with DNA sequencing and were transformed into BL21(DE3) for over-expression of mutant proteins.

PCR overlap-extension method

Mutagenesis, using PCR overlap-extension method, is achieved by performing PCR with specially designed oligonucleotide primers that harbor the desired substitutions, insertions or deletions in their sequence. The two overlapping fragments are fused together in a subsequent extension reaction. The inclusion of terminal primers in the extension reaction amplifies the fused product by PCR. During my work presented in the thesis, two separate overlapping DNA fragments were amplified using mutagenizing primers, having an overlap of minimum 30 base-pairs, along with the either side flanking region primers. Both overlapping DNA fragments were purified from gel-excised bands and were extended with overlap extension PCR, without adding primers. After 5 cycles of overlap-extension reaction, the terminal primers were added to PCR reaction to increase the yield. The amplified product was cloned into desired vector as discussed above.

SECTION-II

PROTEIN PURIFICATION FROM HIGH EXPRESSION SYSTEM

Cell culture of bacterial expression system

The *E. coli* BL21(DE3) cells, harboring the desired construct, were grown in LB media in presence of suitable antibiotics (100 mg/L) at 37^{o}C on an orbitek shaker at 200 rpm. The cell density was monitored by measuring the optical density at 600nm wavelength (OD_{600}) with spectrophotometer. The protein expression of recombinant genes, under *lacUV5* control, was induced with IPTG (0.5 mM to 1mM) in late log phase of growth (~0.8 OD_{600}). After 4 hour of induction, cells were harvested with centrifugation at 6000g for 10min. The cell pellet was stored at -20^{o}C till further use. The expressions of the proteins were analyzed by loading the uninduced and induced samples on SDS-PAGE.

Preparation of cell lysate

The cell pellet was resuspended in lysis buffer [25 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM DTT, 0.1 mM EDTA, 5% Glycerol]. Lysozyme (0.3 mg/mL) was added to the cells suspension in presence of protease-inhibitor [one tablet of protease-inhibitor cocktail (Roche) per 100 mL of cell suspension] and maintained on ice for 1 hr. The cell suspension was further lysed with ultra-sonication for 10 minutes on ice in pulse mode (2 sec on and 5 sec off) using ultrasonic sonicator (Trans-o-sonic, India). The cell lysate was cleared of cellular debris by centrifugation at 20000g for one hr. Both the soluble (supernatant) and insoluble (pellet) fractions were analyzed on SDS-PAGE. All translin proteins and translin-trax complexes were observed in soluble fraction of cell lysate while trax alone was observed in insoluble fractions.

Protein purification

The protein purification was achieved using low-pressure liquid-chromatography system (Bio-Rad Biologic Lp) at room temperature. The elution from the column matrix was monitored online for absorbance at 280nm wavelength. The purification progress was monitored by SDS-PAGE. All the proteins, analyzed/characterized in the present studies, were purified with at least three stage purifications using column chromatography methods. Further, a single major peak was isolated using size-exclusion chromatography (gel-filtration column) before crystallization experiments or DNA binding assays of the proteins. The detailed purification protocols have been mentioned in the respective chapters. All the purified proteins were concentrated to 5-10 mg/mL by ultra-filtration method using Amicon Ultra (Millipore, USA) centrifugal filter devices of 30kDa molecular weight cut-off. The concentrated proteins were stored at 4^{o}C.

Matrix-Assisted Laser Desorption Ionization Time-Of-Flight (MALDI-TOF) Mass spectrometry analysis

The integrity of all the purified proteins was confirmed by molecular weight estimation with MALDI-TOF (Axima-CFR at IIT, Mumbai). The protein samples were diluted to 50-100 pmol/µL concentration in buffer [20 mM Tris-HCl pH 8.0, 25 mM NaCl] and mixed with sinapinic acid matrix. The mixture was spotted on a copper grid and subsequently loaded on the MALDI-TOF for mass analysis.

Estimation of protein concentration

The protein concentrations were estimated by modified Lowry's method (Miller, 1959). The linear working range of the method extends from 5 to 25 µg/mL. Typically, 200 µL

of freshly prepared alkaline copper reagent [1% Copper sulphate, 50 µL; 2% Sodium-Potassium tarterate, 50 µL and 900 µL DW] were added to 200µL of dilute protein sample (5 to 25 µg protein) and kept at room temperature for 10 min. Subsequently, 600 µL of Folin's reagent (1:10 dilution) was added to the mixture and incubated at 50^{o}C for 10 min. The absorbance was monitored at 660 nm wavelength. Bovine serum albumin (BSA) was used as control.

The procedure was repeated for six different (known) concentrations of BSA protein. The standard curve was obtained by plotting the O.D$_{660}$ v/s the known concentration of BSA protein. The concentrations of the unknown proteins were estimated based on their OD$_{660}$ value using the standard curve obtained using known concentration of BSA protein (Figure 2.2).

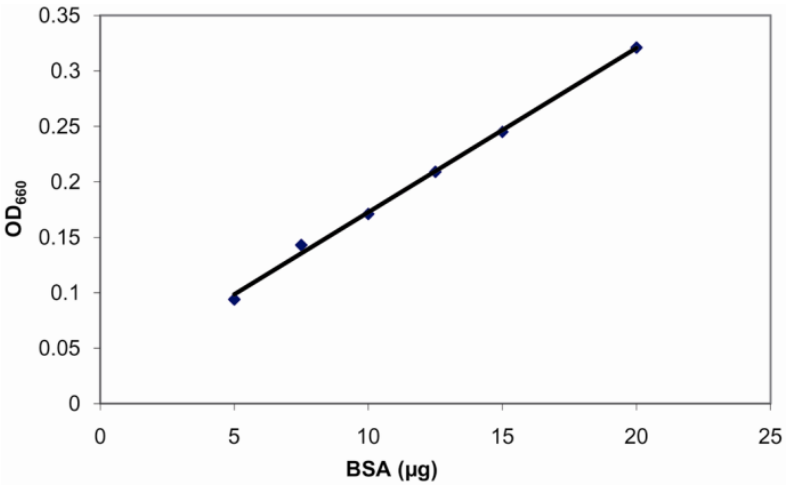


Figure 2.2 Standard curve for protein estimation using Lowry's method.

SDS-PAGE analysis

During the studies presented in the thesis, SDS-PAGE (Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis) was routinely used to monitor proteins during

purification and to estimate the protein subunit molecular weight. The subunit compositions of purified complexes and their relative ratios were also determined based on the relative intensity of the bands. Usually 12% resolving and 5% stacking gels were prepared. Protein samples (10 µL) were mixed with equal volume of SDS sample buffer and were boiled for 5 min. The samples were loaded into the wells of stacking gel and resolved on vertical gel electrophoresis at constant voltage. The protein bands were visualized by Coomassie Blue Staining. The gels were incubated with staining solution (40% methanol, 10% acetic acid, 0.25% coomassie blue R-250) on a gel rocker. The background was cleared with destaining solution (40% methanol, 10% acetic acid).

Silver staining

To detect very small amount of protein (a few nano-grams) gels were stained with silver staining method. Milli Q water was used for all the wash steps and for preparing the reagents. All the steps were performed on the gel-rocker. The gel was incubated in a fixation solution of 50% methanol and 5% acetic acid for 20 min followed by 15 min incubation in 50% methanol. It was subsequently rinsed with water 3 times at intervals of 5 min each. The gel was incubated in freshly prepared sodium thiosulphate solution (0.2 g/L) for 1 min followed by wash with water. The gel was transferred to silver nitrate solution (0.2 g/100 mL) for 30 min. The gel was rinsed briefly with water for 1 min and developed with pre-chilled developing solution (3.0 g sodium carbonate and 25 µL of formaldehyde in 100 mL). The silver staining reaction was terminated with EDTA (14 g/L) solution for 5 min. The gel was subsequently stored in water with 10 % glycerol.

Electrophoresis mobility shift assay (EMSA)

Electrophoresis mobility shift assays (EMSA) were performed to assess the DNA binding activity of translin/translin-trax proteins. For radio-labeled DNA oligos, assays were performed on native-PAGE (Polyacrylamide Gel Electrophoresis under native-conditions). The 4.5% gels were prepared in 0.5X TBE buffer (1.5 mL of 30% acrylamide, 1 mL of 5X TBE, 100 µL of APS and 20 µL of TEMED to final volume of 10 mL). The pre-incubated protein:DNA mixtures were added to 6X gel loading dye and were resolved on vertical gel electrophoresis at constant voltage. The gels were dried under vacuum at 65^{o}C and auto-radiographed on X-ray film (Kodak, India). X-ray film was developed with standard protocol and scanned with gel-documentation system from Syngene (Synoptics, Cambridge, England).

For unlabeled DNA oligos, protein:DNA complex was detected with ethidium bromide staining. For this purpose, 1.5 % agarose gels were prepared in 1X TAE buffer, as mentioned in section-I of this chapter. The protein:DNA mixtures were added with 6X gel loading dye and were resolved on horizontal gel electrophoresis at constant voltage. The intensity of gel shifted bands and residual DNA were measured with UV-transilluminator using gel-documentation system from Syngene (Synoptics, Cambridge, England).

Protein crystallization

In protein X-ray crystallography, the first essential step is to grow diffraction quality crystals. Crystallization is a phase transition, from liquid to solid phase, phenomenon. It is a complex and multi-parametric process and there are no any set rules or 'magic bullet' to produce the crystals. A typical phase diagram has been illustrated here (Figure

2.3). The first step in finding crystallization condition is to set up screening trials, exposing the protein to variety of agents (precipitants, buffers and stabilizers), to find the first lead condition. Commercially available crystal screens are usually good enough for initial screening. There are many methods used in protein crystallization like, batch method, vapour diffusion method, liquid diffusion and dialysis, to bring the protein solution into nucleation zone and then to metastable zone (Figure 2.3). Vapour diffusion methods, hanging drop and sitting drop, are the most widely used methods in which protein solution is equilibrated against the higher precipitant concentration by diffusion of vapours. With the use of robotics in crystallization, micro-batch method is also a method of choice where robots can dispense thousands of microbatch trials in nanolitre volumes. In this method, precipitant reagents are mixed with protein solution and incubated under low density paraffin oil. Seeding methods are also very useful in improving the crystal quality and crystal size. In these methods, pre-formed nuclei of the same protein or the homologous protein are added into the crystallization drop in the meta-stable zone.

During my studies, presented in the thesis, I used hanging drop and sitting drop vapour-diffusion method for protein crystallization. For hanging drop vapour-diffusion experiments, typically, 2 μL of the protein solution and 2 μL of precipitation solution were mixed on cover slip. The thin glass cover slips were siliconized to prevent the spreading of the drop. These cover slips were equilibrated against the 500 μL of the precipitating solution. Initial attempts for crystallization of translin/translin-trax complex proteins were made using commercially available crystallization screens like crystal screen, crystal screen 2 and SaltRx kits from Hampton research and protein-complex crystallization kit from Sigma. The detailed crystallization conditions have

been given in the respective chapters. Sitting drop vapour diffusion method was used to grow big size crystals. Typically, 10-15 µL of protein solution was added to micro-bridges (Hampton research) with equal volume of precipitation solution and were equilibrated against the 500 µL of the precipitating solution. The wells were sealed with cover slips.

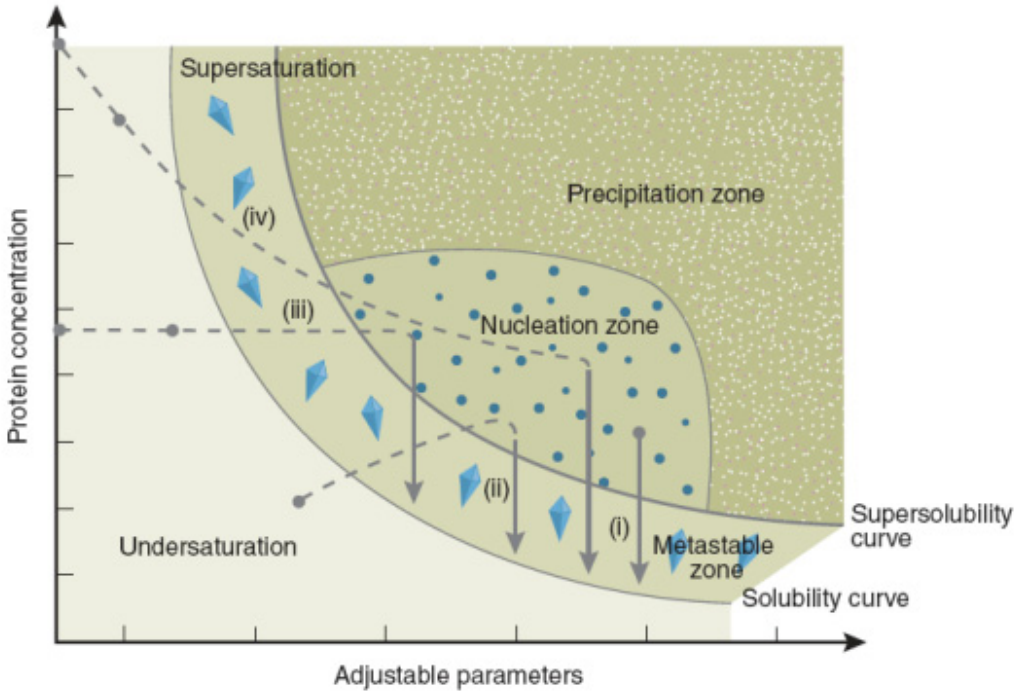


Figure 2.3 Protein crystallization phase diagram. Adjustable parameters include precipitant or additive concentration, pH and temperature. The four major crystallization methods are represented: (i) microbatch, (ii) vapor diffusion, (iii) dialysis and (iv) liquid diffusion. Each involves a different route to reach the nucleation and metastable zone. The Figure has been reproduced from Chayen et al. (2008).

Post crystallization treatments to improve the diffraction quality

Very often, crystals are not very well ordered and are unsuitable for diffraction studies. Recent studies have reported a significant improvement in diffraction properties of the

crystals (as much as 12 Å to 2.6 Å in case of *E. coli* YggV and, from no diffraction to 3.0 Å in *Candida albicans* 3-dehydroquinate dehydratase) by soaking, dehydration, annealing and many more similar approaches, collectively called post crystallization treatments (Heras & Martin, 2005). I attempted various methods to improve the diffraction properties of wild-type drosophila translin crystals as well as drosophila mutant translin crystals. Dehydration of the drosophila mutant translin crystals improved the resolution from 3.9 Å to 3.3 Å. The crystals were transferred to increasing concentration of precipitant in three steps (an increase of 10% in each step) as illustrated by Heras & Martin (2005).

Data processing and structure solutions

Data processing is the way to assign (*h k l*) indices to each measured spot (data indexing) and to estimate the intensity of each spot on absolute scale (integration and scaling). The structure factor amplitude is proportional to the square root of the intensity. As discussed in the introductory chapter, we estimate the amplitude of structure factors from the diffraction intensity experiments. The phase information is retrieved using indirect methods. In the presented work, data processing was performed using *MOSFLM* (Powell, 1999) and *HKL2000* (Otwinowski & Minor, 1997). The molecular replacement method was used to get the initial phases using AMORE (Navaza, 1994) and MOLREP (Vagin & Teplyakov, 1997) programs. The coordinates were subsequently refined by REFMAC5 (Vagin et al., 2004) with intermittent model building using COOT (Emsley et al., 2004). The detailed procedures have been described in the relevant chapters.

Chapter 3

Crystal Structures of the drosophila P168S mutant translin

INTRODUCTION

The human translin, a 27kDa protein, binds to single-stranded DNA (ssDNA) and RNA. It recognizes chromosomal breakpoint sequences at ssDNA ends with varying affinity (Sengupta et al., 2002) and is observed to be recruited to the nucleus concomitant with the induction of double strand breaks by DNA damaging agents and in many cases of lymphoid neoplasms that are robustly associated with T-cell receptor or immunoglobulin gene translocations (Aoki et al., 1995; Kasai et al., 1997). The mouse translin has been proposed to function in regulating the expression of a variety of mRNA sequences by regulating RNA translocations and localization within and perhaps even across cells (Kwon et al., 1991; Han et al., 1995b). It has been proposed that GTP modulates the binding affinity of mouse translin towards RNA or DNA (Chennathukuzhi et al., 2001b). Translin selectivity towards RNA or DNA is also modulated on interactions with another translin-like protein known as trax. The translin protein bears functional nuclear export signal sequence. The trax protein possesses nuclear localization signal sequence and is post-transcriptionally stabilized by translin protein (Aoki et al., 1997; Chennathukuzhi et al., 2001a). The heteromeric translin-trax complex has been reported as a key component of RNA induced silencing complex (RISC) and possesses RNase activity (Liu et al., 2009, Ye et al., 2011). Translin and trax proteins are well conserved in eukaryotes and are found in vertebrates including human, mouse, chicken and Xenopus; in insects like *Drosophila melanogaster* (drosophila), in plants like *Arabidopsis thaliana,* and also in *Schizosaccharomyces pombe* (S. pombe) fungi.

Two basic nucleic acid binding motifs (basic-1 and basic-2) have been identified from the mutational studies on human and mouse translin (Aoki et al., 1999; Chennathukuzhi et al., 2001a). It is surmised that the DNA-binding domain of human translin is formed by the combination of its basic regions in the multimeric structure and

the loss of multimeric structure results in abrogation of its DNA-binding abilities (Aoki et al., 1999). Prior to this work, the crystal structures of human and mouse translin proteins, which share more than 98% sequence identity, were known (Pascal et al., 2002; Sugiura et al., 2004). The octameric structures of human and mouse translin proteins in the crystals, however, are not consistent with the octameric ring structure observed in electron microscopic studies (VanLoock et al., 2001). It is imperative that more structural information on various other members of translin protein family will help resolve the discrepancy.

Drosophila translin shares 50% sequence identity with human and mouse orthologs. During cloning of drosophila *translin* gene, a PCR mediated mutation was inadvertently incorporated in the clone that resulted in proline to serine substitution at position 168 (PCR generated error). The drosophila P168S mutant translin, which was inadvertently reported as wild-type drosophila translin, was DNA-binding incompetent (Sengupta et al., 2006). The proline residue, though conserved very well in eukaryotes, was not thought to participate in nucleic acid binding. *In silico* modeling of the mutant drosophila translin structure, based on available human and mouse translin crystal structures, did not suggest why drosophila translin P168S mutant should lack the affinity towards nucleic acids. The proline (P167 in human translin) was in *trans* conformation and was situated near the N-terminus of helix. Thus proline to serine substitution could have been accommodated in the fold. Since all other known translin homologs were DNA-binding competent, the lack of activity in drosophila mutant translin prompted us to investigate structure-function of drosophila mutant translin.

In this chapter, I have described crystal structure determination of the drosophila P168S mutant translin from two crystal forms. The atomic coordinates of these have been submitted to the Protein Data Bank under the accession codes 2QRX and 2QVA

(Gupta et al., 2008). The tertiary and quaternary structures of the protein in the two crystal forms are identical, but differ markedly from those of human and mouse orthologs. The C-terminal residues were disordered in both the drosophila mutant crystal structures. An important conclusion from the analysis of all the known translin structures could be deduced; up-down dimer is likely to be evolutionary conserved in translin proteins.

METHODS

Cloning and expression of drosophila P168S mutant translin

All basic recombinant DNA procedures were performed as described in chapter 2 (Materials and Methods) of this thesis. The *p*QE30-*dtsn168s* construct, used to express drosophila P168S mutant translin with N-terminal 6XHis tag, was obtained from Prof. B.J. Rao, Tata Institute of fundamental Research (TIFR), Mumbai. The construct was transformed into *E. coli* XL1-Blue for large scale purification of drosophila P168S mutant translin with 6XHis tag at its N-terminus (6XH-dtsnP168S).

The drosophila P168S mutant *translin* was also sub-cloned into *Nde*I/*Bam*HI sites of *p*ET21a expression vector (Novagen, USA) to express the mutant protein without any tag or additional residues. The drosophila *translin* open reading frame (ORF) was PCR amplified with *PfuTurbo* DNA polymerase using primers (forward; CATCACGGA**CATATG**TCGAACTTCGTGAACTT and reverse; ACT**GGATCC**TTATTCGGTTGCAGGAACAGC) and *p*QE30-*dtsn168s* as template. The restriction digestion sites incorporated in primer sequence are shown in bold and underlined text. The resultant *p*ET21a-*dtsn168s* vector was transformed into *E. coli* BL21 (DE3) cells for expression of drosophila P168S mutant translin (dtsnP168S). The nucleotide sequence of *translin* genes of *p*QE30-*dtsn168s* and *p*ET21a-*dtsn168s*

constructs were confirmed by complete sequencing of *translin* ORFs using an automated DNA sequencer.

Purification of drosophila P168S mutant translin protein

The *E. coli* XL1-Blue cells transformed with *p*QE30-*dtsn168s* construct were grown in LB media in presence of ampicillin antibiotic (100 mg/L) at 37^{o}C for expression and purification of drosophila P168S mutant translin with N-terminal histidine tag (6XH-dtsnP168S). The expression of translin mutant protein was induced in late log phase of growth (~0.8 OD_{600}) with 0.5 mM IPTG. After 4 hour of induction, cells were harvested with centrifugation at 6000g for 10 min. The cells were suspended in lysis buffer [25 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM DTT, 5% Glycerol]. Lysozyme (0.3 mg/mL) was added to the cells suspension in presence of protease-inhibitor (one tablet of protease-inhibitor cocktail (Roche) per 100 ml of cell suspension) and kept on ice for 1 hour. The cell suspension was sonicated for 10 minutes on ice in pulse mode (5sec on and 10sec off). The cell lysate was cleared of cellular debris by centrifugation at 20000g for one hour. The protein purification was achieved using low-pressure liquid-chromatography system (BioRad Biologic Lp) at room temperature. The protein elution was monitored online by recording absorbance at 280nm wavelength. The purification progress was monitored by SDS-PAGE. The 6XH-dtsnP168S was purified using immobilized metal-chelating affinity chromatography (IMAC). The metal-chelating affinity matrix (Ni-IDA) was charged with 100 mM $NiSO_4$ after the wash steps. Then the matrix was manually packed into glass column (BioRad, USA) and equilibrated with buffer A1 [25 mM Tris-HCl pH 8.0, 100 mM NaCl and 50 mM imidazole]. The supernatant of cell lysate was loaded onto the Ni-IDA column matrix and washed with buffer A1 till the absorbance at 280nm (OD_{280}) reached to ~0.01. The bound proteins were eluted with a linear gradient of imidazole (50–400 mM) in buffer A1 over six

column volumes. The 6XH-dtsnP168S protein eluted at ~250 mM concentration of imidazole. The protein was further purified with anion exchange chromatography. The protein eluted from Ni-IDA column was dialyzed against the buffer A2 [25 mM Tris-HCl pH 8.0, 100 mM NaCl and 5% glycerol] and loaded onto pre-packed Q-sepharose column (BioRad, USA) pre-equilibrated with buffer A2. After extensive wash with buffer A2, the bound proteins were eluted with a linear gradient of 100–500 mM NaCl in buffer A2. The drosophila mutant translin protein showed a very strong binding towards anion exchanger and eluted at ~350mM of NaCl. The purified protein was dialyzed against storage buffer [25 mM Tris-HCl pH 7.0, 100 mM NaCl] and concentrated to 5 mg/mL using centricorn (Millipore, USA). The protein was stored at 4^{o}C.

E. coli BL21(DE3) cells transformed with *p*ET21a-*dtsn168S* were used for the over-expression of the drosophila P168S mutant translin without any additional residues (here onward will be described as dtsnP168S). The dtsnP168S was purified by conventional chromatography methods. The cell lysate was dialyzed against buffer A2 and purified using Q-sepharose anion exchanger. The bound translin protein was eluted with a linear gradient of 0.1 – 1 M NaCl in buffer A2. The translin protein eluted at ~300 mM of NaCl. The peak fractions were dialyzed against buffer A3 [20 mM sodium acetate, pH 5.6] and loaded onto S-sepharose (BioRad, USA) cation exchanger. The translin protein was observed in the unbound fraction. For further purification, the protein was dialyzed against buffer A4 [1.5 M ammonium sulphate, 0.1 M sodium phosphate, pH 7.0] and loaded on Phenyl-sepharose matrix. The bound proteins were eluted with the reverse gradient of ammonium sulphate (1.5 – 0 M) in buffer A4. The dtsnP168S protein bound strongly with Phenyl-sepharose matrix and eluted at 50 mM ammonium sulphate. The purified protein was dialyzed against buffer C2 [10 mM sodium phosphate buffer pH 7.0,

100 mM NaCl and 5% glycerol] and concentrated to 5 mg/mL using ultra-filtration method. The protein was stored at 4^{o}C.

Crystallization of drosophila P168S mutant translin and data collection

For crystallization trials of drosophila mutant translin, the protein could be concentrated only upto 5 mg/mL by ultra-filtration using centricon-10 (Millipore, USA) as the protein was observed to aggregate at higher concentrations. Initial crystallization attempts were made in two independent sets. In the first set, 6XH-dtsnP168S mutant protein was dialyzed against buffer C1 [10 mM Tris-HCl pH 6.8, 150 mM NaCl and 5% glycerol]. In second set, the mutant translin was dialyzed against buffer C2 [10 mM sodium phosphate buffer pH 7.0, 100 mM NaCl and 5% glycerol]. Initial screening for both set of proteins was performed with formulated crystallization screens, crystal screen and crystal screen2 of Hampton Research and JB classic 1 to 10 screens of Jena biosciences, using hanging-drop vapor-diffusion method. Typically, 2 μL of the protein solution and 2 μL of precipitation solution were mixed on a siliconized cover slip and equilibrated against 500 μL of the same precipitation solution in a reservoir. Orthorhombic crystals of size 500x300x150 μm^3 were obtained in condition [10% PEG-4000, 10% isopropanol, 100 mM sodium citrate pH 5.6, 5% glycerol] at 283K with 6XH-dtsnP168S protein in storage buffer C1. However, 6XH-dtsnP168S protein in storage buffer C2 did not show any indication of crystallization at this condition. The hexagonal bipyramid shaped crystals of size 300x300x150 μm^3, as beautiful as diamonds, were obtained with optimized lab condition S2 [2 M ammonium sulphate, 2% MPD, 100 mM Sodium MES pH 6.5, 5% glycerol] at 298K. The hexagonal crystals were obtained with 6XH-dtsnP168S protein in storage buffer C2. The diffraction intensity data for 6XH-dtsnP168S mutant translin crystals were recorded on Mar345 imaging plate using Cu $K\alpha$ radiation from a rotating

anode X-ray generator equipped with mirror optics. The orthorhombic and hexagonal crystals diffracted to low resolution of nearly 3.4Å and 4.0Å, respectively. Extensively washed hexagonal crystals from 6XH-dtsnP168S were dissolved in solution (20mM Tris, pH~11) and protein, extracted with ZipTip-C_{18} (Millipore), was analyzed for its integrity using MALDI-TOF (Axima-CFR MALDI-TOF-MS). The molecular weight of dissolved crystals was determined to be 28.4 kDa that corresponded to full length protein.

Since removal of disordered regions in protein sequence results in the improvement of diffraction quality of crystals, we purified the dtsnP168S protein, lacking any additional residues or any affinity tag, and attempted to crystallize. The hexagonal crystals could be obtained under similar conditions, S2, as described above. The diffraction intensity data for dtsnP168S protein crystals were recorded on R-AXIS IIC imaging plate using Cu *K*α radiation from a rotating anode X-ray generator at room temperature. The hexagonal crystals of dtsnP168S protein diffracted to 3.6 Å.

The diffraction intensity data from both the crystals were indexed and integrated using MOSFLM (Powell et al., 1999). The integrated data were scaled, merged and truncated using SCALA and TRUNCATE programs of CCP4 (CCP4 suite, 1994). The data collection statistics are summarized in **Table 3.1**. The space group for orthorhombic crystals (6XH-dtsnP168S) was identified to be $P222_1$ with unit cell parameters a=98.58, b=96.62, c=153.41 Å, whereas the hexagonal crystals of dtsnP168S protein belonged to space group $P6_222$ with unit cell parameters a=152.78 and c=94.01 Å. The Matthews coefficient (Matthew, 1968) of 4.3, 3.2, 2.6 and 2.1 $Å^3$ Da^{-1}, corresponding to 3 – 6 copies of the monomers of molecular weight 28.4 kDa in the asymmetric unit of the orthorhombic cell, were in the acceptable range. The unit cell of hexagonal crystals was consistent with 1 – 3 copies of the monomer (Mr, 27 kDa) in the asymmetric unit corresponding to the Matthews coefficient of 5.9, 2.9 and 2.0 $Å^3$ Da^{-1}, respectively.

The resolution of the diffraction intensity data, for hexagonal crystals of 6XH-dtsnP168S mutant translin, could be improved with post-crystallization treatments like soaking and dehydration (Heras & Martin, 2005). The crystals, grown with ammonium sulphate as precipitant [2 M ammonium sulphate, 2% MPD, 100 mM Sodium MES, 5% glycerol, pH 6.5] were soaked in higher precipitant concentration by serial transfer to increasing concentration of ammonium sulphate in three steps (10% increase in each step). The three dimensional diffraction intensity data for the dehydrated crystals were recorded on a CCD detector using BM14-U beam-line of European synchrotron radiation facility (ESRF, France) at 100K with 1^{o} oscillation per image..The dehydrated crystals diffracted to 3.2Å resolutions. The unit cell volume (a=b=150.0, c=89.9, $P6_222$) of dehydrated crystals reduced by ~8 % compared to untreated crystals (a=b=152.8, c=94.0, $P6_222$). However, mosaicity value increased to ~1^{o} in the cryo dataset compared to 0.35^{o} in room temperature data set. The synchrotron intensity data were processed and scaled with HKL2000 **(**Otwinowski & Minor, 1997**)**.

Structure solution and refinement

A) Structure of dtsnP168S protein in hexagonal ($P6_222$) space group

BLAST (Altschul et al., 1997) search of the Protein Data Bank (PDB; Berman et al., 2000*)* revealed the availability of atomic coordinates of human translin (PDB ID, 1J1J) and mouse translin (PDB ID, 1KEY), which shared nearly 50% sequence identity with drosophila translin. The initial phases for the drosophila translin mutant crystals were obtained by the molecular replacement method using the program AMoRe (Navaza et al.,1994) and packing criterion with atomic coordinates of human translin structure (PDB code, 1J1J; Sigiura et al., 2004) as a search model. The use of atomic coordinates of the complete monomer corresponding to residues 1–217 of human translin did not

yield appropriate solution for both the crystal forms during the molecular replacement trials as the solutions were rejected on the basis of packing criterion. The search for the molecular replacement solution was then carried out using several models (truncated models) in which atomic coordinates of residues at the C-terminal residues were removed from the coordinate list of the search model. The initial phases were successfully obtained for the hexagonal crystals by using the truncated model with coordinates for 1–170 residues of human translin monomer. Only one monomer was observed in the asymmetric unit of the crystals belonging to space group $P6_222$ that corresponded to high solvent content of nearly 79%. The initial phases were reasonably accurate such that 3–187 residues of drosophila translin P168S mutant could be fitted into the electron density maps. Notably, the phases were obtained with model having 1–170 residues and 171–187 residues didn't have any model bias. The coordinates were subsequently refined by REFMAC5 (Vagin et al., 2004) with intermittent model building using COOT (Emsley et al., 2004). Refinement of the model, performed using REFMAC5 with bulk solvent correction and maximum likelihood target, was monitored by R_{work} and R_{free} (Brunger et al., 1997). A weight factor of 0.05 was used for X-ray term during the refinement. Although, only one monomer was observed in the asymmetric unit of hexagonal cell, application of crystal symmetry revealed that the P168S mutant of drosophila translin existed as a tetramer in the crystal form.

B) Structure solution of 6XH-dtsnP168S translin crystallized in space group $P222_1$

The partially refined model of drosophila mutant translin with 3 – 187 residues was used as a search model in the molecular replacement calculations for solving the structure of 6XH-dtsnP168S protein that crystallized in space group $P222_1$. The

molecular replacement solution for each of the monomer expected in the asymmetric unit was carried out iteratively and four monomers could be located in the asymmetric unit of orthorhombic cell. Improvement of the molecular replacement phases was achieved by four-fold non-crystallographic symmetry (NCS) averaging using the DM program (Cowtan et al., 1994). The refined phases were not combined with the calculated phases from the model obtained by molecular replacement method so as to reduce model bias in the phases. The electron density maps computed with NCS-averaged improved phases were readily interpretable. A monomer of the 6XH-dtsnP168S translin molecule was fitted into the figure-of-merit weighted electron density maps computed using the refined phases. The four monomers in the asymmetric unit were subsequently refined by REFMAC5 using tight NCS restraints and intermittent model building using COOT software. A weight factor of 0.06 was used for X-ray term during the REFMAC refinement cycles. Application of the crystal symmetry showed that 6XH-dtsnP168S mutant translin also formed tetramer like that observed in the structure of the dtsnP168S protein crystallized in $P6_222$ space group.

RESULTS

Purification of P168S mutant of drosophila translin

Drosophila P168S mutant translin with 6XHis tag at its N-terminal (6XH-dtsnP168S) was expressed in *E. coli* XL1-Blue cells. The protein was purified with Ni-affinity column chromatography followed by anion exchange chromatography (Figure 3.1). MALDI-TOF analysis of the purified 6XH-dtsnP168S protein confirmed the integrity of the protein. Native-like drosophila mutant translin, dtsnP168S protein, was expressed and purified from *E. coli* BL21(DE3) cells transformed with vector *p*ET21a-*dtsn168s*. The dtsnP168S protein was purified with conventional chromatography methods. The

final purification was achieved by anion exchange (Q-sepharose), cation exchange (S-sepharose), followed by hydrophobic column (Phenyl-sepharose) chromatography (Figure 3.2). In both the expression systems protein expression was very high and drosophila protein was found in soluble fractions of cell lysate. The purification yield of the protein was ~10 mg per liter of bacterial cell culture.

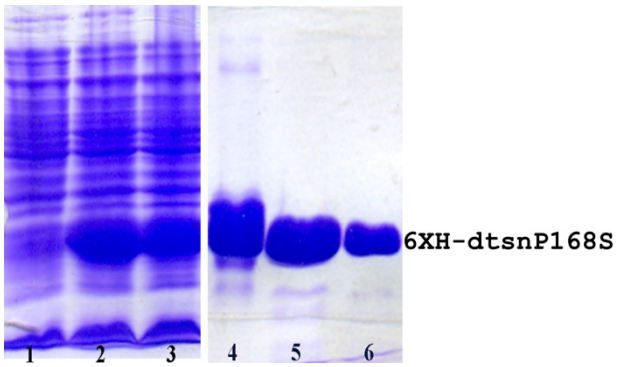
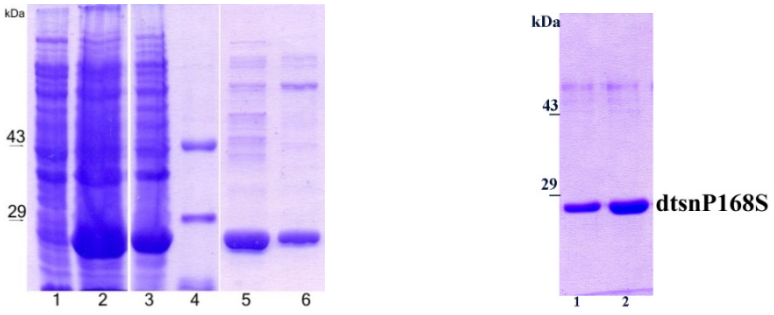


Figure 3.1 Expression and purification of drosophila P168S mutant translin with poly histidine-tag (6XH-dtsnP168S). Uninduced cells (lane 1) and induced cells (clone 1, lane 2; clone 2, lane 3). Protein fraction eluted after purification with metal affinity (lane 4) and anion exchange chromatography (lane 5 & lane 6).

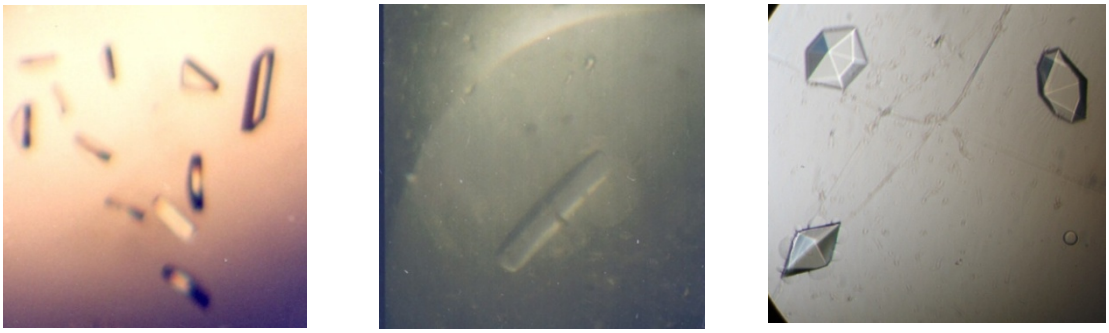


(A) **(B)**

Figure 3.2 Expression and purification of drosophila P168S mutant translin without 6Xhistidine-tag (dtsnP168S). **A)** Uninduced, lane 1; induced cells, lane 2; soluble cell lysate, lane 3; molecular weight marker, lane 4; protein fraction eluted after purification with anion exchange, lane 5; and cation exchange chromatography, lane 6. **B)** Protein fractions eluted after purification with phenyl sepahrose column chromatography, lane 1 & lane 2.

Crystallization & data collection

The orthorhombic crystals of 6XH-dtsnP168S protein used for diffraction studies were grown from 10% PEG-4000, 10% isopropanol, 100 mM sodium citrate pH 5.6, 5% glycerol at 283K. The crystal dimensions were typically 0.4X0.25X0.15 mm^3 (Figures 3.3A & 3.3B). The crystal diffracted to 3.4 Å resolution using Cu *K*α at laboratory X-ray source. The crystals belonged to space group $P222_1$. The hexagonal crystals of dtsnP168S, belonged to space group $P6_222$, were grown in crystallization solution [2 M ammonium sulphate, 2% MPD, 100 mM Sodium MES pH 6.5, 5% glycerol] at 298K. The crystal dimensions were typically 0.3X0.3X0.15 mm^3 (Figure 3.3C). The crystal diffracted to 3.6 Å resolution at laboratory X-ray source. The crystals were dissolved and were analyzed by mass-spectroscopy. The crystals were found to be composed of full length protein, as mass determined by MALDI-TOF matched exactly with the estimated molecular mass from the amino acid sequence of the protein.



(A) **(B)** **(C)**

Figure 3.3 Crystals of drosophila P168S mutant translin. Crystal were grown using hanging drop vapor diffusion method. **(A)** & **(B)** Orthorhombic crystals of poly histidine tagged drosophila P168S mutant translin (6XH-dtsnP168S) of space group $P222_1$. **(C)** The hexagonal crystals (space group $P6_222$) of drosophila P168S mutant translin (dtsnP168S).

Crystal structure analysis

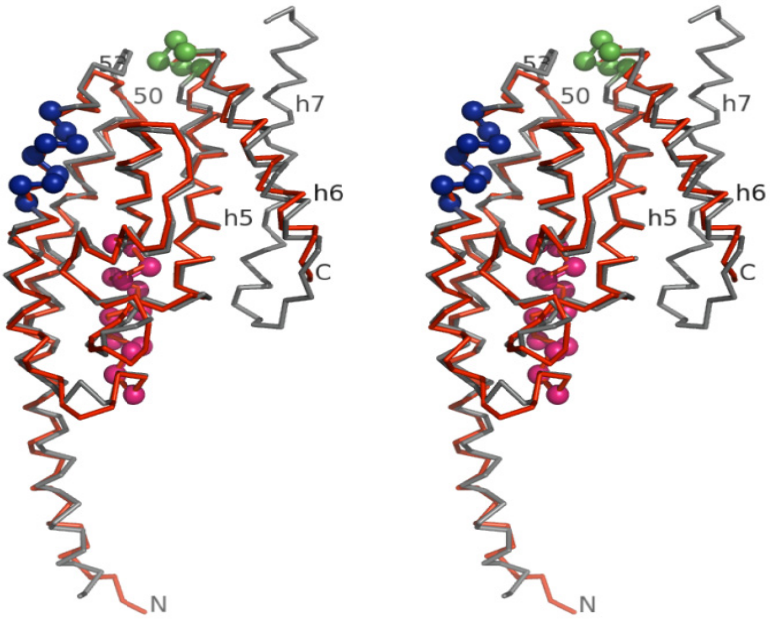
The structures of the drosophila translin P168S mutant were solved in the two crystal forms using the molecular replacement method and were refined against diffraction data upto 3.4Å resolution. The four monomers observed in the asymmetric unit of the orthorhombic ($P222_1$) crystals have the same structure and rms difference between corresponding C^{α} atoms of two monomers, typically, is 0.4Å. In comparison, the asymmetric unit of hexagonal ($P6_222$) cell is composed of a single monomer of the mutant protein. Interestingly, the drosophila translin mutant exists as a tetramer in both the crystal forms.

The final models of monomer structures consist of residues 3–187 (Figure 3.4A). Electron density was well defined for these residues (Figure 3.5). The atomic coordinates for the other N-terminal (14 residues for the 6XHis-tagged protein and 2 residues for the mutant without histidine tag) and 48 C-terminal residues were not included in the refined models, as these could not be identified in the electron density maps. The absence of the electron density for the 48 C-terminal residues was not because the crystals were composed of truncated proteins, but as the C-terminal residues were disordered in both the crystals forms. The MALDI-TOF-MS analysis of thoroughly washed and dissolved crystals showed the presence of only full-length mutant translin protein.

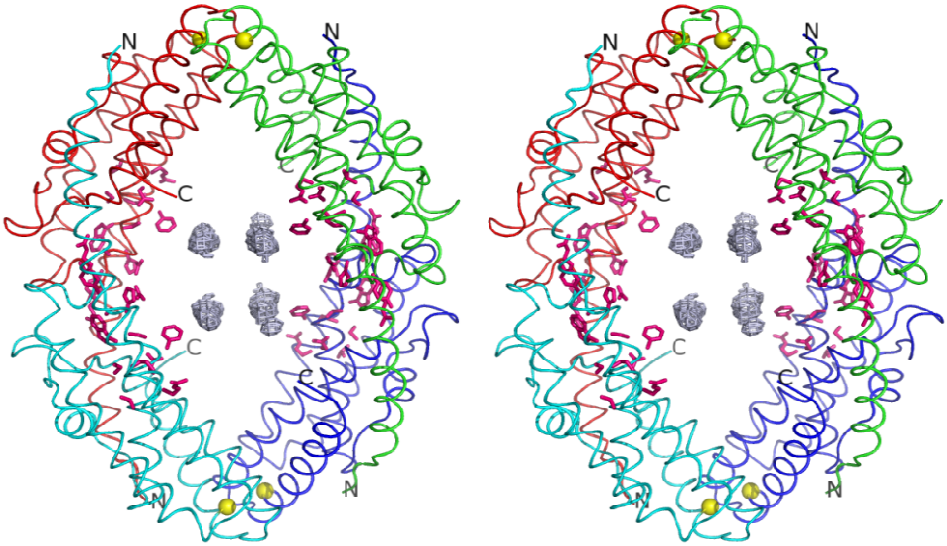
Four long sections of unaccounted electron density were observed in the central pocket of tetramer cavity of both drosophila translin crystal structures (Figure 3.4B) and these sections of electron density were not connected with the electron density for the C-terminal residue of the fitted chains in the refined structures. A poly-alanine helix of up to 10 residues could be fitted in each of these sections. The coordinates of these atoms were not refined and were not included in the refined models, however, as

unambiguous assignment of atoms in the electron density was not possible. The two structure models of drosophila translin mutant have an R-factor below 22% against all the observed data with $F/\sigma(F) \geq 0$ (Table 3.1). The evaluation using PROCHECK (Laskowski et al., 1993) revealed good stereochemistry of the structures, with nearly 90% residues in the core region of the Ramachandran plot (Table 3.1).

Hexagonal crystals, dehydrated with higher precipitant concentration, diffracted to 3.2Å. Fourier maps were computed after partial refinement. Connectivity between modeled C-terminal residues and four section of unaccounted electron density at the central pocket of tetramer could be observed at 1σ level in Fourier maps computed with partial refined structure (Figure 3.6). However, the C-terminal residues could not be fitted in the observed density with confidence. The structure refinement against 3.2 Å was not pursued further, as it did not provide additional information compared to 3.4 Å resolution structures.



(A)



(B)

Figure 3.4 Stereoview of the C^{α} trace of drosophila P168S mutant translin structure

A) Stereoview of the C^{α} trace of drosophila P168S mutant translin monomer structure (red; residues 3 and 187 identified with N and C labels respectively) superposed onto the structure of human translin protein (grey). The basic-1, basic-2 and putative GTP binding motifs of human translin sequence are shown as blue, red and green spheres, respectively. The shortened loop-1 of drosophila translin is identified with residues numbered 50 and 53. The α7 helix of human translin (labeled as h7) is not modeled in drosophila translin mutant structures as it is disordered in the crystals. Also marked are α5 and α6 helices (labeled as h5 and h6). **B**) A cartoon of the drosophila P168S mutant translin tetramer. The translin molecules of the tetramer are shown in cyan, red, green and blue colors, respectively with N and C terminus of each marked as N and C. The amino acids of basic-2 motifs are shown as magenta sticks. The positions of the four columns of unaccounted electron density observed at the center of the tetrahedral cavity are also shown as gray contours. The amino acid residue at position 168 is marked with yellow sphere. The figure was prepared using PyMol software (Delano WL, 2002).

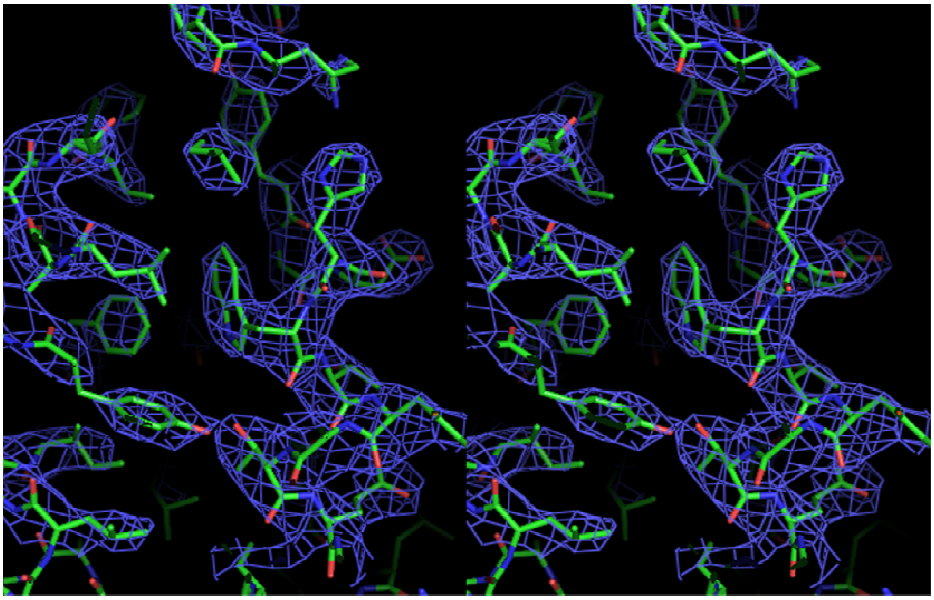


Figure 3.5 Stereo-view of the section of the final Fourier map (blue contours drawn at 1.5σ level) showing residues ^{87}RYSDHWTFI^{95} of drosophila translin, which align with the Basic-2 region residues of the human translin protein. The protein carbon, nitrogen and oxygen atoms are shown in green, blue and red respectively. The figure was prepared using PyMol software.

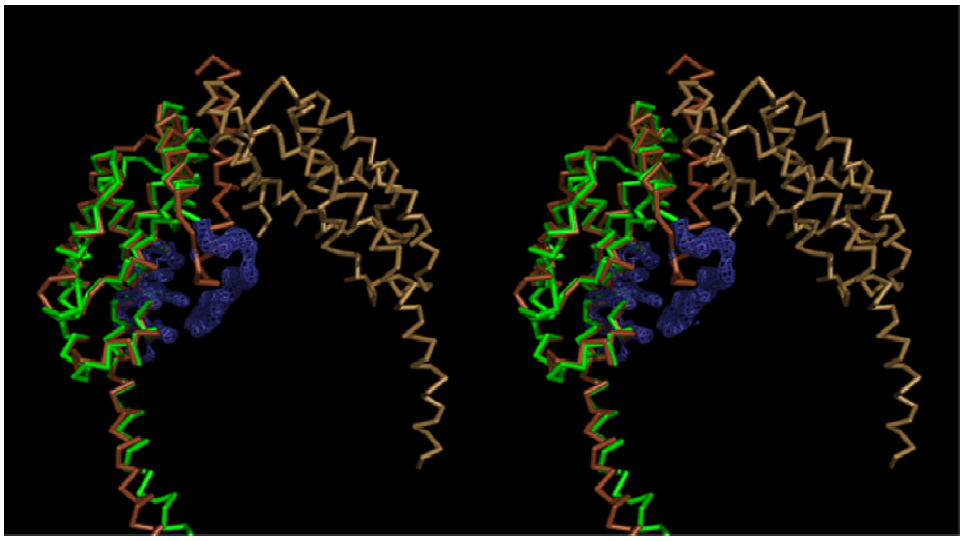


Figure 3.6 Stereo-view of the side-by-side dimer of drosophila mutant translin (green and sand colour). Also one monomer of human translin superposed onto drosophila mutant translin is shown in brown color. Section of the Fourier map, showing that electron density traces of C-terminal region. It can be observed from electron density maps (blue contours drawn at 1.0σ level) of partial refined structure of dehydrated hexagonal crystals that disordered region residues adopt altered conformation compared to α7 of human translin.

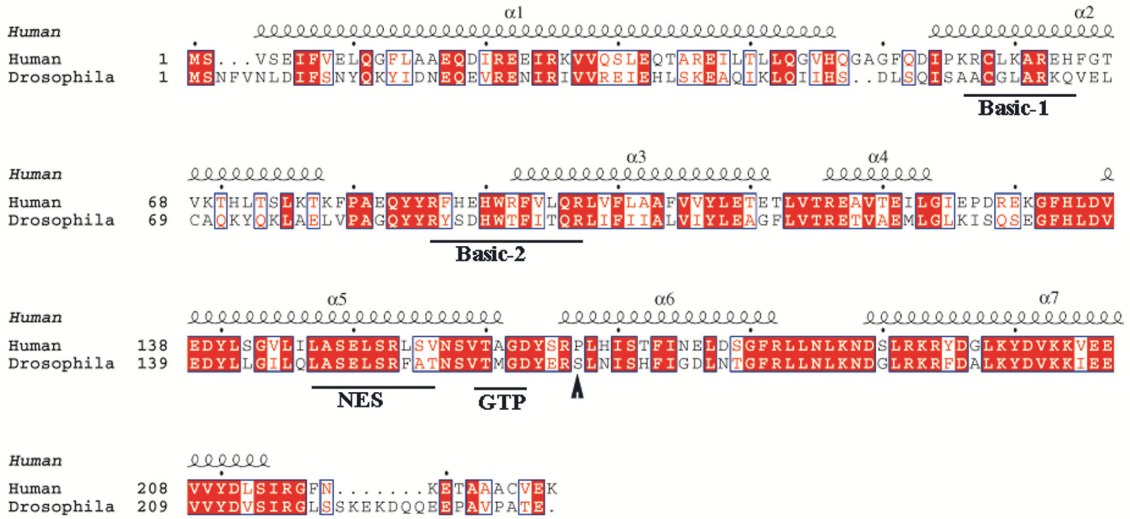


Figure 3.7 Sequence alignment of human translin with drosophila translin (with P168S mutation) proteins. The alignment was done using CLUSTALW (Thompson et al., 1994). The identical residues are highlighted and similar substitutions are boxed. The P168S mutation has been marked as ▲. Also shown are the helices (α1–α7), basic-1, basic-2, nuclear export signal (NES) and putative GTP binding (GTP) motifs for human translin sequence. The figure was prepared using ESPript (Gouet et al., 2003).

Monomer Structure

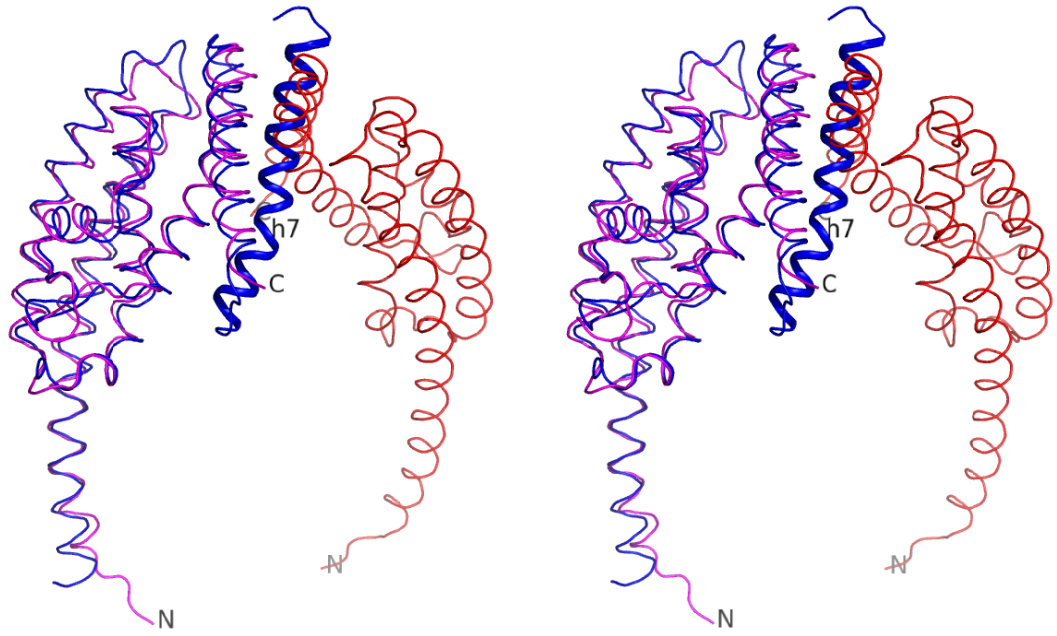
The monomer structures of the drosophila translin P168S mutant in the two crystal forms are identical with typical rms difference of 0.4Å between corresponding 736 main-chain atoms of the two monomers. Six helices (α1–α6; nomenclature as per human translin structure) can be identified in the monomer structure of the drosophila mutant (Figures 3.4A & 3.7). The structure of translin monomers match with the corresponding residues of human and mouse translin proteins (Figure 3.4A); the differences are mainly localized in the N-terminal residues of the α1 helix, the loops connecting helix pairs α1/α2 and α2/α3 and a turn region comprising of residues ^{128}ISQSE^{132}. The rms difference between the equivalent 648 back-bone atoms of the

monomers (corresponding to structurally conserved region of 167 residues) of drosophila translin mutant and its human and mouse orthologs is approximately 1.1 Å.

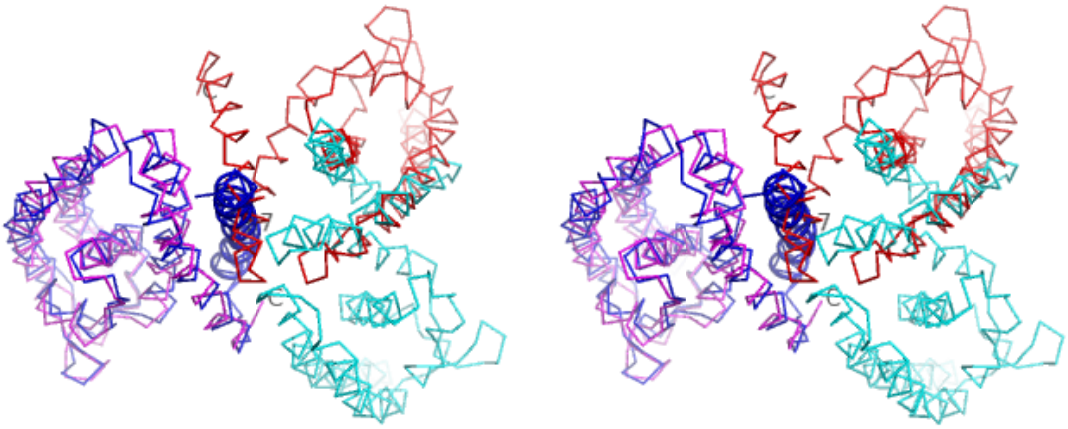
The polypeptide segment ^{47}QGAGFQ^{52} (loop-1) connects α1 and α2 helices in human translin. The loop is shortened to a tight turn (^{50}SDLS^{53}) in drosophila translin mutant due to two residues deletion in its amino acid sequence. This could have functional significance. The electron density for the loop-1 residues is well defined in both crystal forms of drosophila translin mutant suggesting that these residues are ordered and not flexible. In contrast the loop-1 residues are disordered in mouse translin crystal structure (Pascal et al., 2002). The loop-1 in drosophila translin structures is placed in juxtaposition to residues ^{160}VTMGD^{164}, which align with the putative GTP binding motif (^{159}VTAGD^{163}) of mouse translin (Figures 3.4A & 3.7). The residues 189–213 in human and mouse translin proteins adopt a helical conformation (α7 helix), which forms a three-helix bundle with α5 and α6 helices (Pascal et al., 2002, Sugiura et al., 2004). The equivalent C-terminal residues are disordered in both the crystal structures of drosophila mutant translin. It is clear from the oligomeric structure of drosophila translin mutant that its C-terminal residues are placed differently as compared to the observed topology of these residues in human or mouse translin, owing to closer proximity of another monomer in the drosophila translin structure (Figures 3.6 & 3.8A).

The analysis of the available crystal structures of translin proteins does not explain well why the C-terminal polypeptide of the drosophila translin mutant adopts topology different from human and mouse translins. The proline residue (corresponding to Pro-168 of drosophila translin) is in *trans* conformation, is surrounded by Ser-155, Ser-158, Val-159, Leu-168, Ile-170, Val-209 and Leu-212 residues, and is situated near the N-

terminus of the α6 helix in the human translin crystal structure. The proline residue is buried within the monomer and does not interact with other subunits of human translin octamer. The side-chain of Ser-168 in the crystal structures of the drosophila translin mutant points towards Ala-156 (equivalent of Ser-155 in human translin sequence) and forms a hydrogen bond with its main-chain oxygen atom. The Pro to Ser change thus could have been accommodated, as also expected from in-silico modeling, without causing any significant disturbance in the tertiary structure of the mutant. The analysis of intramolecular interactions of the α7 helix of human translin further does not explain a large structural change seen in the topology of the C-terminus residues. These interactions were analyzed using distance criterion and contacts of α7 were defined as those 3.7 Å or less from the other residues of the monomer. The amino acid residues of the α7 helix form several hydrophobic and electrostatic contacts with amino acids of the α5 and α6 helices. Except for Tyr-195 and Asn-175, the other contact residues of the α7 helix of human translin are strictly conserved in drosophila translin mutant. The residue Tyr-195 of α7 stacks against Phe-181 of α6 and the side-chain atoms of Asn-175 of α5 and Glu-201 of α7 interact in human translin. The hydrophobic interactions between the residue pairs of α6 and α7 helices should be possible in drosophila protein as well, as a Phe substitutes for Tyr-195 in the latter protein. The interactions between residues equivalent to Asn-175/Glu-201 pair of human translin shall not be possible in drosophila translin, however, as the Asn residue is substituted by Gly in drosophila translin sequence.



(A)



(B)

Figure 3.8 The side-by-side dimers in the crystals structures of drosophila mutant and human translin proteins. **A**) The cartoons of two monomers of drosophila mutant translin are shown in magenta and green colors, respectively. The C^{α} trace of one monomer of human translin superposed onto drosophila mutant translin is shown in blue color. The α7 helix of human translin (labeled as h7) is shown as thick ribbon and 187^{th} residue of drosophila protein is marked with C. The α7 helix of human translin monomer overlaps with the second monomer of the side-by-side dimer of drosophila protein in the superposed structures. **B**) Drosophila translin monomers are shown in magenta and red, and two monomers of human translin side-by-side dimer are shown in blue and cyan colors.

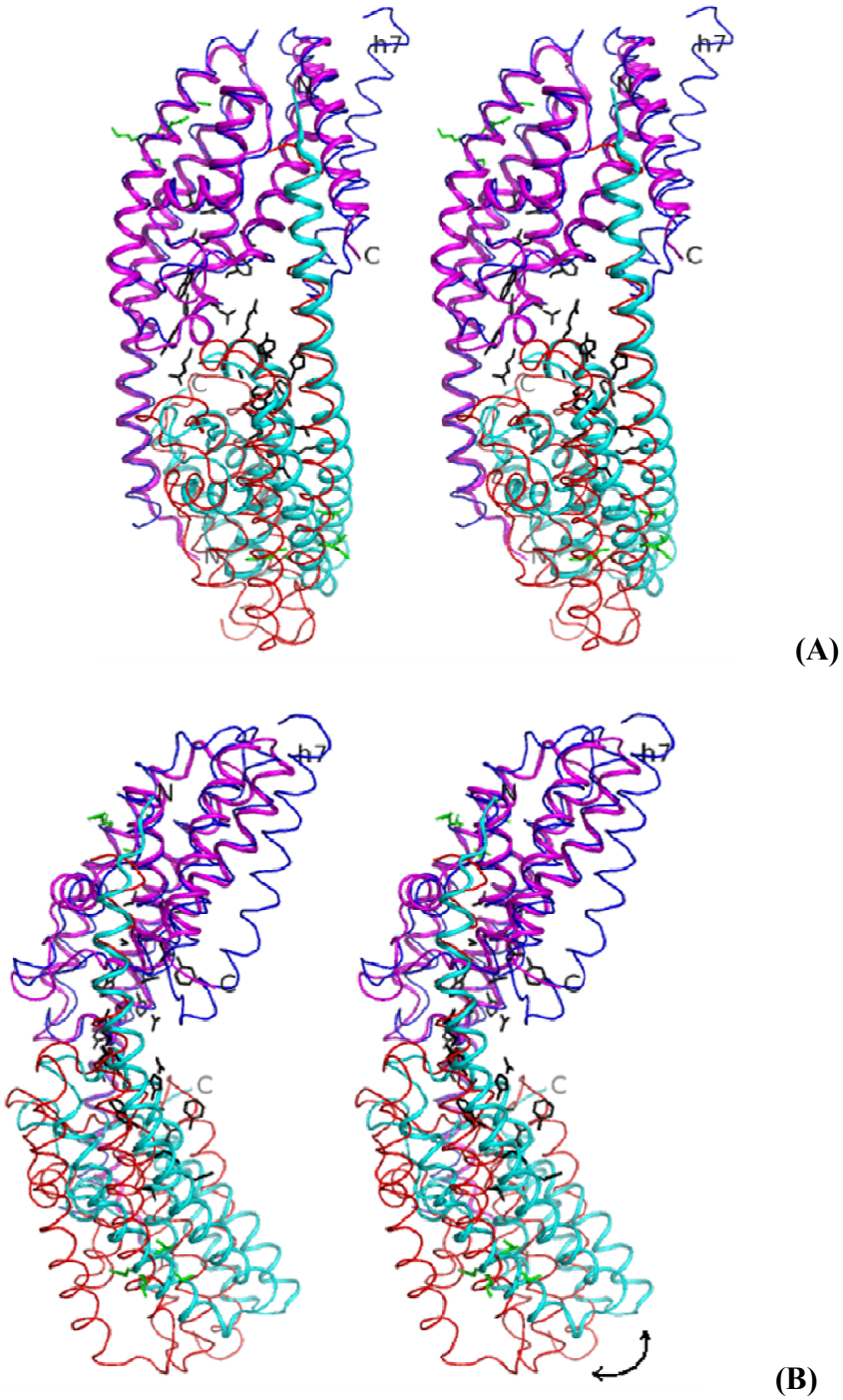


Figure 3.9 Different views (A and B) of the cartoons of two monomers in up-down dimer configuration of drosophila translin mutant shown in magenta and cyan colors, respectively. The α1 helix of one monomer interacts with α6 of the second molecule. Also shown in sticks are amino acids of basic-1 (green) and basic-2 (black) motifs. The C^{α} traces of the two monomers of human translin, superposed onto drosophila translin mutant using DaliLite server (Holm & Park, 2000), are shown in blue and red colors. The α7 helix of human translin is labeled as h7. A decrease in the curvature of the up-down dimer of human translin mutant is shown by an arrow.

Quaternary Structure

The drosophila translin mutant protein exists as a tetramer in both the crystal forms (Figure 3.4B). The tetramers, generated by the application of crystal symmetry of the two crystal forms, superpose on to each other with an rms difference of 1.0 Å for the 732 equivalent C^{α} atoms. The protein interfaces, surfaces and assemblies (PISA) service at EBI (Krissinal et al., 2002) estimates a gain of nearly –66.6 kcal/mol in solvation free energy on the formation of the tetramer and characterizes the tetrameric structure of drosophila translin mutant as a stable unit. In comparison, an octameric structure, which can be visualized as formed by two tetramers, is observed for human and mouse translin proteins (Pascal et al., 2002, Sugiura et al., 2004). The tetramer of drosophila translin mutant, however, is distinctly different from the two possible tetramers of human or mouse translin proteins.

Two types of dimers can be identified to form the tetrameric structure of drosophila translin mutant. Akin to the dimers identified from the octameric quaternary structures of human and mouse translin proteins (Pascal et al., 2002, Sugiura et al., 2004), one of the dimer results from side-by-side juxtaposition of the two monomers (side-by-side dimer; Figure 3.8A), and the second dimer is formed due to up-and-down placement of the two monomers (up-down dimer; Figure 3.9A).

The two monomers constituting a side-by-side dimer are placed closer to each other in drosophila translin mutant (Figures 3.8A & 3.8B), as compared to side-by-side dimers in human and mouse translin structures. Closer juxtaposition of the two side-by-side monomers in human or mouse translin protein is restricted due to steric hindrance by its α7 helix. In drosophila mutant translin interactions between Tyr-165/Leu-169 pair and the Ser-157 residues of the two monomers characterize the side-by-side dimer

interface, and the interaction mode differs totally from human and mouse translin orthologs. The non-conservation of interactions at the side-by-side dimer interface has also been observed in the homologous human and mouse translin crystal structures and it was suggested that the relative positions of translin monomers in multimeric structure could easily be changed (Sugiura et al., 2004). It can be visualized from the present structures that the dramatic change in the side-by-side dimer interface of drosophila translin mutant is due to the altered topology of the C-terminal tail of the monomers (Figures 3.6 & 3.8B).

Interestingly, the up-down dimer of drosophila translin mutant resembles the up-down dimer of human or mouse protein, and equivalent residues are buried at these dimer interfaces. Similar to human and mouse translin proteins, the interface of drosophila translin up-down dimer is characterized by the interactions between the extended N-terminal α1 helix of one monomer with the residues constituting loop-3, and the α5 and α6 helices of the other monomer of the dimer. The up-down dimer is also stabilized by a salt-bridge between Arg-87 of loop-2 and Glu-139 of α5 in the structure of drosophila translin mutant and in human and mouse translins. The structural comparison thus reveals conservation of up- down dimer in varying multimeric states of translin proteins.

Nearly 1750 $Å^2$ (~16%) of the solvent accessible area is buried at the up-down interface on dimer formation in drosophila translin mutant. This matches with the loss of solvent accessible area of the up-and-down dimer of human translin. The PISA service at EBI, using the atomic coordinates of drosophila translin mutant and human translin protein, estimates a gain of nearly -24 kcal/mol in solvation free energy on the formation of up-down dimer of drosophila and human translin proteins, and suggests

the complexation significance score of 1.0 for the up-down dimer. In contrast, a loss of nearly 570 $Å^2$ of solvent accessible surface area and a gain of nearly -9 kcal/mol in solvation free energy for the side-by-side dimer formation of drosophila translin mutant is estimated. Similarly, a loss of nearly 450 $Å^2$ solvent accessible surface area and a gain of -5.7 kcal/mol in solvation free energy for human translin protein are estimated. The comparison and analysis of available translin structures thus clearly suggest that the up-down dimer, conserved in the known translin structures irrespective of their oligomeric status, is more stable than the side-by-side dimer.

The up-down inter-subunit interactions result in the clustering of the amino acids of the basic-2 nucleic acid binding motif in the inner surface of the curved up-down dimer (Figures 3.7, 3.9A & 3.9B). The inner curvature of the up-down dimer of drosophila translin mutant is more than that of human translin (Figure 3.9B). A change of nearly 17° in inter-subunit angle was estimated from the angle between α3 helices of human and drosophila mutant translin up-down dimers in the superposed structures. The increase in the curvature of the up-down dimer of drosophila mutant could be due to the interaction between two up-down dimers at the poles of tetrameric ellipsoid. These interactions have been noted here as the side-by-side dimer interface and are not conserved in human and mouse translin orthologs.

The size of drosophila translin mutant molecule (tetramer) is smaller as compared to the size of human or mouse translin octamer. On the basis of dissociation pattern predicted by the PISA server, two up-down dimers form a tetramer of drosophila translin mutant. These up-down dimers are related by twofold symmetry. The structural superposition of drosophila translin tetramer onto human translin octamer shows that the octamer of human or mouse translin can also be constructed by pseudo 4-fold rotation of up-down dimer (Figure 3.10). The analyses thus suggest that multimeric

structures of translin orthologs may differ by the number and relative position of the conserved up-down dimer units. Importantly, interactions at the side-by-side dimer interface shall not be conserved between different multimeric structures.

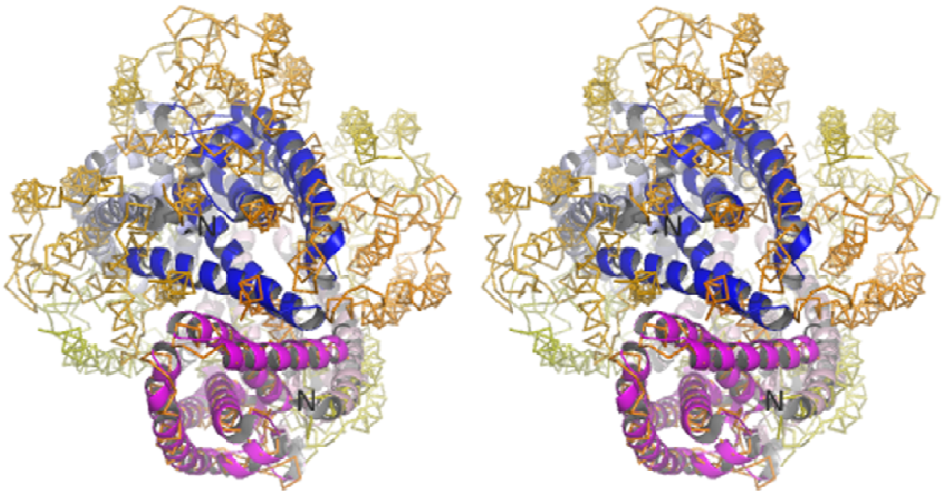


Figure 3.10 Superposition of drosophila translin tetramer onto human translin octamer. The cartoons of the four molecules of drosophila translin tetramer are shown in magenta and blue shades. The C^{α} trace of the eight monomers of human translin octamer is shown in different shades of orange color. View down the pseudo 4-fold rotation axis of the human translin octamer

DISCUSSION

Translin, a ssDNA/RNA binding protein, is widely conserved in eukaryotes. The drosophila translin, having 235 residues, shares nearly 50% sequence identity with human and mouse orthologs. We solved the three dimensional crystal structure of DNA-binding incompetent P168S mutant of drosophila translin from two crystal forms. Crystal structure of the mutant drosophila translin monomer matches up to 187 amino acids with the corresponding residues of human and mouse orthologs. Despite high sequence identity (~50%), the crystal structure of drosophila translin P168S mutant

differ dramatically at its C-terminal regions from those of human and mouse proteins. Residues 187–235 of drosophila translin mutant are disordered in both the crystal forms, obtained with totally different crystallization conditions, as the electron density for these residues was not well defined in Fourier maps. However, the residues 190–213 of human or mouse translin adopt a helical conformation and the resulting helix (α7) forms a three-helix bundle together with the α5 and α6 helices. It can be observed from electron density maps (Figure 3.6) that disordered region residues adopt altered conformation compare to human and mouse translin. It is further confirmed from the quaternary structure of drosophila translin mutant that its C-terminal polypeptide cannot trace the topology observed in human and mouse translin proteins. The perfect match between both the structures of drosophila P168S mutant translin suggests that the observed structural differences of drosophila mutant translin monomer, as compared to those of human and mouse orthologs, are not due to crystallization artifacts.

An altered tertiary structure of drosophila translin mutant, interestingly, leads to dramatic changes in its quaternary structure, as compared to those of human and mouse orthologs. The drosophila translin mutant is an open doughnut shaped tetramer the central cavity of which is accessible from two sides (Figure 3.4B), in contrast to closed octameric crystal structure of human and mouse translin proteins (Figure 3.10). The tetramer is stabilized by a number of inter-molecular hydrogen bonds and salt bridges, and is conserved in both the crystal forms. The PISA server also characterized the tetrameric structure of drosophila translin mutant as a stable unit with nearly -66.6 kcal/mol gain in solvation free energy upon formation of the tetramer.

Up-down dimer is the evolutionary conserved unit

The biochemical studies of mouse translin suggested that dimer is the minimum structural unit needed for DNA- and RNA-binding (Wu et al., 1998). It is logical to expect that the functional dimer is conserved in translin proteins due to high sequence-function similarity, even if translin from different species may exist in varying multimeric states. Two possible dimers, up-down and side by side dimers, have been discussed from the crystal structures of human and mouse translin proteins (Pascal et al., 2002, Sugiura et al., 2004), though the conserved unit was not identified clearly.

The side-by-side dimer, presumably, in mouse and human translin can be stabilized by Cys-225, though Cys appears to be dispensable both for dimerization and DNA- and RNA-binding (Wu et al., 1998). The Cys residue is also not strictly conserved in the known translin sequences. In drosophila translin mutant structure the placement of and interactions between the two monomers of the side-by-side dimer differ completely from those of human and mouse orthologs. Also, the relative positions of the monomers in two side-by-side dimers in human differ significantly from the mouse translin protein structure. In contrast, the up-down dimer is conserved in the known translin structures irrespective of their oligomeric status, and the contact residues at the up-down dimer interface of drosophila translin mutant match with those of human or mouse translin protein. Significantly higher solvation free energy gain upon formation of the up-down dimer assembly is estimated for each of the human, mouse and drosophila mutant translin proteins, as compared to solvation free energy gain for the side-by-side dimers. The occurrence of more number of hydrogen bonds and salt bridges at the up-down dimer interface suggests also that this dimer is energetically more stable. Further, the up-down dimer is identified by the PISA server as the most stable unit in the

dissociation pattern of mutant translin tetramer. Taken together, the up-down dimer can thus be considered as an evolutionarily conserved primary unit in translin proteins.

It can be reasoned from the structural analysis that the tetrameric quaternary structure of drosophila translin mutant results from interactions of two up-down dimers and the octameric structure of human and mouse translin, in comparison, can be generated by the repetition of four up-down dimer units. This model is useful in describing the structural features of wild-type drosophila (discussed in chapter4) and chicken translin proteins. The chicken protein is observed to form a ring-shaped structure composed of 10 subunits of 27 kDa translin polypeptide (Aoki et al., 1999). From the plasticity observed in drosophila translin mutant structure, it can be thought that decamers of chicken translin are formed by five up-down dimers.

The crystal structures of drosophila translin mutant also reveal that curvature of up-down dimer may be linked with the number of these repeat units assembled into a given multimeric state; the curvature of a up-down dimer of drosophila translin tetramer is more than that of human and mouse octameric translin. This dynamicity in curvature of up-down dimer may help to accommodated more number of up-down dimers and resulting in a ellipsoid of the radius larger than that observed for human and mouse translin. This proposal thus rationalizes EM data that chicken translin ring is of larger size as compared to the rings of human and mouse translin.

The recently available crystal structures of heteromeric translin-trax complex from human and drosophila (Ye et al., 2011; Tian et al., 2011) can also be well explained with the hypothesis of up-down dimer. In the heteromeric complex structures different stoichiometry ratio of translin and trax were observed. These complexes can be formed by combination of translin-trax heteromeric up-down dimer and translin-translin homomeric up-down dimers. Four up-down heteromeric dimers result in $translin_4$-$trax_4$

composition, while replacement of one or two heteromeric dimers with homomeric translin dimers would result in $translin_5$-$trax_3$ or $translin_6$-$trax_2$ composition, respectively.

The crystal structure of drosophila P168S mutant translin was also compared with the recently available crystal structure of drosophila translin-trax complex (Tian et al., 2011). The drosophila mutant structure matches up to 187 amino acids with the monomer structure of wild-type drosophila translin. However, C-terminal residues of wild-type drosophila translin, in the heteromeric complex structure, adopt the same topology as observed for human and mouse translin but differ dramatically with drosophila mutant translin. Earlier trax has been reported to rescue the nucleic acid binding properties of DNA-binding incompetent mutant of human translin (Chennathukuzhi et al., 2001a). It would be interesting to see whether the C-terminus of wild-type drosophila translin adopts the topology observed in heteromeric translin-trax complex or is it stabilized in the observed conformation on interaction with trax? Thus the crystal structure of wild-type drosophila translin would be useful to understand the biology of translin proteins.

In conclusion, the present study has demonstrated plasticity in the structure of translin proteins and it is suggested here that the up-down dimer of translin molecules is the evolutionarily conserved unit. The present studies also suggested that different quaternary organization in drosophila P168S mutant translin might be responsible for its DNA-binding incompetence. The observed change in quaternary structure of the mutant presumably is due to altered topology of its C-terminal polypeptide.

Table 3.1 Crystallographic data and refinement statistics

	6XH-dtsnP168S	dtsnP168S
Data statistics		
Unit Cell	a=98.58, b=96.62, c=153.41 Å	a=152.78, c=94.01 Å
Space group	$P222_1$	$P6_222$
Solvent content (%)[a]	61.68	78.8
Resolution limit (Å)	3.4	3.6
Total reflections	59405	20696
Unique reflections	18817	7311
Completeness (%), overall/outer shell[b]	91.4/76.5	93.7/90.2
R_{merge}, overall/outer shell[b]	0.099/0.354	0.106/0.412
Mean I/ mean σ(I), overall/outer shell[b]	11.8/2.6	13.3/2.8
Wilson B ($Å^2$)	73.2	51.0
Refinement and model statistics		
Resolution range (Å)	20 – 3.4	20 – 3.6
Final $R_{work,}$ overall/outer shell[b]	0.203/0.277	0.215/0.239
Final R_{free}, overall/outer shell[b]	0.269/0.386	0.249/0.238
Number of non-hydrogen atoms	5925	1502
Average thermal parameter ($Å^2$)	42.7	65.4
Ramachandaran plot[c]	93.9/5.6/0.3/0.3	89/9.8/1.2/0
Root-mean-square deviation from ideality		
Bond lengths (Å)	0.015	0.017
Bond angles (^{o})	1.583	1.804
B, bonded main chain ($Å^2$)	0.630	0.780
B, bonded side chain ($Å^2$)	1.542	2.068

S.D is the standard deviation

[a]As estimated from the Matthews coefficient corresponding to four monomers in $P222_1$ crystals and one monomer in $P6_222$ crystals.

[b]The highest resolution shells for 6XH-P168S mutant translin and P168S mutant translin are 3.40–3.49 Å and 3.60–3.79 Å, respectively.

[c]Percentage of residues in most favoured/ additionally allowed/ generously allowed/ disallowed regions of the Ramachandran plot.

Chapter 4

Biochemical and biophysical characterization of drosophila translin & human translin-trax complex

Part A

Biochemical characterization of wild-type drosophila translin

Introduction

Translin, a ssDNA and RNA binding protein, is well conserved in eukaryote evolution. Translin binds to consensus sequences found at breakpoint junction of chromosomal translocations and recruits to the nucleus in DNA damaging events (Aoki et al., 1995; Kasai et al., 1997). Translin also binds with conserved 3′-untranslated regions (3′-UTRs) and regulates the mRNA transport and mRNA translation (Kwon et al., 1991; Han et al., 1995). GTP binding modulates the nucleic acid binding properties of translin such that GTP binding reduces the RNA, not the DNA, binding ability of translin (Chennathukuzhi et al., 2001b). Translin forms a multimeric structure that is octamer in human and mouse translin while chicken translin is expected to form a decamer, and loss of multimeric structure in human translin abrogates the nucleic acid binding activity (Aoki et al., 1999; Pascal et al., 2002). Two basic regions, basic-1 and basic-2, were identified on translin sequence and substitution in either of these resulted in loss of RNA binding activity, while only basic-2 region was observed to be crucial for DNA-binding activity (Aoki et al., 1999).

Our earlier studies (described in chapter 3 of present thesis) demonstrated that drosophila P168S mutant translin forms a stable tetramer in crystal form, compared to the DNA-binding competent octamer of human and mouse translin. The drosophila mutant was DNA-binding incompetent. It was proposed that change in oligomeric status that is due to altered C-tail conformation, might be responsible for the lack of DNA-binding activity in drosophila mutant translin. To establish the role of multimeric structure in translin proteins and to find the role of conserved proline (Pro-168), we characterized the wild-type drosophila translin for its DNA-binding properties and multimeric status. We also synthesized a chimeric translin, a chimera of N-terminal

region of drosophila translin with C-terminus of human translin. The results emphasized the critical role of C-terminal residues in the formation of DNA-binding competent oligomer (Gupta et al., 2008).

MATERIALS AND METHODS

Cloning of wild-type translin of drosophila melanogaster

The *p*QE9-*tsn* and *p*QE30-*dtsn168s* constructs (obtained as a kind gift from Dr. M. Kasai, National Institute of Health, Japan; and Prof B. J. Rao, Tata Institute of Fundamental Research, Mumbai, respectively) were used to purify the human translin protein and drosophila P168S mutant translin (as described in chapter 3 of present thesis) proteins, respectively.

The expression vector *p*QE30-*dtsn*, used to express wild-type drosophila translin fused with N-terminal 6XHis tag, was synthesized with Quick-Change site-directed mutagenesis kit (Stratagene, USA) using *p*QE30-*dtsn168s* construct as template and mutagenizing primer pairs (CGACTACGAGCGT**C**CCCTGAATATCTCCCAT and ATGGGAGATATTCAGGG**G**ACGCTCGTAGTCG). The mutation incorporated in primer sequence has been shown in big font letter and the underlined letters denote the mutated codon. The resultant nicked circular DNA, after incubation with *Dpn1* enzyme to digest the parental DNA template, were transformed into competent *E. coli* XL1-Blue cells. The *p*QE30-*dtsn* plasmid, isolated from *E. coli* XL1-Blue, was transformed into *E. coli* BL21 (DE3) cells for over-expression of wild-type drosophila translin.

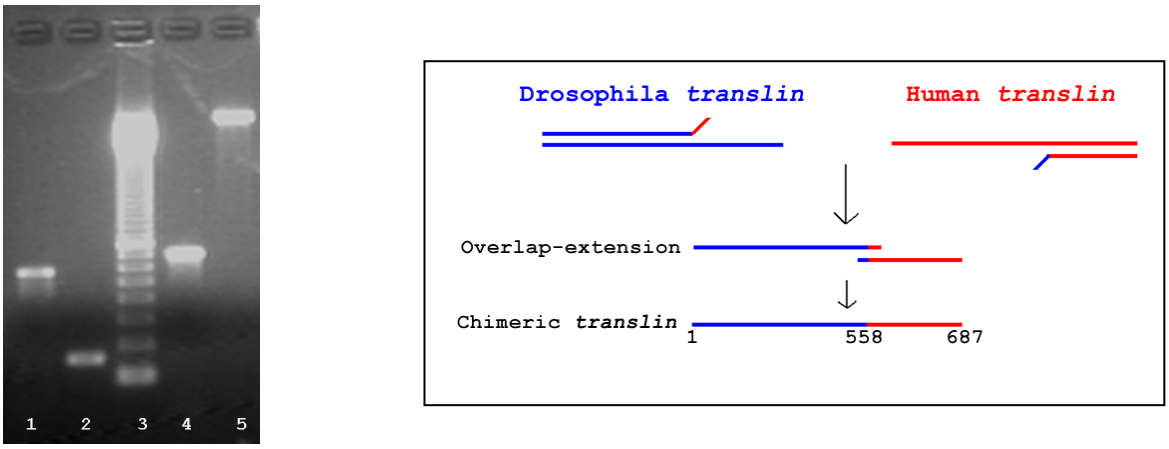
The drosophila *translin* gene was also subcloned into *NdeI* and *BamHI* restriction sites of the *p*ET28a plasmid to express the wild-type drosophila translin with cleavable histidine tag (N-terminal 6XHis tag followed by thrombin proteases cleavage site). The

drosophila *translin* gene was PCR amplified with *PfuTurbo* DNA polymerase (Stratagene, USA) using primer pairs (forward; CATCACGGACATATGTCGAACTTCGTGAACTT and reverse; ACTGGATCCTTATTC GGTTGCAGGAACAGC) and *p*QE30-*dtsn* as template. The *NdeI* and *BamHI* restriction sites were incorporated in the primer sequence (shown as underlined text). The resultant *p*ET28a-*dtsn* construct was transformed into *E. coli* BL21 (DE3) cells for the expression of the wild-type drosophila translin protein with cleavable 6XHis tag.

Construction of Chimeric-translin

A chimeric *translin* gene was constructed using primer overlap extension method. A DNA fragments of 583 base-pairs (bp) (encoding 187 N-terminal residues of drosophila translin with P168S substitution along with flanking region sequence) was PCR amplified with *PfuTurbo* DNA polymerase using *p*QE30-*dtsn168s* template and primer pairs (forward, CATCACGGACATATGTCGAACTTCGTGAACTT and reverse, AGTCATTTTTCAGGTTCAGCAGACGGAAGC) (lane 1, Figure 4.1A). Another DNA fragment of 153 bp (encoding 43 C-terminal residues (186–228) of human-translin along with flanking region sequences) was amplified using primers (forward; GTCTGCTGAACCTGAAAAATGACTCCCTG and reverse; TGACGGATCC TTATTTTTCAACACAAGCTG) and *p*QE9-*tsn* as template (lane 2, Figure 4.1A). Both the DNA fragments shared the complimentary sequence of 24 bases at their terminals. These overlapping DNA fragments were purified from gel-excised bands and were extended with overlap extension PCR. After 5 cycles of overlap-extension reaction, the terminal primers (forward, CATCACGGACATATGTCGAACTTCGTGAACTT and

reverse; TGACGGATCCTTATTTTTCAACACAAGCTG) were added to PCR reaction to increase the yield. The resultant chimeric *translin* gene (*Ctranslin*) (lane 4, Figure 4.1A) was digested with *NdeI* and *BamHI* restriction enzymes and subcloned in-frame with N-teminal 6XHis tag sequence of *p*ET28a vector pre-digested with *NdeI* and *BamHI* restriction enzymes (lane 5, Figure 4.1A). The *p*ET28a-*Ctranslin* construct was transformed into *E. coli* BL21 (DE3) cells for over-expression and purification of chimeric translin. The nucleotide sequence of all constructs was confirmed by DNA sequencing using an ABI automated DNA sequencer.



(A) **(B)**

Figure 4.1 Synthesis of chimeric *translin*. A) Two DNA fragments, one of 583 bp (lane 1) amplified from *drosophila translin* and another fragment of 153 bp (lane 2) amplified from *human translin*. These two overlapping DNA fragments were extended to form a chimeric *translin* (lane 4). *p*ET28a vector digested with *NdeI* and *BamHI* restriction enzymes (lane 5). 100 bp ladders (lane 3) were used as marker. **B)** The schematic presentation of chimeric *translin*. DNA templates of drosophila translin and human translin have been shown in blue and red colour, respectively. The chimeric *translin* with 687 bp encodes for 229 residue long chimeric translin protein.

Purification of translin proteins

E. coli BL21 (DE3) cells transformed with *p*QE30-*dtsn* were used for the over-expression of the wild-type drosophila translin with N-terminal 6XHis tag. The cells were grown in LB media in presence of ampicillin (100 mg/L) antibiotic. The expression of wild-type drosophila translin protein was induced with 0.5 mM IPTG in late log phase of growth (~0.8 OD_{600}). After 4 hour of induction, cells were harvested with centrifugation at 6000g for 10min and were suspended in lysis buffer [25 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM DTT, 5% Glycerol] in presence of Lysozyme (1 mg/mL) and protease-inhibitor [one tablet of protease-inhibitor cocktail (Roche, USA) per 100 mL of cell suspension] and were maintained on ice for 1 hour. The cells suspensions were lysed by ultra-sonication for 10 minutes on ice in pulse mode. The cell lysate was cleared of cellular debris by centrifugation at 20000g for one hour. The protein purification was achieved using low-pressure liquid-chromatography system (BioRad Biologic Lp) at room temperature. The supernatant of cell lysate was loaded onto the metal chelating affinity matrix (Ni-IDA, Amersham-Pharmacia) pre-equilibrated with buffer A1 [25 mM Tris-HCl pH 8.0, 100 mM NaCl and 50 mM imidazole]. After loading, the protein column was washed with buffer A1 till the OD_{280} reached ~ 0.01. The bound proteins were eluted with a linear gradient of imidazole (50–400 mM) in buffer A1. The wild-type drosophila translin protein eluted at ~200 mM concentration of imidazole. The eluted fractions of the protein were directly loaded onto pre-packed Q-sepharose column (BioRad, USA) pre-equilibrated with buffer A2 [25 mM Tris-HCl pH 8.0, 100 mM NaCl]. The bound proteins were eluted with a linear gradient of 100–500 mM NaCl in buffer A2. The wild-type drosophila translin eluted at ~350mM of NaCl. The purification progress was monitored by SDS-PAGE. The

purified wild-type drosophila translin was dialyzed against storage buffer [25 mM Tris-HCl pH 8.0, 100 mM NaCl] and concentrated to 12 mg/mL using centricorn (Millipore, USA). The protein was stored at 4^{o}C.

The chimeric translin (Ctranslin) protein was also expressed in *E. coli* BL21 (DE3) cells and purified with Ni-IDA followed by anion exchange (Q-sepharose) chromatography. Human translin and drosophila P168S mutant translin (6XH-dtsnP168S) were purified from *E. coli* BL21 (DE3) cells harboring *p*QE9-*tsn* and *p*QE30-*dtsn168s* expression vectors, respectively. All translin proteins, purified with immobilized metal-chelating affinity chromatography and ion-exchange chromatography, were treated with DNaseI and RNaseA enzymes for overnight at 20^{o}C. All 6XHis tagged translin proteins were purified from the reaction milieu by using Ni-IDA matrix.

E. coli BL21 (DE3) cells harboring *p*ET28a-*dtsn* were used for over-expression of the wild-type drosophila translin fused with cleavable 6XHis tag at its N-terminal. The protein purification was achieved with two stage purification as described above. The purified protein was incubated with thrombin proteases (Amersham pharmacia, USA) overnight at 20^{o}C for cleavage of N-terminal 6XHis tag. The reaction mixture was loaded on Ni-IDA matrix and the wild-type drosophila translin (without any additional residue) eluted in unbound fractions.

Gel filtration analysis

The molecular weights of the purified translin proteins were measured by Superdex^{TM} 200 10/300 GL column (GE Healthcare, USA) coupled with high pressure liquid chromatography system (HPLC PU-2089 from JASCO, Japan) at flow rate of 0.5 mL/min in buffer [25 mM Tris-HCl pH 8.0, 100 mM NaCl]. The Superdex^{TM} 200

column was calibrated with gel filtration molecular weight markers (Amersham-Pharmacia; Carbonic anhydrase, 29 kDa; Alcohol dehydrogenase, 150 kDa; β-Amylase, 200 kDa; Ferritin, 440 kDa). The molecular mass of translin proteins was determined based on their elution time on the gel-filtration column using the calibration curve. The eluted peak of each independent gel filtration experiment was adjudged on SDS-PAGE.

DNA-binding assay

DNA-binding properties of translin proteins were assessed with electrophoresis mobility shift assays (EMSA). The wild-type drosophila translin and chimeric translin proteins at two concentrations (1500 and 3000 pmol of monomeric proteins), and the human and 6XH-dtsnP168S mutant translin proteins (3000 pmol of monomeric protein) were incubated with a 24-mer ssDNA (Bcl-CL1; 5' GCCCTCCTGCCCTCCTTCCGCGGG 3') (150 pmol) for 30 min in 15 µl of binding buffer B1 [25 mM Tris–HCl, pH 8.0, 1 mM EDTA and 0.1 M NaCl]. The products were resolved on 1.5% agarose gel in TAE buffer using gel electrophoresis pre-stained with ethidium bromide. The protein-DNA complexes with retarded mobility, compared to free DNA, were visualized under UV-light using gel-documentation system from Syngene (Synoptics, Cambridge, England). The agarose regions containing gel-shifted complexes were excised and adjudged on 12% SDS-PAGE to confirm the composition.

GTP binding studies of drosophila translin

Dot-blot assay was performed to assess the affinity of wild-type drosophila translin towards GTP. Human translin (2 µg), wild-type drosophila translin (2 µg) and BSA (5 µg) proteins were blotted independently on a PVDF membrane. Subsequently, the

membrane was incubated with 10 pmol of radiolabeled [α-^{32}P]GTP in 15 mL of binding buffer B1 [25 mM Tris-HCl, pH 8.0, 1 mM EDTA and 0.1 M NaCl] in a petri dish at room temperature on a gel rocker. The membrane was washed with the binding buffer B1 three times for five minutes each. The dried membrane was exposed to X-ray film for autoradiograph. The membrane was also stained with 0.1% Ponceau S dye.

Crystallization of wild-type drosophila translin and data collection

The purified drosophila wild-type translin protein, fused with 6XHis tag, was concentrated to 8 mg/mL in storage buffer [25mM Tris–HCl, pH 8.0 and 0.1M NaCl]. Initial attempts to crystallize drosophila translin protein were made using hanging drop vapor diffusion method with commercially available crystallization screens like crystal screen, crystal screen 2 and SaltRx kits from Hampton research and protein-complex crystallization kit from Sigma. Crystals of drosophila translin protein were observed in number of conditions with precipitant like PEG-4000 and ammonium sulphate. However, attempts to improve the crystal size resulted in highly amorphous crystals as observed from the birefringence properties. The optically good quality crystals were obtained using micro-seeding methods from crystallization buffer [22% PEG 3350, 100 mM Tris-HCl pH 8.5]. The crystals diffracted to limited resolution of 8-10 Å using Cu-Kα at laboratory X-ray source. Further attempts to improve the crystal quality were made with wild-type drosophila translin protein (without poly histidine tag). However, crystals, obtained with crystallization buffer [20% MPEG-2000, 0.1 M Na-Hepes pH 7.5, 0.2 M NaCl], also did not diffract to better than 6 Å resolution.

RESULTS

Expression and purification of translin proteins

Wild-type drosophila translin was expressed in *E. coli* BL21 (DE3) cells and purified from soluble fractions of cell lysate using metal chelating affinity (Ni-IDA) matrix followed by anion exchange chromatography. Final purification was achieved with gel-filtration column (Figure 4.2A). Expression and purification yields were high. Human translin and drosophila P168S mutant translin were also purified from bacterial expression system using similar protocol.

The crystal structure analysis of drosophila P168S mutant translin revealed that C-terminus region of mutant protein adopted altered conformation compared to human translin. We hypothesized that altered conformation of C-terminus resulted in tetrameric oligomerization of P168S mutant protein, compared to octamer of human translin, and that resulted in the loss of DNA-binding activity in the mutant. To confirm the role of C-terminus region, a synthetic chimeric translin protein of 229 residues was engineered using recombinant DNA technology. The chimeric translin had 1–186 residues from N-terminus of drosophila translin (1–186) with P168S substitution, and 187–229 residues from C-terminus (186–228) of human translin that harbored a cysteine residue (equivalent to Cys-225 of human translin) proposed earlier to stabilize human translin oligomer (Wu et al., 1998). The chimeric translin was expressed in *E. coli* BL21 (DE3) cells transformed with *p*ET28a-*ctranslin.* Interestingly, the chimera protein folded properly and was found in soluble fraction of cell lysate. Chimeric protein was purified with two stage purification using Ni-affinity and anion exchange chromatography (Figure 4.2B).

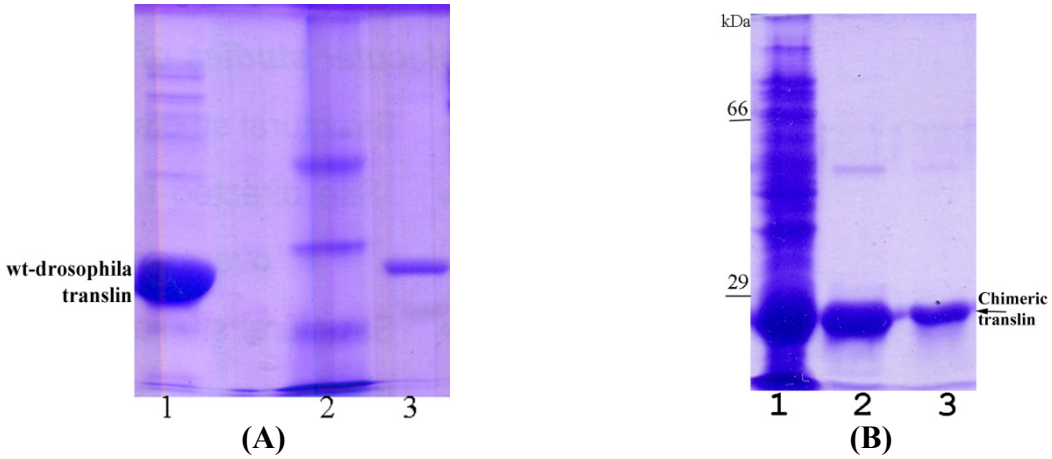
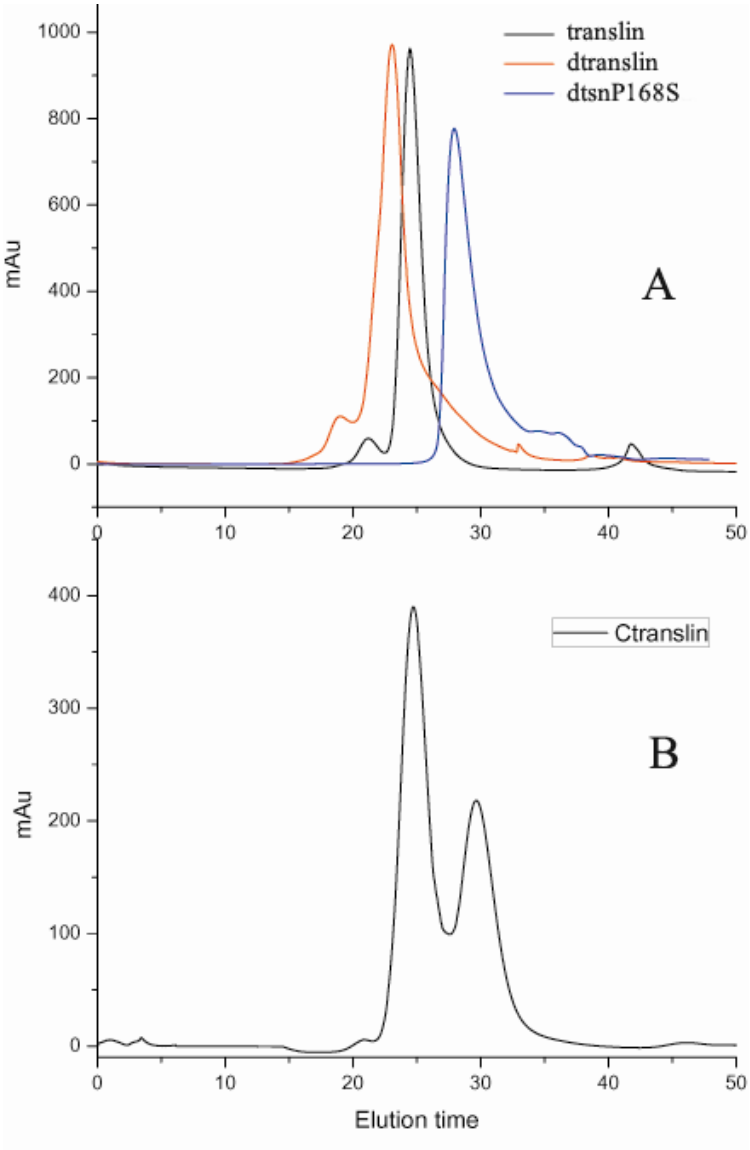


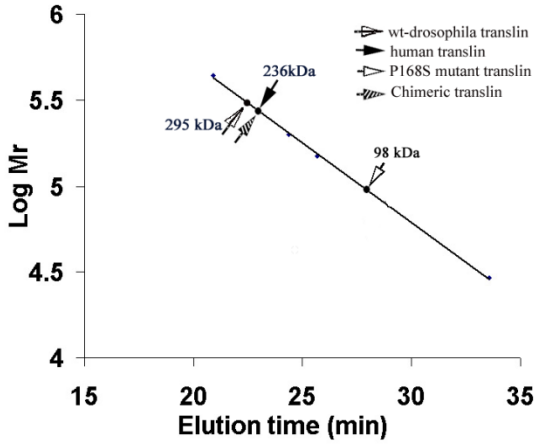
Figure 4.2 Purification of translin proteins. Fractions of translin protein eluted at various steps during protein purification were analyzed on 12% SDS-PAGE. **A)** Lane 1 and lane 3 are wild-type drosophila translin eluted from anion exchange and gel-filtration chromatography, respectively. Lane 2 showing the molecular weight markers of 20 kDa, 29 kDA, 43kDa, respectively. **B)** Purification of chimeric translin. Lane 1 was loaded with soluble fraction of cell lysate, while lane 2 and lane 3 were loaded with chimeric translin purified after metal affinity column and anion exchange chromatography, respectively.

Oligomeric status of drosophila wild-type translin and chimeric translin

The molecular weight of the translin proteins was determined based on its elution time on the molecular sieve Superdex^{TM} 200 10/300 GL column. The molecular weight of wild-type drosophila translin was estimated to be 295 kDa based on its elution profile in gel-filtration column that corresponds to decamer (Figure 4.3). However, a possibility to exist as octamer also cannot be ruled out given the resolution of gel-filtration chromatography. In contrast, the molecular weight of drosophila P168S mutant translin was estimated to be 98 kDa from the elution profile that corresponds to tetramer distinctly differing from other translin orthologs. It also confirms that tetrameric structure, observed in crystal structure, was not due to crystallization artifacts. In comparison, the molecular weight of human translin was determined to be 236 kDa

consistent with the reported octameric oligomerization. Interestingly, a major peak in the elution profile for the chimeric translin was observed corresponding to the molecular weight of 236 kDa very much similar to that observed for human translin. However, a small peak was also observed in the elution profile of chimeric translin corresponding to a dimer with estimated molecular weight of 62 kDa (Figure 4.3). The major peak in the elution profile of all translin proteins were harvested for analyses of their DNA-binding activity.





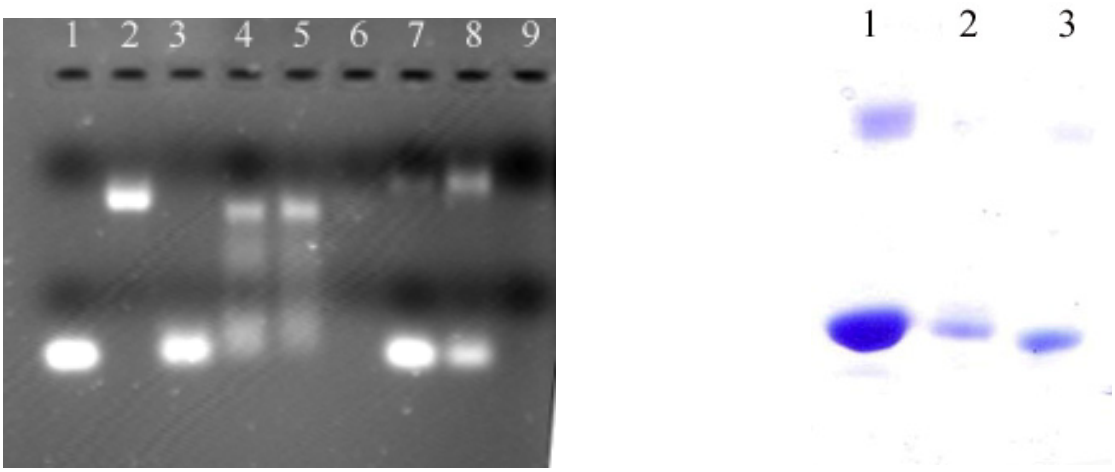
(C)

Figure 4.3 Gel-filtration analyses of translin proteins A) Elution profile of human translin (translin), wild-type drosophila translin (dtranslin) and drosophila P168Smutant translin (dtsnP168S) **B)** Elution profile of chimeric translin (ctranslin). **C)** Determination of molecular mass of translin proteins using calibration curve. The figure shows the standard curve obtained from the elution time of gel filtration molecular weight markers (Carbonic anhydrase, 29 kDa; Alcohol dehydrogenase, 150 kDa; β-Amylase, 200 kDa; Ferritin, 440 kDa). The molecular weights of 236 kDa are the best estimates from the elution profiles of human and stable multimer of chimeric proteins, whereas molecular weights of 295 and 98 kDa are expected for wild-type and P168S mutant of drosophila translin proteins, respectively.

DNA binding properties of translin proteins

The DNA-binding activity of the translin proteins were checked with a 24-mer Bcl-CL1 ssDNA in a gel-shift assay and were compared with the DNA-binding activity of the purified human translin, under identical conditions of assay. As observed earlier (Sengupta et al., 2006), the drosophila P168S mutant translin did not bind to this specific ssDNA (lane 3; Figure 4.4A). Intriguingly, the wild-type drosophila translin (lanes 4 and 5; Figure 4.4A) and 236 kDa chimeric proteins (lanes 7 and 8; Figure 4.4A) showed distinct DNA-binding activity. The drosophila translin-ssDNA complex migrated faster as compared to the human complex (lane 2; Figure 4.4A), while the migration of the gel-shifted complex of chimera was very similar to that of human

translin-ssDNA complex. The intensity of the gel-shifted complexes however revealed weaker binding of wild-type drosophila and chimeric translins with the 24-mer Bcl-CL1 ssDNA, compared to the human protein. The SDS-PAGE analysis (Figure 4.4B) of the gel pieces excised from the agarose gel and containing the gel-shifted complexes showed that all the analyzed gel-shifted complexes were genuinely due to translin proteins.

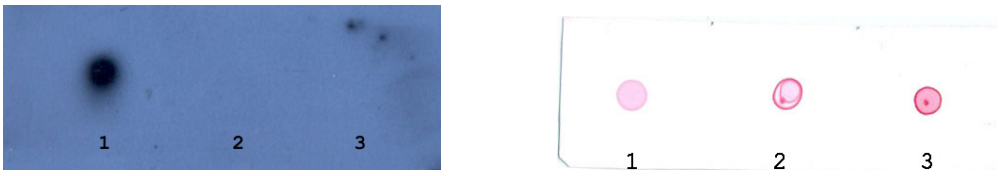


(A) **(B)**

Figure 4.4 Analyses of the DNA-binding activity of the translin proteins. **A)** Bcl-CL1 24-mer ssDNA (150 pmol) was incubated with the known concentrations of RNase A and DNase I treated translin proteins. The mixtures were analyzed on 1.5 % agarose gel and stained with ethidium bromide; lane 1, only DNA; lane 2, human translin (3000 pmol); lane 3, drosophila 6XH-P168S translin mutant (3000 pmol); lanes 4 and 5 for wild-type drosophila translin (1500 and 3000 pmol, respectively); lanes 7 and 8 for 220 kDa chimeric translin (1500 and 3000 pmol, respectively); lanes 6 and 9, only the wild-type drosophila and chimeric translin proteins respectively and no DNA. **B)** Analysis of gel-shifted complex on SDS-PAGE. The gel-shifted ssDNA-translin complexes were excised from the agarose gel and the gel pieces were loaded onto the 12% SDS-PAGE gel and subjected to electrophoresis; the excised agarose gel bands of human, wild-type drosophila and chimeric translin proteins (lanes 1–3, respectively).

Drosophila translin does not bind with GTP

A putative GTP binding site (^{161}VTAGD^{165}) was predicted on the human translin sequence and binding of GTP have been reported to modulate the nucleic acid binding properties of human translin (Chennathukuzhi et al., 2001b). Drosophila translin has the equivalent site (^{162}VTMGD^{166}) with Ala to Met change. We checked the purified translin proteins for their affinity towards GTP. The GTP binding studies were performed with radiolabeled [α-^{32}P]GTP using dot blot assays (Figures 4.5A&B). Human translin showed very strong affinity towards GTP as reported earlier. However, wild-type drosophila translin did not bind with GTP under similar condition.



(A) **(B)**

Figure 4.5 GTP binding studies of translin proteins with dot-blot assay.

A) Protein samples were blotted onto a PVDF membrane. The membrane was incubated with [α-^{32}P] GTP in binding buffer [25 mM Tris–HCl pH 8.0, 1 mM EDTA and 0.1 M NaCl]. The membrane was washed extensively with binding buffer and exposed on X-ray film overnight for autoradiography. Dot-blot 1, human translin (2 µg); dot-blot 2, drosophila wild-type translin (2 µg); dot-blot 3 control BSA (5 µg). **B)** Staining of the PVDF membrane with 0.1% Poaceau S dye.

Crystallization of wild-type drosophila translin

Wild-type drosophila translin was crystallized using hanging and sitting drop vapor diffusion methods. Crystals of drosophila translin could be grown in number of conditions (Figure 4.6). The crystals were composed of wild-type drosophila translin as tested on SDS page (data not shown). However, most of them were highly amorphous. Attempts were made to obtain diffraction quality crystals with wild-type drosophila translin (without poly histidine tag) and using micro seeding methods. However, the best obtained crystals did not diffract to better than 6 Å resolution using laboratory X-ray source.

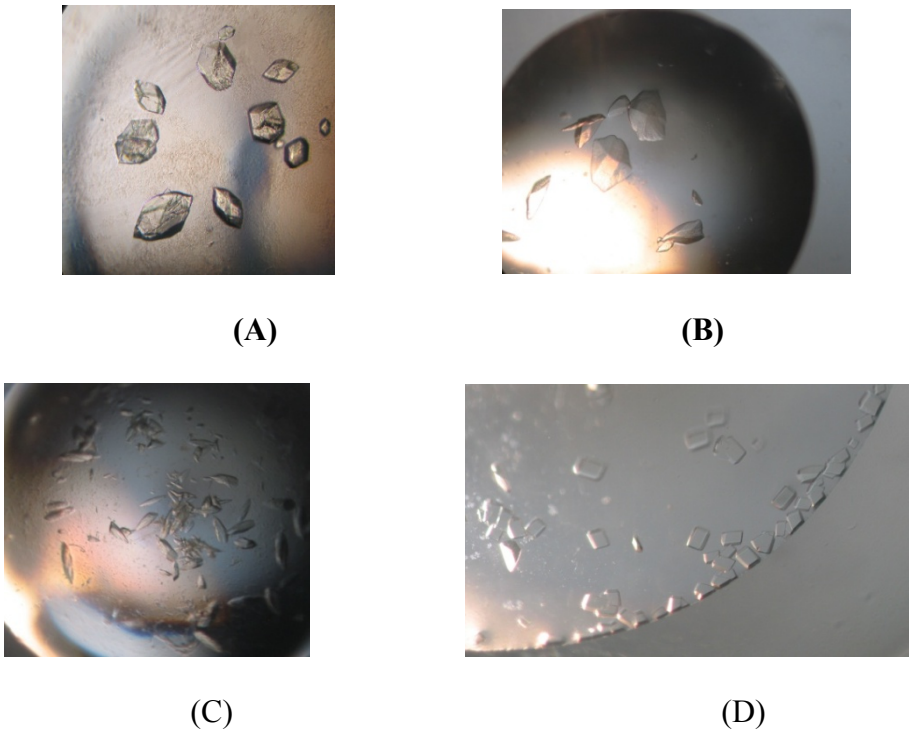


Figure 4.6 Crystals of wild-type drosophila translin proteins. **(A)** – **(C)** crystal were obtained with 6XHis tagged drosophila translin and diffracted to 8–10Å. The crystals in **(C)** were grown using micro-seeding method. **(D)** Crystals were obtained with wild-type drosophila translin without poly histidine tag and diffracted to 6 Å resolution only.

DISCUSSION

In the present study, we have described the purification of wild-type drosophila translin and characterized the protein for DNA binding activity. The purified wild-type drosophila translin protein was observed to bind to ssDNA sequence as determined using gel-retardation experiments. In comparison the drosophila P168 mutant translin that was mistakenly reported as wild-type drosophila translin (Sengupta et al., 2006), was found to be DNA-binding incompetent. It was observed from the multiple sequence alignment that Pro-168 is conserved in the translin orthologs, though its role had not been discussed. The DNA binding studies of drosophila translin proteins suggested that the conserved proline residue may be crucial for nucleic acid binding activity; either being involved in interactions with the nucleic acid substrate or in maintenance of the translin fold.

To elucidate the role of conserved proline, we analyzed the purified translin proteins for its multimeric state using gel-filtration chromatography. The drosophila P168S mutant translin was observed to form tetramer in both crystals and solution form compared to the octameric status of human translin. In comparison, drosophila wild-type translin was observed to exist as octamer/decamer on the molecular sieve column. The results suggested that difference in multimeric state of mutant and wild-type protein might be responsible for the loss of DNA-binding activity in mutant translin. However, the role of proline-168 in multimerization of translin proteins was not clear from the structural analysis of mutant translin.

In the crystal structure, the C-terminus residues of mutant translin were found to be disordered and adopted altered conformation compared to human translin. It was proposed that tetrameric conformation of mutant translin might be due to its disordered

C-terminus (Gupta et al., 2008). To confirm this hypothesis and to ascertain the role of conserved Pro-168, we synthesized a chimeric protein of 229 residues by swapping the C-terminus of drosophila P168S mutant translin with the corresponding residues of human translin. The hypothesis was also based on another observation that Cys-225 in human translin could stabilize the C-terminus of chimeric translin due to disulphide interactions (Wu et al., 1998). As expected, a 236 kDa octamer of the chimeric translin protein was observed in the gel filtration experiments. Interestingly, the octameric chimeric protein, harboring Ser at position 168 (instead of Pro) and nucleic acid binding motifs of drosophila translin, showed DNA-binding activity with ssDNA Bcl-CL1 sequence on gel-shifted assay, which suggested that the oligomeric status, and not the conserved Pro at 168, is critical for DNA-binding activity of translin proteins. Further, the elution profile in gel filtration experiments and mobility of the gel-shifted complex of chimera were similar to that observed for human translin, and differed from the wild-type drosophila translin reported here. These findings suggested that the C-terminal residues predominantly encode the oligomeric status of translin proteins and the Pro residue at position 168 may be required for the stability of translin fold. The presence of another peak in gel-filtration experiments corresponding to a dimer of chimeric protein also suggested that proline is required for stable octameric state.

DNA-binding activity of human translin has earlier been reported to mediate through the basic region residues, basic-2, found at the inner curvature of octameric translin as the substitutions in the basic-2 region (residues 86–97, Figure 4.7) of human translin completely inhibited its DNA binding activity (Aoki et al., 1999). These basic region residues are not strictly conserved in drosophila and other translin orthologs. In the present studies, wild-type drosophila translin was observed to be DNA-binding

competent. However, the binding affinity was weaker as compared to human protein. The DNA binding studies were performed with the ssDNA probe used earlier for the DNA binding studies of human translin (based on consensus sequences found at chromosomal breakpoint junction) (Aoki et al., 1995). It is possible that the nucleic acid binding preference is altered in the wild-type drosophila translin. The S. pombe translin also had been shown to bind to transcripts $(GU)_n$ with much higher affinity compared to the corresponding $d(GT)_n$ sequence (Jacob et al., 2004).

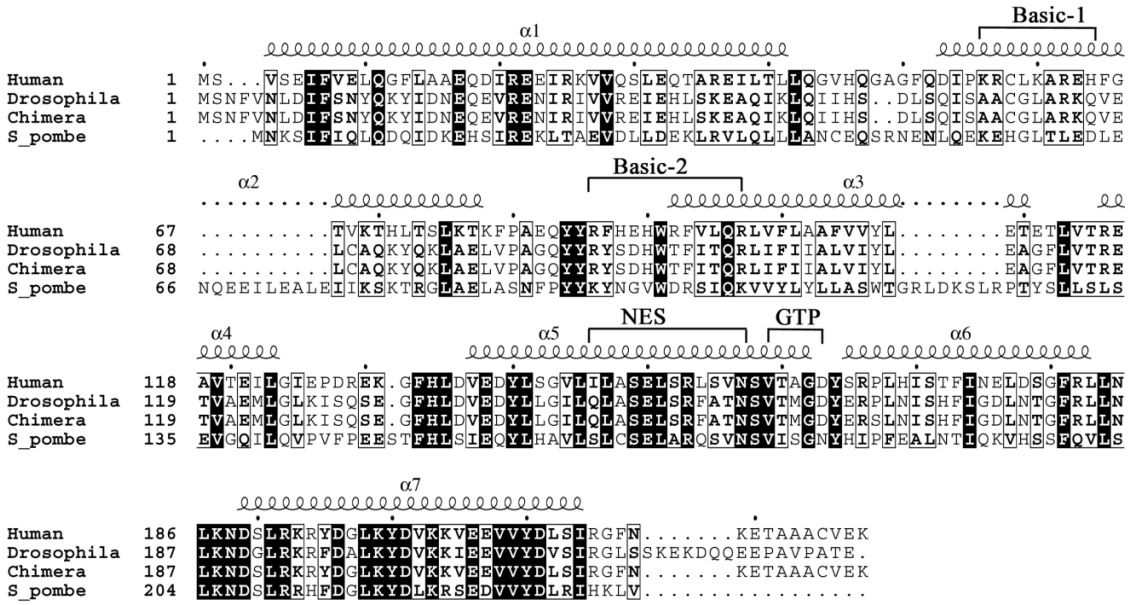


Figure 4.7 Sequence alignment of human translin (Human) with wild-type drosophila (Drosophila), chimeric (Chimera), and S. pombe (S_pombe) translin proteins. The alignment was done using ClustalW (Thompson et al., 1994). The identical residues are highlighted and conservative substitutions are boxed. Also shown are the Basic-1, Basic-2, nuclear export signal (NES) and putative GTP binding (GTP) motifs for human translin sequence. The secondary structure (helices α1–α7) of human translin is also shown. The figure was prepared using ESPript (Gouet et al., 2003).

GTP, which has been shown to modulate the binding affinity of mouse translin towards RNA (Chennathukuzhi et al., 2001b), did not show binding towards wild-type drosophila translin. The crystal structures of the P168S mutant of drosophila translin suggested that the polypeptide segment (^{47}QGAGFQ^{52}, loop-1) that connects α1 and α2 helices in human translin, is shortened to a tight turn (^{50}SDLS^{53}) in drosophila translin due to two residues deletion in its amino acid sequence (Figure 4.7) (discussed in chapter 3 of the present thesis). This loop-1 is placed in juxtaposition to putative GTP binding site in translin structures. The rigid conformation of loop-1 may contribute in blocking the interaction of GTP with drosophila translin. This shortening of loop region was observed in insects (*Aedes aegypti* and *Bombyx morii*) and plant translin (Arabidopsis, Rice, Grapes etc) in the multiple sequence alignment. It shall be interesting to see if these translin orthologs interact with GTP.

In conclusion, the present studies strongly imply that **(1)** Wild-type drosophila translin is DNA-binding competent, **(2)** The DNA-binding propensity of translin proteins critically hinges on the oligomeric status of the protein that may be encoded by the C-terminus residues, **(3)** Pro-168 is required for the stable fold of translin proteins and **(4)** GTP binding activity observed for human and mouse translin may not be conserved in many translin orthologs.

Part B
Biochemical and biophysical characterization of translin-trax complex

Introduction

Translin **a**ssociated factor-**X** (trax), a 33 kDa protein, was identified as an interacting partner of DNA/RNA binding protein, translin, using yeast two hybrid experiments (Aoki et al., 1997; Wu et al., 1999a). Trax is post transcriptionally stabilized by translin and deletion of *translin* gene results in the total loss of trax protein in mouse, drosophila and *S. pombe* (Chennathukuzhi et al., 2003a; Claussen et al., 2006; Jaendling et al., 2008). Co-precipitation of trax with translin in immuno-precipitation studies suggested that translin-trax heteromeric complex is the predominant cellular species (Finkenstadt et al., 2000). Both translin and trax were components of RNA binding complex purified from brain and testis extracts (Taira et al., 1998; Finkenstadt et al., 2002). Translin-trax complex regulates mRNA trafficking and translation in neurons and plays a role in synaptic plasticity (Li et al., 2008). Recently, heteromeric translin-trax complex was reported to possess RNase activity and was identified as a key component of RNA induced silencing complex (RISC) (Liu et al., 2009; Ye et al., 2011). Trax has also been proposed to be associated with DNA repair as trax interacts with C1D protein following exposure to γ-irradiations. C1D is an activator of DNA-PK (DNA-dependent protein kinase) that is essential for the repair of DNA double strand breaks (Erdemir et al., 2002a). The trax has also been found to interact with several other proteins involved in variety of cellular functions (Jaendling et al., 2010; Li et al., 2008).

Here we have reported the characterization of trax using translin-trax complex. We designed a bacterial co-expression system to purify human translin-trax complex to homogeneity in high yield. We modeled the partial structure of trax using bioinformatics tools. The derived structural model was supported by biophysical methods like size exclusion chromatography and circular dichroism. We also synthesized a heterologous

complex of drosophila translin with human trax to understand *in situ* mechanism of translin-trax complex formation.

METHODS

Recombinant plasmids

The *p*QE9-*tsn* and *p*ET28a-*trx* constructs, used to express 6XHis tagged human translin and 6XHis tagged human trax, were kind gifts from Professors M. Kasai, National Institute of Health, Tokyo, Japan and N.B. Hecht, University of Pennsylvania, USA, respectively. The *p*ET21a-*dtsn168s*, used to express drosophila P168S mutant translin in native form was that used in earlier studies (described in chapter 3 of this thesis).

The construct, *p*ET21a-*tsn,* was generated by cloning human *translin* into the *NdeI* and *BamHI* restriction sites of *p*ET21a plasmid. The *translin* cDNA was PCR amplified with *PfuTurbo* DNA polymerase using primer pairs (forward, ATCACGCTCATATGTCTGTGAGCGAGATC; reverse, TGACGGATCCTTATTT TTCAACACAAGCTG) and *p*QE9-*tsn* as template. The restriction sites incorporated in the primers are underlined. The construct was transformed into *E. coli* DH5α cells used as cloning host.

The *p*ET28a-*trx229* construct*,* used to express truncated trax (having 32–163 and 180–277 residues of trax and cysteine at position 226; equivalent to Cys-225 of human translin), was synthesized using overlap-extension method. Two overlapping DNA fragments were amplified in two independent PCR reaction using *p*ET28a-*trx* template and primer pairs (forward, TCATCTGCTAGCGTGATGTTGGCCTTTAAATC and reverse, CAGGTGTGACTCTCAGCTCTTTCCCATTGTCTTCAGTCG) and (forward, GAAGACAATGGGAAAGAGCTGAGAGTCACACCTGTCG and reverse, CATTTC

TGAAGCTTATGAAAACACACATGCCAACATATGTTTTGG), respectively. Both DNA fragments, purified from gel-excised bands, were extended with overlap-extension PCR and cloned into *NheI* and *BamHI* restriction sites of *p*ET28a vector. The construct *p*ET28a-*trx229* was expressed in *E. coli* BL21 (DE3) cells.

Co-expression and purification of translin-trax complexes

The *E. coli* BL21 (DE3) cells harboring *p*ET28a-*trx* construct were co-transformed with *p*ET21a-*tsn* to co-express human 6XHis-trax (will be referred as TRAX throughout the text) and translin (untagged) proteins. Since both the *ColE1*-derived plasmids have same origin of replications and thus incompatible, the co-transformants were selected and maintained under the selection pressure of two antibiotics ampicillin and kanamycin, the resistance for which was conferred by the *p*ET21a and *p*ET28a vectors, respectively. The cells were grown in LB media in presence of ampicillin 100 mg/L and kanamycin 50 mg/L. The cell culture was induced with 1mM IPTG in late log phase of growth (OD_{600} ~0.8). The cells were harvested and lysed in buffer (50 mM Tris-HCl pH 8.0, 200 mM NaCl, 5% glycerol, 1 mM DTT) using ultra-sonication as discussed in 'Materials and Methods' chapter of this thesis. The translin-TRAX complex was purified by loading the cell lysate onto the chelating-sepharose FF column (Ni-IDA) pre-equilibrated with buffer A1 (25 mM Tris-HCl pH 8.0, 100 mM NaCl and 50 mM imidazole). The bound proteins were eluted with a linear gradient of imidazole (50–500 mM). The translin-TRAX complex eluted at ~250 mM concentration of imidazole. For further purification, the eluted fractions were loaded on anion exchange (high-Q sepharose) column, pre-equilibrated with buffer A1 (25 mM Tris-HCl pH 8.0, 100 mM NaCl). The proteins were eluted with linear gradient of NaCl in buffer A1. The human

translin-TRAX complex eluted as a single peak at ~325 mM NaCl concentration. The purified complex was incubated with DNase 1 and RNase A overnight at 20^{o}C followed by purification using Ni-IDA column.

A poly-histidine tagged truncated trax with 229 residues (6XHis-trax229; TRAX229), lacking large insertions compared to human translin observed in sequence alignment of translin and trax, was also expressed in *E. coli* BL21 (DE3) cells harboring *p*ET28a-*trx229*. The truncated TRAX229 protein was found in inclusion bodies alone. The inclusion bodies were solubilized with 8 M urea. The truncated protein was purified in denatured form using Ni-IDA column matrix in presence of 8 M urea. Attempts to refold the TRAX229 protein using dilution method or dialysis methods were unsuccessful. However complex of truncated TRAX229 with human translin (translin-TRAX229), synthesized from bacterial co-expression system, could be purified from soluble fractions of cell lysate with the similar protocol used to purify wild-type translin-TRAX complex. A heterologous complex of drosophila P168S mutant translin with human TRAX (dtsnP168S-TRAX) was also synthesized and purified from *E. coli* BL21 (DE3) cells co-transformed with *p*ET28a-*trx* and *p*ET21a-*dtsn168s* expression vectors using the similar protocol.

Gel filtration analysis

The Superdex^{TM} 200 10/300 GL column (GE Healthcare) coupled with high pressure liquid chromatography system (HPLC PU-2089 from JASCO, Japan) was used for molecular weight determination of the proteins/complexes. The Superdex^{TM} 200 column was pre-calibrated with gel filtration molecular weight markers. The purified proteins/complexes were loaded on the gel-filtration column in buffer [25 mM Tris-HCl

pH 8.0, 100 mM NaCl]. The eluted peaks of each independent gel filtration experiment were adjudged on SDS-PAGE.

DNA-binding assay

The DNA binding activity of translin-TRAX complexes were assayed with electrophoresis mobility shift assay (EMSA) using [γ-^{32}P] labeled Bcl-CL1 ssDNA probe. The probe (2000 nM) was 5′ end-labeled using T4 polynucleotide kinase (5 U/reaction) and [γ-^{32}P]ATP (50 μCi) in a 100 μL reaction buffer at 37 ºC for 30 minutes followed by the heat inactivation of the enzyme at 65 ºC for 30 minutes. The radiolabeled DNA probe was purified by spin column MicroSpin™ G-25 (GE-healthcare). The translin (50 nM of octameric human translin and 100 nM of tetrameric drosophila P168S mutant translin) and translin-TRAX complexes (50 nM of 300 kDa heteromeric complex) were incubated with the labeled-DNA probe (100 nM) in binding buffer B1 [25 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM EDTA] for 30 min in 20 μL volume. The reaction mixture was separated on a 4.5% native-PAGE in 0.5X TBE buffer. The gels were dried under vacuum and exposed to X-ray film overnight. The intensity of the shifted DNA-protein complexes was estimated using gel documentation system from Syngene (Synoptics, Cambridge, England). Pure translin variants were used as control in the gel-retardation experiments.

Circular Dichroism analysis and estimation of secondary structure

To estimate the secondary structure content of the trax protein, the far-UV circular dichroism (CD) spectra for the purified homomeric translin and translin-TRAX complexes were recorded with JASCO spectro-polarimeter (model J-810) in the

wavelength range of 200–260 nm using quartz cuvette of 1mm path length. Each sample was diluted to the concentration of 0.3 mg/mL to record the spectrum at similar protein concentrations. The observed spectra were deconvoluted into their secondary structure content using K2D2 (Perez-Iratxeta & Andrade-Navarro, 2008) software packages. The secondary structure for human trax was predicted using PSIPRED, a protein structure prediction server (McGuffin et al., 2000). The secondary structure of human translin was also analyzed from the crystal structure (PDB code, 1J1J).

Structure prediction of the trax protein

A BLAST search of the protein data bank (PDB) with the human trax protein sequence resulted in three structures with limited sequence homology; human translin (PDB code, 1J1J; Sugiura et al., 2004), mouse translin (PDB code, 1KEY; Pascal et al., 2002) and the partial structure of drosophila translin (PDB codes, 2QVA and 2QRX; Gupta et al., 2008). The human translin protein shares pair-wise sequence identity of nearly 22 % with human trax protein. The partial structural model (residues 34–264) of human trax was constructed by the comparative modeling approaches (Sali & Blundell, 1993) using an automated protein structure homology modeling server, SWISS-MODEL (Schwede et al., 2003; http://swissmodel.expasy.org/). The structural model was analyzed using Swiss-PdbViewer (Guex & Peitsch, 1997).

Results

Secondary and tertiary structure prediction of the trax protein

The human trax protein share nearly 22% sequence identity and 33% sequence similarity with human translin protein (Figure 4.8). The secondary structure for human

trax was predicted using PSIPRED, a protein structure prediction server (McGuffin et al., 2000). Trax was predicted to be all-helical protein and all of the helices were found to align with the observed helices in translin structure (Figure 4.8). The additional residues on the N- and C-terminal residues of trax, as compared to translin sequence, were predicted to adopt random coil conformation (Figure 4.8).

The partial structural model of human trax (Figure 4.9) was constructed using SWISS-MODEL server (Schwede et al., 2003). The server used human translin structure as a template and yielded partial structural model consisting of 34–264 residues of trax. The partial homology model seems to be reliable as the constructed structural model matches with the predicted secondary structure. Also the alignment of trax and template sequences used for threading matched well with the results of multiple sequence alignment of trax and translin orthologs. The model was further validated using circular dichroism and gel-filtration experiments.

To validate that the core structure of trax was similar to translin, we constructed a truncated trax, lacking additional N- and C-terminal residues and the 17-residue long insertion between α4 and α5 helices (loop-4) compared to translin sequence (Figures 4.8 & 4.9). This translin-like truncated trax had 229 residues, those aligned with translin sequence along with a Cys residue at position corresponding to Cys-225 of human translin. The Cys-225 was proposed to stabilize the translin homo dimers (Wu et al, 1998).

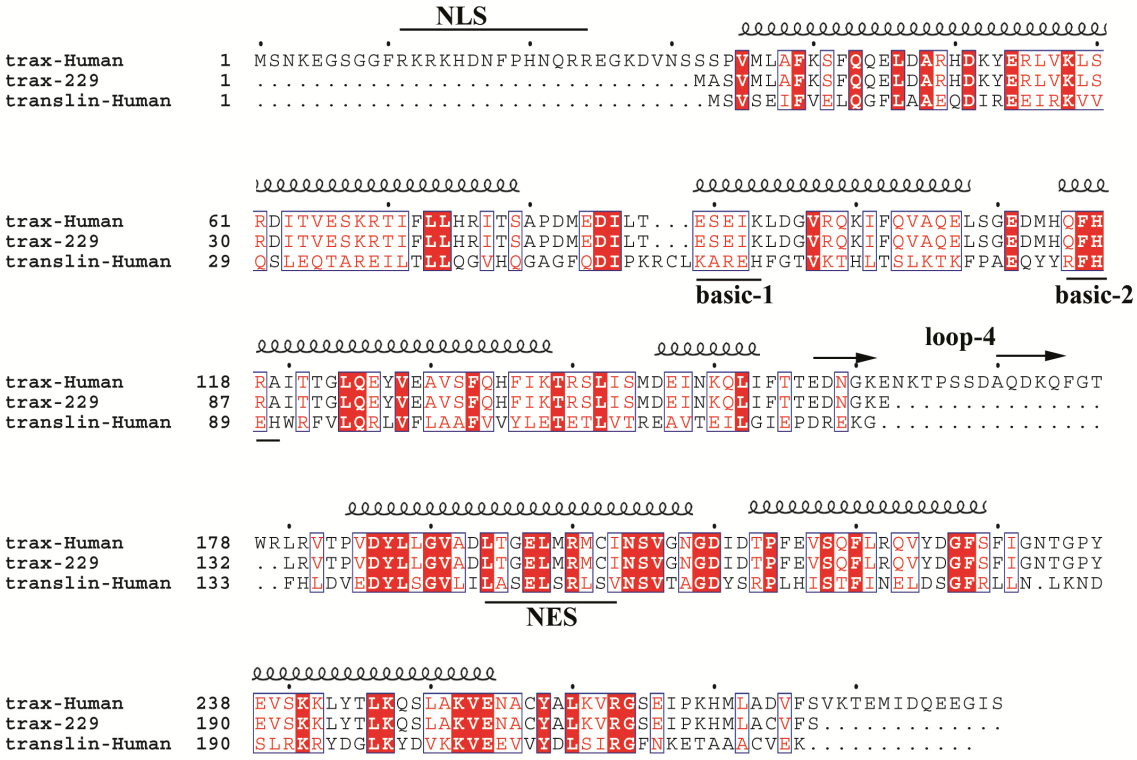


Figure 4.8 Alignment of trax, truncated trax (trax229) and translin sequences. The amino acid sequences (UniProtKB/Swiss-Prot entries) of the human trax and human translin (**Q99598** and **Q15631)** were aligned by ClustalW (Thompson et al., 1994). The identical amino acid residues are shaded and similar residues are boxed. The nuclear localization signal (NLS) on the trax sequence is identified. Also marked are two basic regions that have been implicated in nucleic acid binding (basic-1 and basic-2) and the nuclear export signal (NES) on translin sequence. The secondary structure of human trax as predicted from its amino acid sequence by PSIPRED server is shown as coils for helices and as an arrow for an extended conformation.

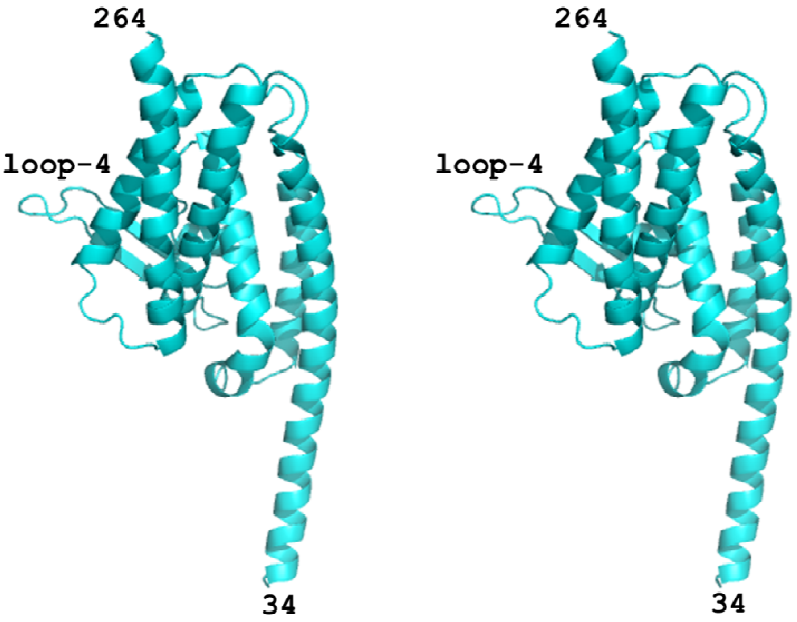
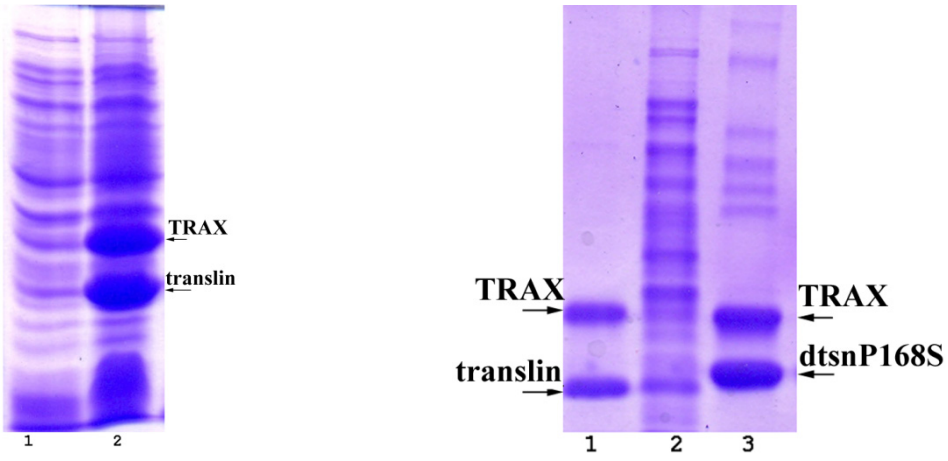


Figure 4.9 Cartoon of the partial structural model of the trax protein (stereo view). Model for residues 34–264 was constructed by comparative modeling methods. A large insertion of 17 residues in the trax sequence, compared to human translin, is marked as loop-4. The figure was prepared using PyMol suite (Delano, 2002).

Purification of translin-trax complexes

Trax is post-transcriptionally stabilized by translin (Yang et al., 2004) and recombinant trax expressed in bacterial expression system was found in insoluble fractions alone (Gupta et al., 2005). However, biologically active heteromeric translin-trax complex was purified from bacterial co-expression system (Gupta et al., 2005). We designed a co-expression system to purify heteromeric complex to homogeneity, free from any contamination of pure translin or trax. The human translin-TRAX complex was synthesized by co-expressing translin (without any additional residues or affinity tag) and 6XHis-trax (TRAX) in a bacterial expression system (Figure 4.10A). The heteromeric complex was purified with metal chelating affinity matrix followed by anion exchange chromatography. The purified translin-TRAX complex was free from

contamination of homomeric translin or TRAX as the homomeric translin, lacking any affinity tag, eluted as unbound fractions on metal chelating affinity column (lane 2; Figure 4.10B) while TRAX alone aggregated in inclusion bodies. The purified complex was estimated to have 1:1 molar ratio of translin and TRAX based on the relative intensities of the bands on 12% SDS-PAGE (lane 1; Figure 4.10B) as quantified by gel documentation system. A heterologous complex of drosophila P168S mutant translin with human trax (dtsnP168S-TRAX) was also purified with similar protocol (lane 3; Figure 4.10B). Interestingly, the truncated trax (TRAX229) folded properly upon co-expression with translin and a heteromeric translin-TRAX229 complex could be purified from soluble fractions of cell lysate. However, truncated trax aggregated into inclusion bodies when expressed alone.



(A) **(B)**

Figure 4.10 Expression and purification of translin-TRAX complex.

A) Expression of human translin and human TRAX proteins in the bacterial co-expression system. Lane 1, uninduced cells; lane 2, induced cells. **B)** Purification of human translin-TRAX complex. Lane 1 and lane 3 are purified human translin-TRAX complex and heterologous complex of drosophila P168S mutant translin with human trax (dtsnP168S-TRAX). Lane 2 unbound fraction from Ni-IDA column matrix during purification of human translin-TRAX complex. A band corresponding to human translin in unbound fractions was observed. Slow migrating bands in lane 3 correspond to higher-oligomers.

Gel-filtration analysis

The molecular weight of translin-TRAX complexes was determined with gel-filtration chromatography using Superdex^{TM} 200 10/300 GL column (GE Healthcare). The major peak of human translin-TRAX complex eluted at a corresponding molecular weight of ~290 kDa. A small fraction of very high molecular weight was also observed in the gel-filtration experiments (Figure 4.11A). The peak fraction of human translin-TRAX complex was analyzed on SDS-PAGE and was found to have 1:1 molar ratio of translin and TRAX. The data suggested that the heteromeric translin-TRAX complex of 290 kDa was best represented as octamer/decamer formed with four or five copies each of translin (Mr ~27.0 kDa) and TRAX (Mr ~35.5 kDa) proteins. The elution profile of heteromeric complex of human translin with translin-like trax (translin-TRAX229) was very similar to homomeric human translin and its molecular weight was estimated to be ~240 kDa corresponding to octameric oligomerization (Figure 4.11A).

To test whether trax on interactions disrupts the quaternary structure of translin oligomer we synthesized a heterologous complex of drosophila P168S mutant translin with human TRAX (dtsnP168S-TRAX). Interestingly, the elution profile of the heterologous dtsnP168S-TRAX complex was similar to human translin-TRAX complex despite that oligomeric state of drosophila mutant translin (tetramer, Mr~98 kDa) is very different from the human translin (Figure 4.11B). Consistent with high mass of drosophila translin (28 kDa, estimated from amino acid sequence), compared with that of human translin (27kDa, estimated), the molecular mass of the heterologous complex was marginally higher (~305 kDa) than that of human translin-TRAX complex.

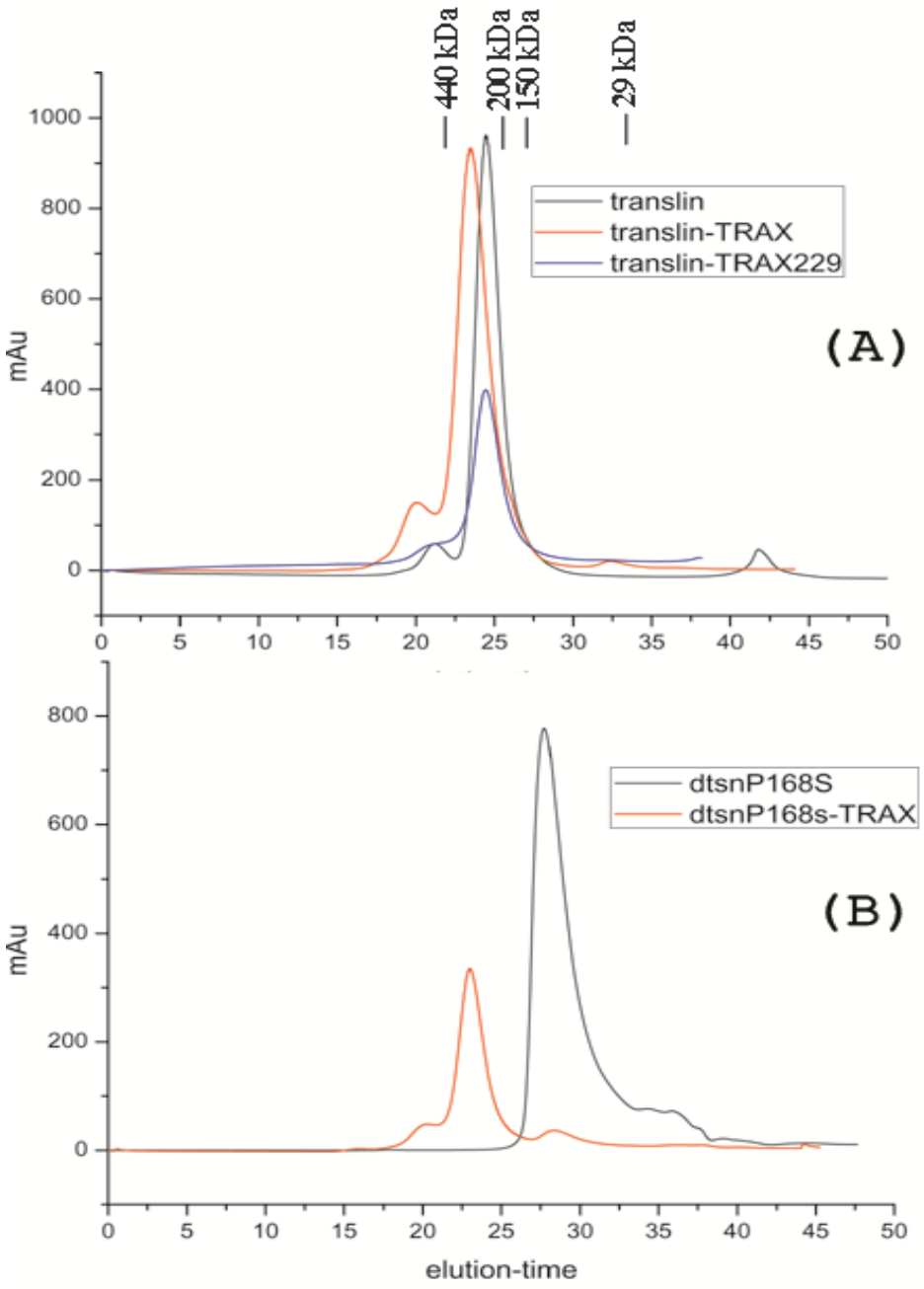


Figure 4.11 Molecular mass determinations of translin-TRAX complex using gel-filtration methods. The purified translin-TRAX complexes were loaded on the gel-filtration column at 0.5 mL/min flow rate. **A)** Elution profile of homomeric human translin, human translin-TRAX complex and complex of translin-like truncated trax with human translin (translin-TRAX229). **B)** Elution profile of drosophila P168S mutant translin (dtsnP168S) and heterologous complex of drosophila P168S mutant translin with human trax (dtsnP168S-TRAX). The molecular mass of the heterologous complex was determined to ~305 kDa, marginally higher than that of human translin-TRAX complex (~290 kDa).

DNA-binding assay

The DNA binding activity of DNase 1 and RNase A treated translin and translin-TRAX complexes were estimated by electrophoresis mobility shift assay (EMSA) using 5'-end radiolabeled 24-mer Bcl-CL1 ssDNA probe. The translin-TRAX complex was observed to bind with ssDNA (lane 3, Figure 4.12). The drosophila P168 mutant translin did not show the ssDNA binding activity (lane 4, Figure 4.12) as discussed earlier in chapter 3 of this thesis. Surprisingly, the heterologous complex of drosophila P168S mutant translin with human TRAX (dtsnP168S-TRAX) showed significant DNA binding activity (lane 5, Figure 4.12).

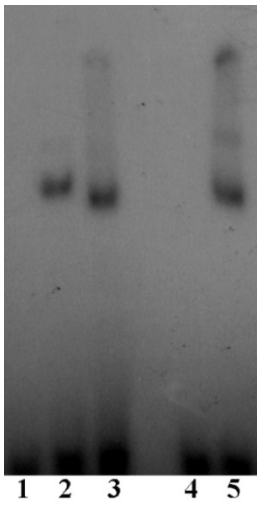


Figure 4.12 Analysis of DNA-binding activity of translin-TRAX complexes.
The DNA binding activity of translin-TRAX complex was assessed using electrophoresis mobility shift assay (EMSA) with radiolabeled ssDNA probe. Prior to EMSA analysis all the purified samples were treated with DNase 1 and RNase A overnight at 20^{o}C followed by purification using Ni-IDA chelating sepharose column. Lane 1, only DNA; lane 2, human translin; lane 3, human translin-TRAX complex (translin-TRAX); lane 4, drosophila P168S mutant translin (dtsnP168S); lane 5, heterologous complex between drosophila mutant translin and human TRAX (dtsnP168S-TRAX).

Circular Dichroism analysis and estimation of secondary structure

To estimate the secondary structure content of trax protein, the far-UV circular dichroism (CD) spectra for the purified homomeric translin and translin-TRAX complexes were recorded in the wavelength range of 200–260 nm (Figure 4.13). The CD spectrum of translin-TRAX complex was nearly identical to that of homomeric translin in the wavelength range of 200–260 nm. The similarity of CD spectra supported the structural model of trax and suggested that conformation of translin protein was not altered on interaction with TRAX. The CD spectrum of heterologous dtsnP168S-TRAX complex also matched well with that of human translin-TRAX complex.

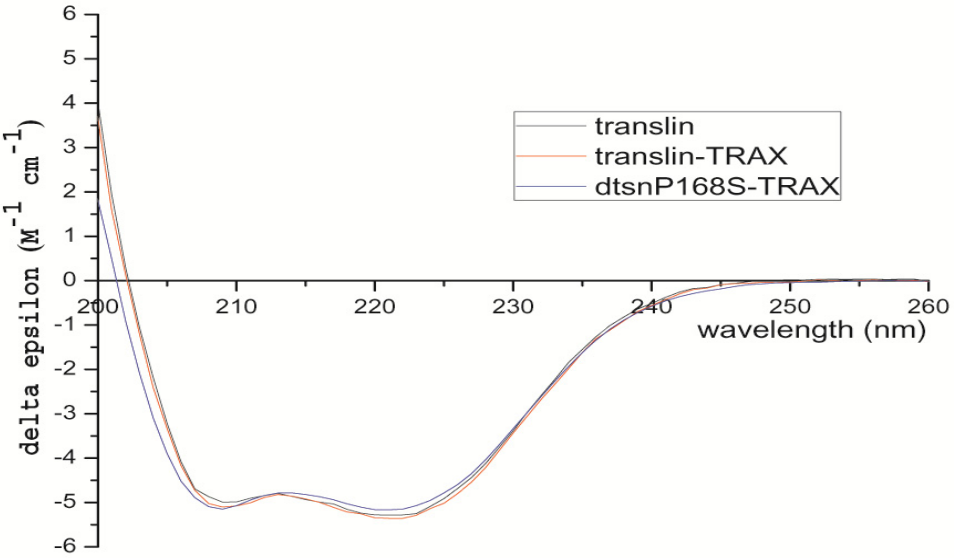


Figure 4.13 Circular Dichroism (CD) analysis. Circular Dichroism analysis of purified translin/translin-TRAX complexes (0.3 mg/mL each) in the wavelength range of 200–260 nm. The characteristic negative bands of α-helical proteins at 222 nm and 208 nm were observed for all the proteins/complexes. CD spectra of human translin (translin) and human translin-TRAX complex (translin-TRAX) and heterologous complex of drosophila P168S mutant translin with human trax (dtsnP168S-TRAX). The CD spectrum of human translin essentially overlapped with translin-TRAX complexes.

DISCUSSION

The trax protein was discovered as a binding partner of translin protein (Aoki et al., 1997; Wu et al., 1999a) and both proteins are highly conserved in eukaryotes. Both Trax and translin were observed in RNA binding complex purified from cell extracts of brain and testis (Finkenstadt et al., 2002). Trax is post-transcriptionally stabilized by translin in mouse, drosophila and S. pombe (Chennathukuzhi et al., 2003a; Claussen et al., 2006; Jaendling et al., 2008). All these data taken together suggests that trax and translin interact closely with each other and translin supports stability of trax fold.

To characterize trax protein, we synthesized human translin-TRAX complex using bacterial co-expression system by co-expressing human translin with human 6XHis-trax (TRAX). The heteromeric complex was purified to homogeneity with very high yield. Translin and TRAX proteins were observed to have 1:1 stoichiometry in the heteromeric complex. The molecular mass of heteromeric translin-TRAX complex was estimated to ~290 kDa using size exclusion chromatography. The estimated molecular mass of the recombinant translin-TRAX complex matched closely to the molecular mass of translin-trax complex (240 kDa) extracted from rat liver (Wu et al., 2003). The recombinant translin-TRAX complex was observed to bind with ssDNA.

Core structure of trax is similar to translin

The secondary structure prediction server predicted trax to be all helical and the predicted helical region of trax sequence matched well with the experimentally determined secondary structure of translin. The partial tertiary structure (34–264 residues) of trax was constructed with homology modeling based method using translin structure as template. The modeled structure of trax was observed to have an additional

loop (loop-4) between helices α4 and α5, compared to translin structure, that is due to a 17-residue long insertion in trax sequence. The nearly identical CD spectra of homomeric translin and heteromeric translin-TRAX complex supported the structural model of trax. Based on the predicted model we constructed a truncated trax having 229 residues those matched with core structure of translin. This translin-like trax, however, was insoluble when expressed alone, yet the truncated trax was found to be properly folded when co-expressed with translin and behaved like translin octamer on gel-filtration column. The analogous behavior on molecular sieve column and identical CD spectrum of translin-TRAX229 complex also strongly supports the model that core structure of trax matches with translin. Intriguingly, the modeled structure of trax matches well with the recently published crystal structure of translin-trax complex (Tian et al., 2011; Ye et al., 2011).

Translin and trax co-oligomerize to form a heteromeric complex

The molecular mass (290 kDa) of heteromeric translin-TRAX complex suggested that it forms an octamer/decamer with four or five copies each of translin and TRAX. Since homomeric human translin exist as an octamer, it suggested that trax disrupts the octameric state of translin or both the proteins co-oligomerize *in situ*. To further understand the mechanism of formation of complex, we synthesized heterologous complex of drosophila P168S mutant translin with human TRAX (dTsnP168S-TRAX). Surprisingly, the heterologous complex dTsnP168S-TRAX was found to be DNA-binding competent. Notably, the drosophila P168S mutant translin is DNA binding incompetent due to its open doughnut shape stable tetramer oligomerization, compared to active octamer of human translin (Gupta et al., 2008). Yet, the CD and gel filtration

characteristics of the heterologous complex matched with those displayed by wild type translin-TRAX complex. Similar results would not be expected if trax recruits to a pre-formed dTsnP168S tetrameric complex that is DNA binding incompetent. On the other hand, if the monomeric form of the dTsnP168S mutant co-oligomerizes with trax to form heterodimers, probably the up-down heterodimer, these heterodimers can be assembled into a complex analogous to human translin-trax complex. The up-down dimer is the most likely confirmation, as it is the stable structure unit in translin-like proteins (discussed in chapter 3; Gupta et al., 2008).

In conclusion, we established protocol for the large-scale synthesis of DNA-binding competent translin-trax complexes using bacterial co-expression system and provided evidence that the structure of trax is very similar to translin. We also showed that translin and trax proteins co-oligomerize *in situ* to form a stable and DNA-binding competent heteromeric complex. Most likely, the translin functions as chaperone for the stable fold of trax.

Chapter 5

Identification of nucleic acid binding sites on translin-associated factor X (trax) protein

INTRODUCTION

Trax protein was initially identified in a yeast two hybrid screen as a protein that interacts with translin, a ssDNA and RNA binding protein (Aoki et al., 1997; Wu et al., 1999a; Aoki et al., 1995). Translin and trax proteins share homology and have been evolutionarily conserved in eukaryotes. Deletion of *translin* in mice leads to complete loss of trax protein, without affecting trax mRNA levels, indicating that the stability of trax protein is dependent on its interaction with translin (Chennathukuzhi et al., 2003a; Yang et al., 2004). Similar results have been reported in *Drosophila melanogaster* (drosophila) and *Schizosaccharomyces pombe* (S. pombe) demonstrating that the close interaction between translin and trax has been conserved over the course of evolution (Claussen et al., 2006; Jaendling et al., 2010). The translin-trax complex binds to both ssDNA and RNA and has been implicated in a variety of cellular functions including DNA repair, spatial and temporal regulation of mRNA translation, dendritic trafficking of BDNF mRNA, and siRNA-mediated silencing (Erdemir et al., 2002a; Chiaruttini et al., 2009; Liu et al., 2009).

To understand how translin-trax complex exerts its cellular effects, it would be helpful to define how it binds to ssDNA or RNA. Initial characterization of translin and trax revealed that translin, but not trax alone, is able to form a homomeric complex capable of binding ssDNA or RNA (Aoki et al., 1999). Two basic regions ^{60}KAREH^{64} and ^{86}RFHEH^{90}, designated basic-1 and basic-2, have been identified on human translin sequence and mutation in basic-2 region caused the complete loss of DNA/RNA binding activity (Aoki et al., 1999). However, translin constructs with mutations in the basic-2 region are able to form heteromeric complexes with trax that are fully active in gel-shift binding assays (Chennathukuzhi et al., 2001a). Also,

heteromeric complex between wild-type translin and trax proteins showed stronger binding to DNA than RNA, however, trax on its own is believed not bind to RNA or DNA (Chennathukuzhi et al., 2001a; Liuis et al., 2010).

We have conducted studies aimed at deducing nucleic acid binding characteristics of trax and in defining the binding sites. Our results show that trax binds to ssDNA and two regions in trax, designated B2 and B3, mediate nucleic acid binding activity of the heteromeric complex. We also identified a new DNA-binding motif (B3) on translin sequence, corresponding to B3 motif of trax, which contributes towards DNA-binding activity of translin.

METHODS

Prediction of DNA-binding residues

The DNA-binding sites in human trax and translin sequences were predicted using the DP-Bind web server (Hwang et al., 2007; http://lcg.rit.albany.edu/dp-bind/). The analysis was performed with PSSM based encoding which uses a position-specific scoring matrix generated by PSI-BLAST (Altschul et al., 1997). The DNA-binding residues were identified from the consensus that retained only high confidence predictions and were categorized in five motifs (B1–B5) on trax and translin sequences.

Site directed mutagenesis

The *p*QE9-*tsn* and *p*ET28a-*trx* constructs, used to express human translin (translin) and human trax with 6XHis tag (TRAX), were those used in earlier studies (Gupta et al., 2005). Mutants of the putative nucleic acid binding motifs B2 and B3 on the TRAX

sequence were generated by Quick-Change site-directed mutagenesis kit (Stratagene, USA). The *p*ET28a-*trxB2* construct, used to express B2 region mutant of human trax with N-terminal 6XHis tag (TRAXB2; ^{115}**Q**F**HR**A^{119} to ^{115}**L**F**NA**A^{119}), was synthesized using complimentary primers pairs (CAGGGGAAGATATGCATCTGTTCAATGCAGCCATTACTACAGG and CCTGTAGTAATGGCTGCATTGAACAGATGCATATCTTCCCCTG) and *p*ET28a-*trx* as template. Similarly, *p*ET28a-*trxB3* construct, used to express B3 region mutant of human trax (TRAXB3; ^{241}**KK**L**Y**^{244} to ^{241}**NT**L**N**^{244}), was synthesized using complimentary primers pairs (CCTTACGAGGTTTCTAATACGCTGAATACCTTG AAACAAAG and CTTTGTTTCAAGGTATTCAGCGTATTAGAAACCTCGT AAGG) and *p*ET28a-*trx* as template.

The *p*QE9-*tsnB2* construct, used to express homomeric human translin (6XHis tagged) with substitution in basic-2 region (translinB2, ^{86}**R**F**H**E**H**^{90} to ^{86}**T**F**N**E**N**^{90}), was generated by the PCR overlap extension method using mutagenizing primer pair (GCTGAACAGTATTACACATTTAATGAGAACTGGAGGTTTGTGTTG and CAACACAAACCTCCAGTTCTCATTAAATGTGTAATACTGTTCAGC) along with vector specific primers using *p*QE9-*tsn* as template. Similarly, the *p*QE9-*tsnB3* construct, used to express homomeric human translin (6XHis tagged) with substitution in basic-3 region (translinB3; ^{192}**RKRY**^{195} to ^{192}**TNSN**^{195}), was also generated by the PCR overlap extension method using mutagenizing primer pairs (CCTGAAAAATGACTCCCTGACGAACAGCAACGACGGATTGAAATATGACG and CGTCATATTTCAATCCGTCGTTGCTCGTCAGGGAGTCATTTTTCAGG).

The constructs *p*ET21a-*tsn,* *p*ET21a-*tsnB2* and *p*ET21a-*tsnB3*, were also synthesized by cloning mutant and wild-type translin genes into the *Nde*I and *Bam*HI

sites of *p*ET21a plasmid to express translin proteins without any tag, to be used for co-expression system. The nucleotide sequences of all the cloned genes were confirmed by DNA sequencing using an ABI automated DNA sequencer. A summary of all the constructs used for testing DNA-binding activity is listed in Table 5.1.

Expression and purification of the proteins

The *E. coli* BL21(DE3) cells harboring *p*ET21a-*tsn* were co-transformed with *p*ET28a-*trx*, *p*ET28a-*trxB2* or *p*ET28a-*trxB3* constructs to express the complexes of wild-type human translin with wild-type or mutant 6XHis-trax (TRAX) proteins. Similarly, the complexes of human mutant translin (translinB2 or translinB3) with TRAX (wild type or mutants) were expressed by co-transforming the *p*ET28a-*trx*, *p*ET28a-*trxB2* or *p*ET28a-*trxB3* into the *E. coli* BL21 (DE3) cells harboring *p*ET21a-*tsnB2* or *p*ET21a-*tsnB3* construct. These all 9 complexes were expressed as 6XHis-TRAX complexed with untagged translin. The co-transformants were selected and maintained using two antibiotics, based on the resistances conferred by the transforming vectors. The homomeric translin proteins (wild-type translin, translinB2 mutant or translinB3 mutant) were expressed with N-terminal 6XHis tag in *E. coli* BL21(DE3) cells harboring *p*QE9 constructs (*p*QE9-*tsn, p*QE9-*tsnB2* or *p*QE9-*tsnB3*). The expression of translin and TRAX proteins was induced by IPTG. Translin proteins or translin-TRAX complexes (Table 5.1) were purified using chelating-sepharose FF column (Ni-IDA) followed by anion exchanger high Q-sepharose column (as discussed in chapter 4). The purified proteins were incubated with DNase 1 and RNase A overnight at 20^{o}C followed by purification using Ni-IDA column. The purification protocol for heteromeric complexes ensured recovery of only the translin-TRAX complexes from

the affinity matrix because translin was untagged and poly-His tagged TRAX was insoluble when expressed alone.

Gel filtration and Circular dichroism analysis

The purified translin proteins and translin-TRAX complexes were loaded on the Superdex^{TM} 200 10/300 GL column (GE Healthcare) for final purification step as well as for molecular weight determination. The Superdex^{TM} 200 column was calibrated with gel filtration molecular weight markers (Amersham-Pharmacia; Carbonic anhydrase, 29 kDa; Alcohol dehydrogenase, 150 kDa; β-amylase, 200 kDa; Ferritin, 440 kDa). The eluted peak of each independent gel filtration experiment was adjudged on SDS-PAGE. The purified proteins were concentrated to 5–7 mg/mL. The protein concentrations were estimated by modified Lowry's method using bovine serum albumin (BSA) as the standard.

The far-UV circular dichroism (CD) spectra for the purified translin proteins (0.3 mg/mL) and translin-TRAX complexes (0.3 mg/mL) were recorded in the wavelength range of 200–260 nm at 25^{o}C with JASCO spectro-polarimeter (model J-810). Each spectrum was averaged for three scans.

DNA-binding assay

Electrophoresis mobility shift assays (EMSA) were performed to assess DNA-binding-activity of the DNase 1 and RNase A treated proteins/complexes using Bcl-CL1 ssDNA (GCCCTCCTGCCCTCCTTCCGCGGG) as probe. The molecular masses of the heteromeric complexes used were ~300 kDa, whereas oligomers of human translin were 236 kDa, respectively. The probe was 5′ end-labeled using T4 polynucleotide

kinase (3 Units/reaction) and [γ-^{32}P]ATP in a 50 µL reaction buffer at 37 ºC for 60 minutes followed by heat inactivation of the enzyme at 65 ºC for 30 minutes. The radiolabeled probe was purified by spin column MicroSpin™ G-25 (GE Healthcare). Translin (50 nM of octameric translin) and translin-TRAX complexes (100 nM of 300 kDa translin-TRAX complex) were incubated with the labeled-DNA probe (100 nM in the reaction mixture) in binding buffer B1 [25 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM EDTA] for 30 minutes. The 10 µL of reaction mixtures were separated by electrophoresis on a 4.5% native-PAGE gel in 0.5X TBE buffer. The dried gels were exposed to X-ray film overnight. The intensity of the shifted DNA-protein complexes was estimated using a gel documentation system from Syngene (Synoptics, Cambridge, England).

The identity of proteins in gel-shifted bands from the EMSA analysis performed with 100-fold higher concentrations of DNA and proteins was adjudged by SDS-PAGE analysis. The DNA in this experiment was stained with ethidium bromide and varying concentrations of translin proteins (1 and 5 µM) and translin-TRAX complexes (2 and 10 µM) were incubated with 10 µM of the 24-mer unlabeled DNA oligo in 20 µL reaction volume for 30 minutes. The products were resolved by electrophoresis on 1.5% agarose gel and stained with ethidium bromide. The agarose regions containing gel-shifted bands were excised and adjudged on 12% SDS-PAGE, which confirmed that both translin and TRAX proteins were components of translin-TRAX:DNA complexes.

Photochemical crosslinking and purification of TRAX-DNA covalent complex

To detect the direct binding of ssDNA with TRAX, the radiolabeled DNA probe was crosslinked with translin-TRAX complex using KrF excimer UV laser (COMPexPro 205) at 248 nm. Translin (50 pmol of octameric translin) and translin-TRAX complexes (50 pmol of the 300 kDa heteromeric complex) were individually incubated with 50 pmol of labeled-DNA probe for 30 min in 50 µL binding buffer (25 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM EDTA). Each of the protein:DNA mixtures were irradiated in a quartz cell (Eppendorf UVette) for 1 minute at 1 pulse/sec imparting about 2400 mJ of total energy. BSA:DNA mixture was also irradiated under identical conditions as a control. To isolate TRAX-DNA crosslinked complex, the irradiated heteromeric complex was disrupted in the presence of 8M urea followed by 5 minutes incubation in boiling water bath. The reaction mixture was loaded onto Ni-IDA spin-column pre-equilibrated with binding buffer (25 mM Tris-HCl, pH 8.0, 8M urea, 50 mM imidazole and 100 mM NaCl). The matrix beads were extensively washed with binding buffer. The bound protein was eluted with elution buffer (25 mM Tris-HCl pH 8.0, 8 M urea, 1 M imidazole and 100 mM NaCl). This protocol resulted in purification of denatured 6XHis-trax (TRAX) and its covalent complexes as the poly-His tag was fused only with trax sequence. The purification of TRAX alone was further confirmed by mass spectrometry of the purified protein component obtained by disrupting pure translin-TRAX complex in an independent experiment. The TRAX-DNA complex along with irradiated translin-DNA complex and prestained molecular weight markers were resolved on 12% SDS-PAGE. The dried gel was autoradiographed on X-ray Film. The positions of TRAX-DNA and translin-DNA complexes on autoradiograph of the SDS-PAGE were estimated from the mobility of the prestained marker proteins. To

adjudge the extent of protein-protein crosslinking, purified translin was mixed with [α-^{32}P]GTP and the mixture was irradiated for 2400 mJ laser energy under conditions identical to that for protein-DNA complexes.

The heteromeric translin-TRAX complex (50 pmol) was also UV-crosslinked with 50 pmol of the 24-mer unlabeled Bcl-CL1 and 43-mer unlabeled DNA oligo (CCTGTAGTAATGGCTGCATTGAACAGATGCATATCTTCCCCTG) under similar conditions as described above. Post-irradiation, TRAX-DNA crosslinked complexes were purified after disruption with chaotropic agents following the procedure as used for purification of 24-mer labeled DNA covalent complex. The covalent complexes of TRAX with 24 and 43-mer ssDNA were resolved on 12% SDS-PAGE. The protein bands were stained with silver-stain to analyze supershift of the TRAX-DNA complex due to higher molecular mass of 43-mer DNA.

Crystallization and structural analysis of translin-DNA complex

Human translin protein was purified with gel-filtration chromatography prior to crystallization experiments. The peak fractions corresponding to 235 kDa molecular weight were harvested and concentrated to 15 mg/mL using ultra-filtration method. The translin protein was incubated with ssDNA probe Bcl-CL1 in 1:2 molar ratios (one octamer of translin versus 2 molecules of 24-mer ssDNA). The translin-DNA complex was crystallized using hanging drop vapor diffusion method. Crystals of size 400 x 200 x 100 μm^3 could be grown in condition C1 [2.5 M sodium format, 0.1 M sodium acetate pH 4.7, 1 mM EDTA] at room temperature. The crystals were transferred to cryo-protectant solution [2.5 M sodium format, 0.1 M sodium acetate pH 4.7, 1 mM EDTA and 30% sucrose]. The three dimensional diffraction intensity data for translin-

DNA complex crystals were recorded on a CCD detector using BM14-U beam-line of European synchrotron radiation facility (ESRF, France) at 100K with 0.5^{o} oscillation per image. The crystals diffracted to 2.45 Å resolution. The intensity data were processed and scaled with HKL2000 **(**Otwinowski & Minor, 1997**)**. The space group for orthorhombic crystals was identified to be $C222_1$ with unit cell parameters a=127.2, b=133.2, c=135.1 Å. The scaled intensities were truncated with CTRUNCATE program of CCP4 software package (CCP4, 1994). The unit cell parameters, however, were similar to those observed for crystals of the native protein (PDB ID 1J1J; Sugiura et al., 2004). The difference Fourier maps were computed with the atomic coordinates of native-translin structure. Refinement of the model was performed using REFMAC5 (Vagin et al., 2004).

RESULTS

Trax is DNA-binding competent

To detect the direct binding of DNA with trax, the radiolabeled ssDNA probe was crosslinked with homomeric translin and heteromeric translin-TRAX complex (a complex of human translin with 6XHis-trax) using UV-laser. The BSA:DNA mixture was used as a negative control in the crosslinking experiment. For unambiguous detection of TRAX-DNA covalent complex, the irradiated heteromeric translin-TRAX complex was disrupted by boiling with 8M urea. TRAX/TRAX-DNA complex was purified with Ni-IDA matrix spin-column using poly-His tag in TRAX. The protocol resulted in purification of TRAX or its covalent complexes, as verified by mass spectrometry. The eluted fraction of the disrupted heteromeric complex containing TRAX and TRAX covalent complexes and irradiated translin:DNA mixture were

resolved on 12% SDS-PAGE. The translin-DNA and TRAX-DNA complexes were clearly detected on autoradiogram. The TRAX-DNA complex (lane 3, Figure 5.1), migrates close to the 44 kDa marker protein and thus corresponds to TRAX molecule (35.3 kDa) covalently linked with one ssDNA moiety (~7.6 kDa). As expected, yield of the covalent complex was very poor. A band of very weak intensity (relative to the unliganded protein) was observed on silver-stained gel SDS-PAGE (Figure 5.2). The migration of this band corresponded to difference in the molecular masses between nucleic acid complexed and uncomplexed TRAX protein. The UV-crosslinked heteromeric complex with 43-mer ssDNA probe in an independent experiment showed expected shift of ~6 kDa in the TRAX-DNA band (Figure 5.2). As compared to TRAX-DNA covalent complex observed in autoradiogram, translin-DNA complex migrates at about 36 kDa (lane 2, Figure 5.1). No band could be detected in lane 4 loaded with BSA:DNA irradiated sample, suggesting that crosslinking of DNA with TRAX was not due to artifact A very high molecular mass complex was also detected in both the lanes. Similar high molecular mass band was also detected at much lower concentration in the translin:[α-^{32}P]GTP mixture that was irradiated under same conditions.

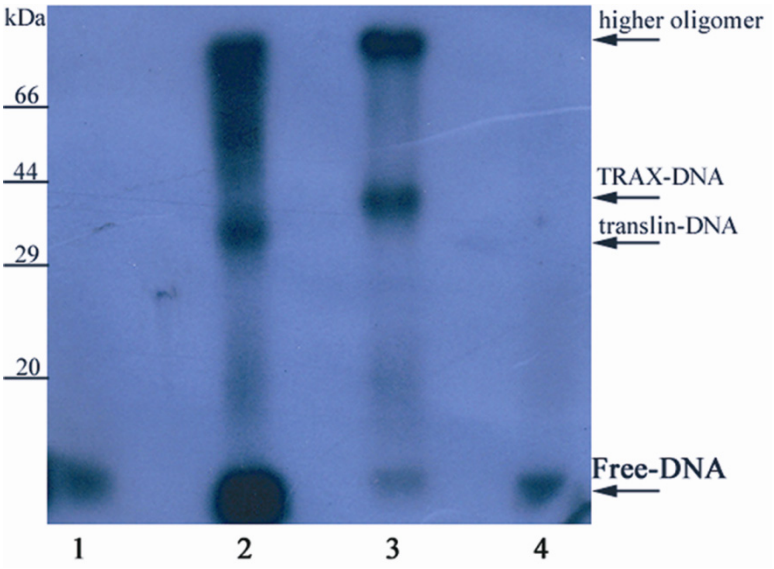


Figure 5.1 Autoradiograph showing protein-DNA crosslinking. Purified translin, translin-TRAX complex and BSA were mixed with [γ-^{32}P]-labeled Bcl-CL1 24-mer ssDNA. The mixtures were irradiated with UV-laser imparting total of 2400 mJ energy of 248 nm wavelength. Post-irradiation, the translin-TRAX complex was disrupted and TRAX was purified using affinity tag. The protein:DNA complexes (translin-DNA, lane 2; TRAX-DNA, lane 3, BSA:DNA, lane 4) and DNA alone (lane 1) were resolved on 12% SDS-PAGE and dried gel was autoradiographed on X-ray film. The translin-DNA and TRAX-DNA complexes differ by expected 6 kDa in molecular mass. The BSA, used as a negative control, did not show presence of protein-DNA covalent complex in the post-irradiated mixture. The intensities of free DNA band in lane 1 and 4 are much weaker, as only 5 µL of the irradiated BSA:DNA mixture was loaded, compared to 20 µL of the translin:DNA mixtures.

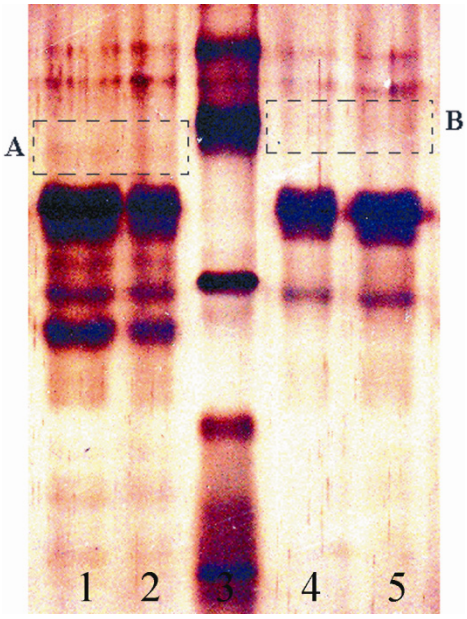


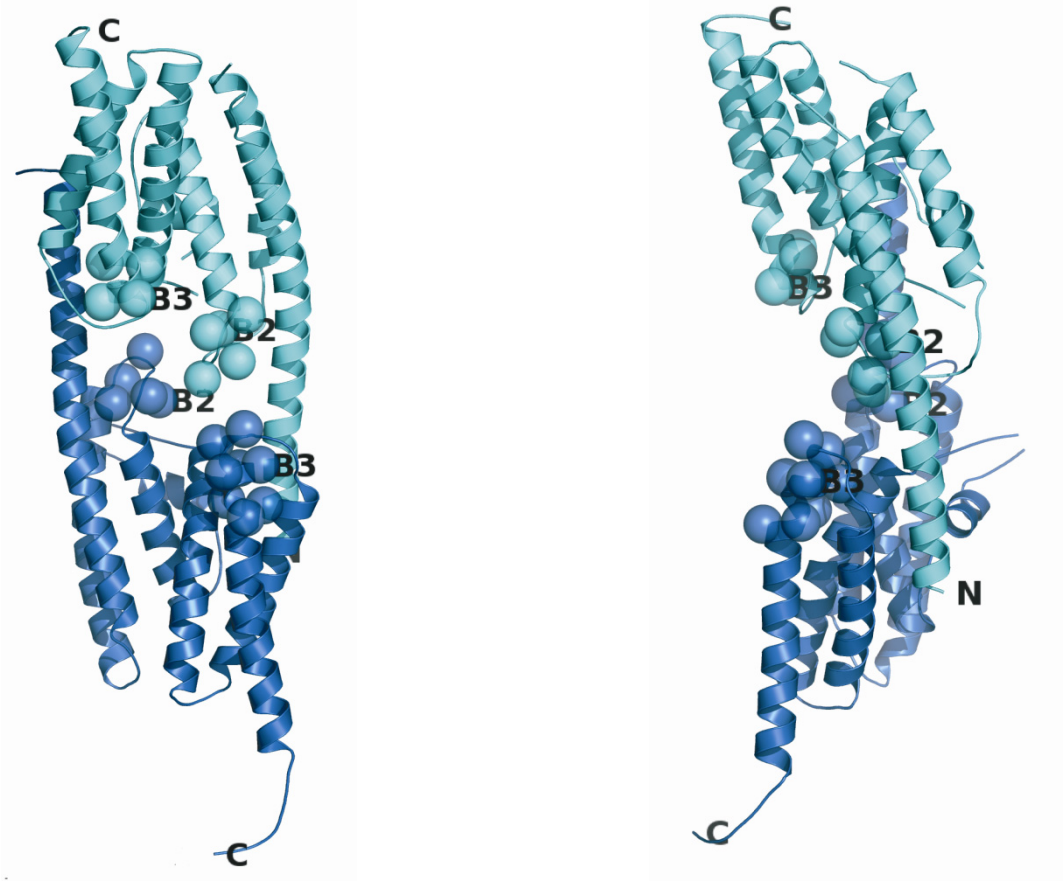
Figure 5.2 Supershift analysis of TRAX-DNA complex. To confirm TRAX-DNA crosslinking, a 43-mer DNA was incubated with translin-TRAX complex and was UV-irradiated. Post-irradiation, the complex was disrupted using chaotropic agents and TRAX-DNA complex alone was purified using poly-His tag available on TRAX sequence alone. The covalent complexes of TRAX with 43-mer DNA (lanes 4 and 5), with 24-mer DNA (lanes 1 and 2) and molecular weight markers (lane 3) were resolved on 12% SDS-PAGE. The protein bands were stained with silver-stain. The relative migration of covalent TRAX-DNA complexes (boxed, weak band intensity suggests poor yield of the covalent complex) corresponded to the molecular mass difference between 43- and 24-mer DNA.

Figure 5.3 Predicted DNA binding motifs on translin and TRAX sequences. The amino acid sequences of the TRAX and translin proteins of human and drosophila (UniProtKB/Swiss-Prot **Q99598**, **Q8INE1**, **Q15631** and **Q7JVK6,** respectively) were aligned by ClustalW. For clarity, only sequences of human TRAX (TRAX) and human translin (translin) are shown. The identical amino acid residues are shaded and similar residues are boxed. The five putative DNA-binding regions (B1–5) predicted by the DP-Bind server and the nuclear localization signal (NLS) on the TRAX sequence are identified. Also marked are five basic regions (B1–5) on translin sequence and its nuclear export signal (NES).

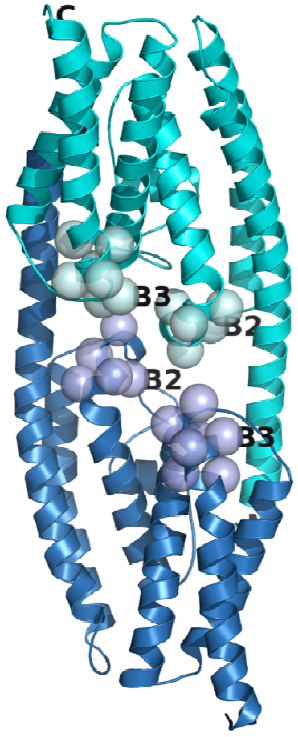
Prediction of DNA-binding sites on trax and translin

Candidate DNA-binding residues in the human trax protein were identified based on the consensus sites detected by the DP-Bind web server. This algorithm predicted most of the 22 N-terminus amino acids as DNA-binding residues, as well as five additional DNA-binding motifs (B1, ^{75}HRITS^{79}; B2, ^{115}QFHRA^{119}; B3, ^{237}YEVSKKL^{243}; B4, ^{247}KQSLAK^{252} and B5, ^{261}KVRGS^{265}). The five motifs (B1–5) were also predicted to be present in human translin sequence (Figure 5.3). The previous studies had demonstrated that substitutions in B2 and B3 motifs in translin, identified here, resulted

in loss of ssDNA-binding activity [Aoki et al., 1999; Eliahoo et al., 2010]. The B2/B3 motifs of trax are disposed similar to those of translin monomers in the structure of translin-TRAX heterodimer (Figure 5.4). Accordingly, to test the role of the B2 and B3 motifs of trax in nucleic acid binding activity of the heteromeric translin-TRAX complex, we generated trax mutants with substitution in predicted B2 or B3 motifs as follows: TRAXB2 - ^{115}QFHRA^{119} to ^{115}LFNAA^{119}; TRAXB3 - ^{241}KKLY^{244} to ^{241}NTLN^{244}.



(A) **(B)**



(C)

Figure 5.4 Different views (A,B) of cartoon showing close proximity of B2 and B3 motifs (spheres) of translin (cyan) and TRAX (blue) in translin-TRAX heterodimer and translin-translin homodimer (C). Figures (A & B) were prepared using atomic coordinates of translin-TRAX complex structure (PDB code, 3PJA), and Figure (C) was prepared with atomic coordinates of human translin structure (PDB code, 1J1J).

Oligomeric status and structural content of mutant translin-TRAX complexes

Prior to testing the impact of these mutations in TRAX on the nucleic acid binding activity of the translin-TRAX complex, we checked whether they altered its size or composition. The molecular masses of the heteromeric mutant complexes formed by translin variants with either wild type TRAX or TRAXB2 or TRAXB3 constructs were determined to be ~300 kDa by gel-filtration chromatography, nearly identical to that of wild-type heteromeric translin-TRAX complex (Figure 5.5). The eluted proteins were

adjudged on 12% SDS-PAGE. The intensities of translin and TRAX bands showed nearly equimolar ratio of the two proteins in the heteromeric complexes (Figure 5.6). By comparison, the molecular mass of homomeric translin complexes generated by either wild-type translin or translin constructs containing mutations in B2 or B3 domains (^{86}RFHEH^{90} to ^{86}TFNEN^{90}; ^{192}RKRY^{195} to ^{192}TNSN^{195}) were determined to be approximately 236 kDa, consistent with earlier reports that mutations in the basic-2 (B2) motif do not affect the octameric status of human translin (Aoki et al., 1999). The migration of translin, translinB2 and translinB3 on gel-electrophoresis under native conditions was also observed to be similar (Figure 5.7).

We compared the CD spectra of homomeric translin and heteromeric translin-TRAX complexes and found excellent agreement between CD spectra of heteromeric and homomeric complexes in the wavelength range of 200–260 nm with the known crystal structures of translin and translin-TRAX complexes (Figure 5.8). The observed characteristic negative bands of α-helical proteins at 222 nm and 208 nm are in conformity with all-helical structure of translin and TRAX proteins. The CD analysis also suggests that mutations in the nucleic acid binding motifs of translin and TRAX do not alter the secondary structural content of the complexes.

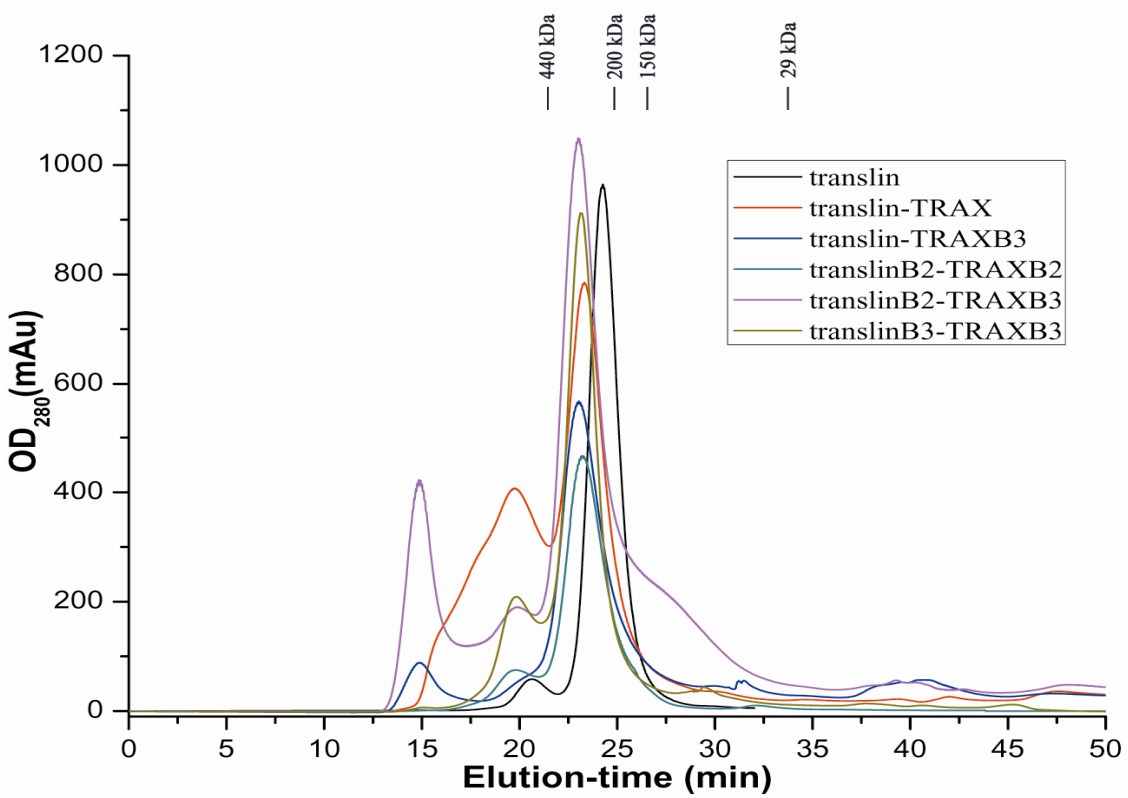


Figure 5.5 A profile overlay of Superdex 200 gel-filtration chromatography of human translin and translin-TRAX complexes. The major peaks corresponding to molecular masses of 295 kDa for the complexes and 236 kDa for the translin were used for DNA-binding assays.

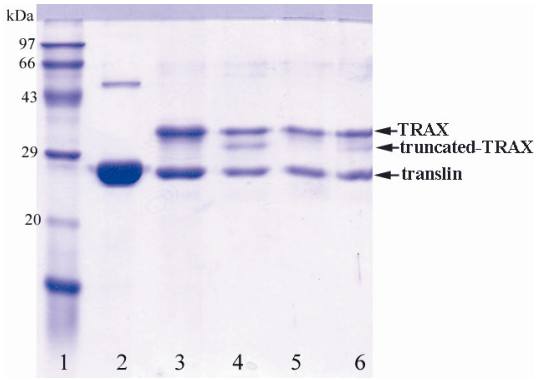


Figure 5.6 The translin/TRAX protein complexes purified by gel-filtration chromatography were adjudged on SDS-PAGE. Molecular weight markers, lane 1; human translin, lane 2; translinB2-TRAXB3 complex, lane 3; translin-TRAXB3 complex, lane 4; translinB2-TRAXB2 complex, lane 5 and wild-type translin-TRAX complex, lane 6. Nearly equimolar stoichiometry of translin and TRAX proteins in the heteromeric complexes could be estimated from the band intensities. Truncation in TRAX protein was observed on storage. The truncation site at the C-terminus was confirmed by N-terminal sequencing and MALDI-TOF analyses. Integrated intensities of both the TRAX bands were summed to estimate translin/TRAX stoichiometry.

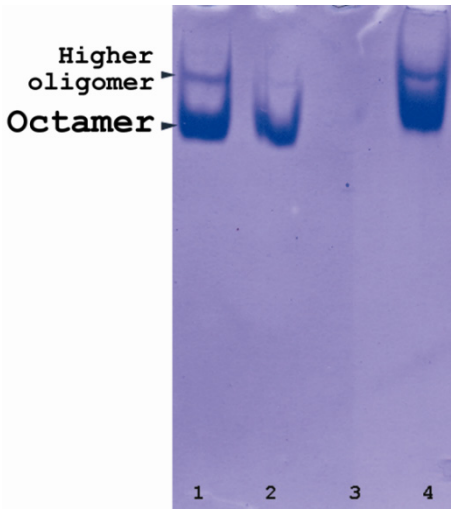


Figure 5.7 Native-PAGE analysis of translin proteins. Human translin and its mutants were resolved on 4.5% polyacrylamide gel under native conditions (Lane 1, translin; lane 2, translinB3; lane 4, translinB2). The proteins purified by three-stage chromatography were those eluted at molecular mass of about 236 kDa from Superdex 200 column. The native-PAGE showed presence of low-abundance high molecular mass oligomer.

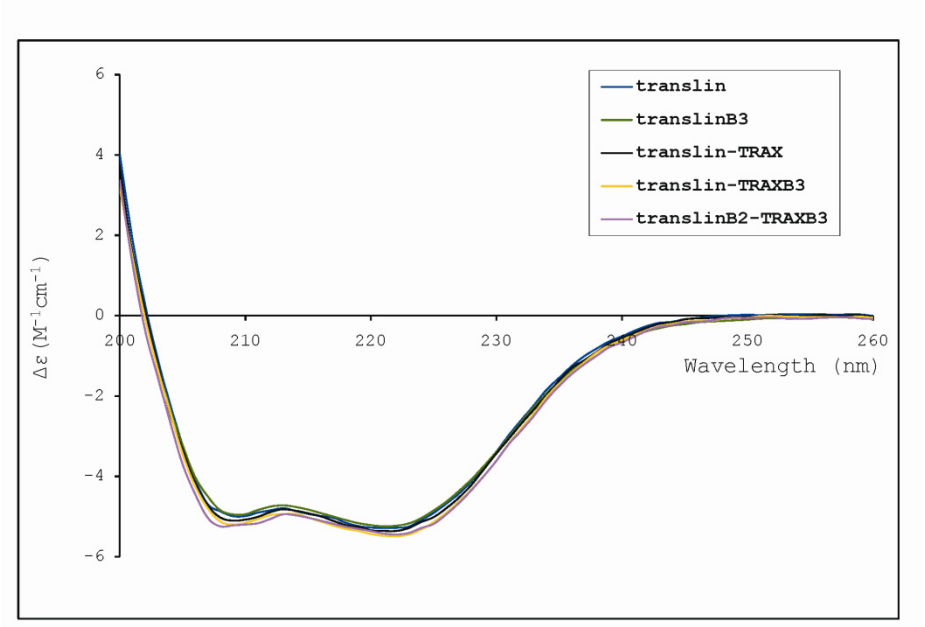


Figure 5.8 Circular dichroism (CD) analysis of the proteins in the wavelength range of 200–260 nm. The CD spectrum of human translin essentially overlaps with those of translin-TRAX complexes. For clarity, CD spectra are shown for a few of the proteins/complexes. The CD spectra presented are of human translin (translin), human mutant translinB3 (translinB3), human translin-TRAX complex (translin-TRAX), and complexes of mutants of human TRAXB3 and translin (translin-TRAXB3 and translinB2-TRAXB3).

DNA-binding activity of mutant translin-TRAX complexes

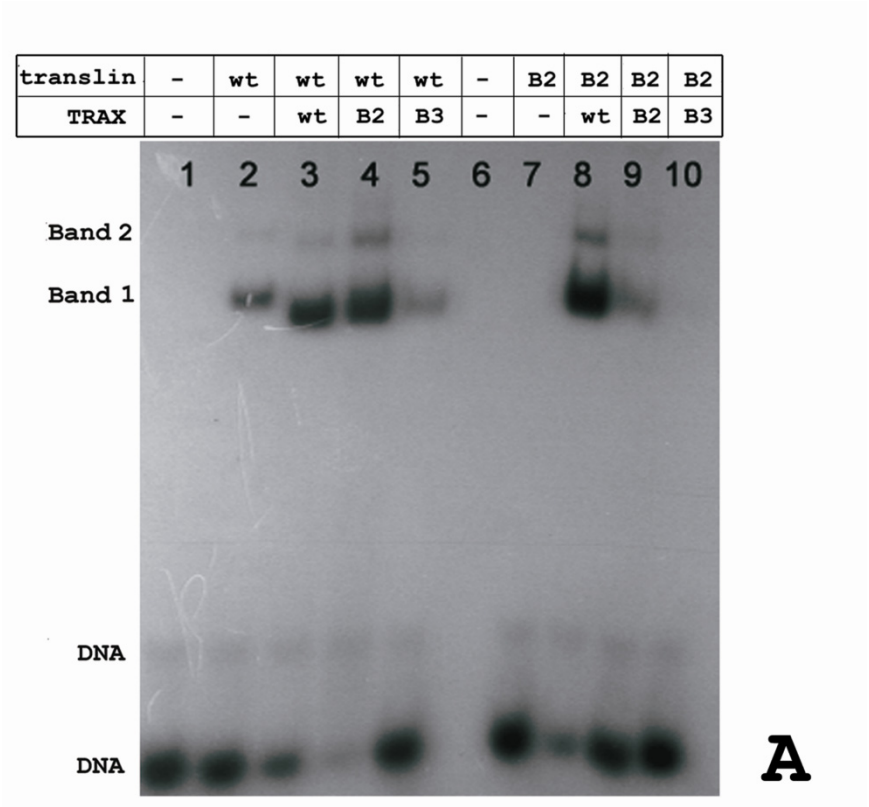
The DNA-binding activities of heteromeric translin-TRAX complexes were monitored by electrophoresis mobility shift assay using the Bcl-CL1 probe, which is a 24-mer ssDNA oligo used in initial studies characterizing the nucleic acid binding properties of translin (Aoki et al., 1995). Two bands with retarded migration (bands 1 and 2) were observed for the active protein complexes, compared to DNA alone (lane 1, Figure 5.9A). However, intensity of the slow migrating band 2 was very weak, that corresponded to higher oligomer observed in gel-filtration and native-PAGE analyses of the pure proteins. Similar band pattern was also observed in earlier studies aimed to deduce the nucleic acid binding properties of translin or translin-TRAX complex

proteins (Aoki et al., 1999; Chennathukuzhi et al., 2001a). The heteromeric translin-TRAX complex displayed enhanced ssDNA-binding activity compared to the homomeric translin complex (lanes 3 and 2; Figure 5.9A). The enhanced activity could be due to additional binding sites contributed by TRAX protein as the total loaded quantity of translin was equal in both the lanes (lane 2; 50 nM of octameric translin; lane 3, 100 nM of octameric translin-TRAX complex). To test the predicted role of the B2 and B3 motifs of TRAX in nucleic acid binding activity, we examined the ssDNA-binding activities of the heteromeric complexes formed by wild type translin with the TRAXB2 or TRAXB3 mutant constructs (Table 5.1). While the binding activity displayed by the translin-TRAXB2 complex is comparable to that of the wild type translin-TRAX complex, the binding activity of the translin-TRAXB3 complex is markedly reduced (lanes 4 and 5, Figure 5.9A).

We further tested the effect of substitutions in B2/B3 motifs of TRAX in DNA-binding ability of the heteromeric complex constructed using translinB2 mutant. Consistent with previous studies, homomeric mutant translinB2 did not show any DNA-binding activity with radiolabeled ssDNA probe (lane 7, Figure 5.9A). Interestingly, wild type TRAX could rescue activity of translin B2 mutation comparable to wild-type translin-TRAX complex (lane 8, Figure 5.9A). However, a decrease in DNA-binding activity of heteromeric translinB2-TRAXB2 complex was observed (lane 9, Figure 5.9A) compared to its wild-type counterpart. Surprisingly, heteromeric complex of translinB2-TRAXB3 was observed to be fully inactive in the study with radiolabeled ssDNA probe (lane 10, Figure 5.9A). To detect weak binding affinities, the assays were also performed with high concentrations of unlabeled DNA (10 μM) and proteins (typically, 1 and 5 μM of octameric translin proteins, and 2 and

10 μM of octamer of heteromeric complex). Although the band intensities showed trend of binding affinities similar to that observed with labeled DNA, however, weak binding was detected even with the heteromeric translinB2-TRAXB3 and homomeric translinB2 mutants (Figure 5.9B). Proteins/complexes alone were also loaded to confirm that detected intensity is not due to artifact.

Since very similar elution profile on gel-filtration column and nearly identical CD spectra were observed for all the complexes, the data suggests that alterations in the DNA-binding activities can be attributed to the substitution in DNA-binding motifs. Taken together it can be concluded both B2 and B3 motifs of TRAX contribute to DNA-binding activity in the context of the heteromeric complex, with mutations of the B3 site having clearly more pronounced effect.



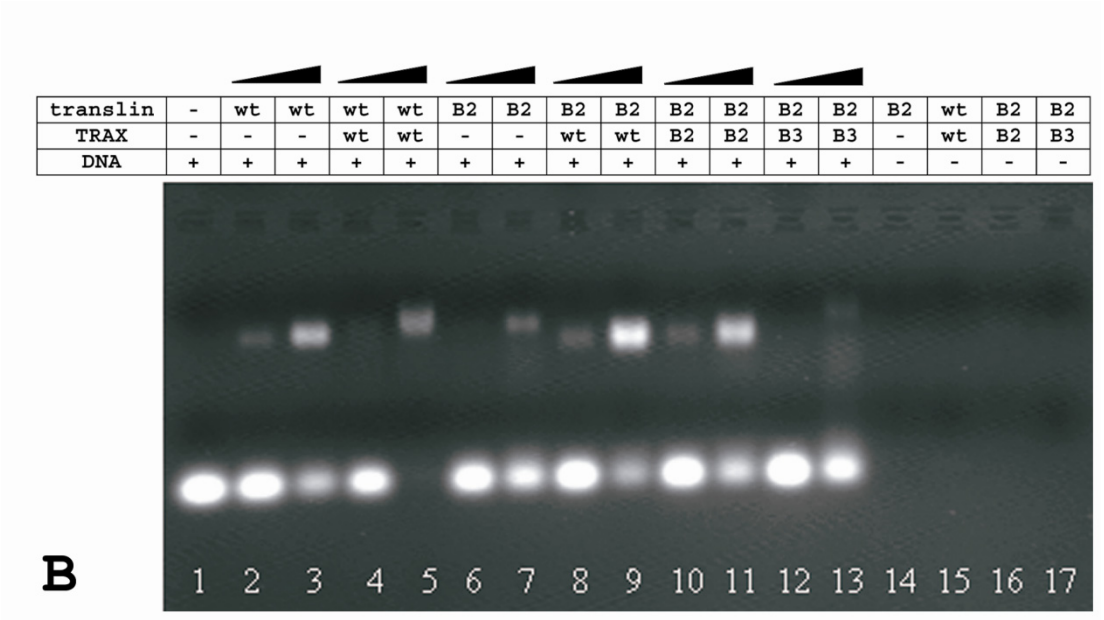
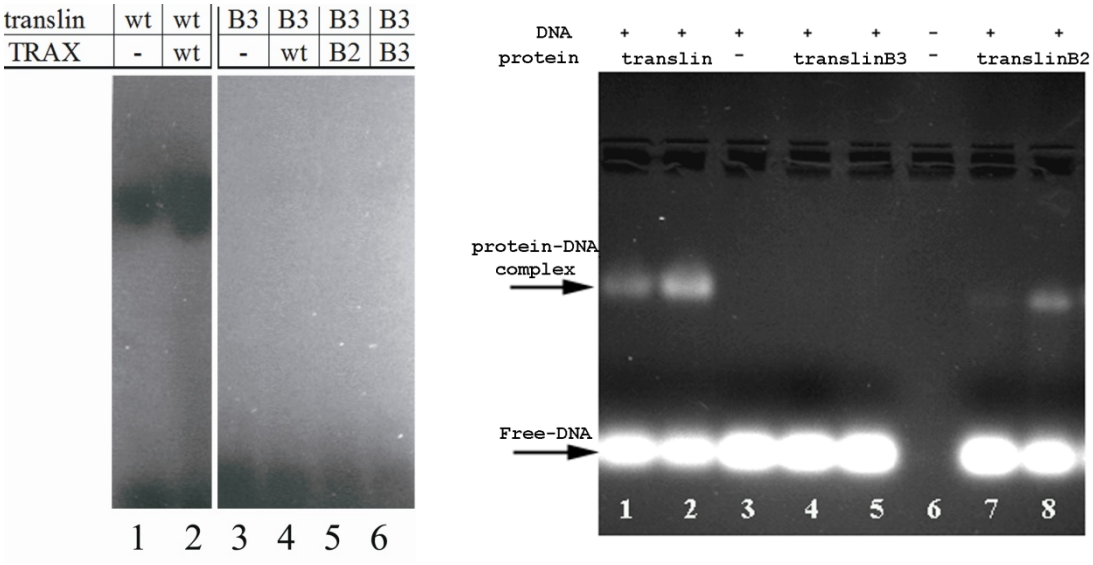


Figure 5.9 DNA-binding activity of the proteins/complexes. A) [γ-^{32}P]-labeled Bcl-CL1 24-mer ssDNA (100 nM) was incubated with human translin (50 nM of octameric translin) and translin-TRAX complexes (100 nM of octameric translin-TRAX complex). The mixtures were resolved on the 4.5% native-PAGE in TBE buffer. Lane 1, only DNA; lane 2, human translin; lane 3, translin-TRAX complex; lanes 4 and 5, complexes of human translin with the B2 and B3 mutants of human TRAX (translin-TRAXB2 and translin-TRAXB3), respectively; lane 6, Blank; lane 7, basic-2 region mutant of human translin (translinB2); lanes 8, 9 and 10, complexes of translinB2 with human TRAX, TRAXB2 and TRAXB3 (translinB2-TRAX, translinB2-TRAXB2 and translinB2-TRAXB3), respectively. **B)** Bcl-CL1 24-mer ssDNA (10 µM) was incubated with human translin (1 µM and 5 µM, lanes 2 and lane 3 respectively), wild-type translin-TRAX complex (2 µM and 10 µM, lanes 4 and 5 respectively), translinB2 mutant (1 µM and 5 µM, lanes 6 and 7 respectively), translinB2-TRAX (2 µM and 10 µM, lanes 8 and 9 respectively), translinB2-TRAXB2 (2 µM and 10 µM, lanes 10 and 11 respectively) and translinB2-TRAXB3 (2 µM and 10 µM, lanes 12 and 13 respectively) and the reaction mixtures resolved on 1.5% agarose gel were stained with ethidium bromide. At higher concentrations the ssDNA-binding activity of translinB2-TRAXB3 complex was weakly detectable with ethidium bromide. This was not observed in autoradiograms with low concentrations of radiolabeled DNA and the protein complex.

Role of predicted B3 region of translin in DNA-binding activity

To assess the role of B3 region (B3, ^{187}KNDSLRKRY^{195}) in translin, we tested the DNA-binding activity of homomeric translinB3 mutant. TranslinB3 mutant also existed as a stable octamer similar to wild–type translin as adjudged by gel-filtration and native-PAGE analysis (Figure 5.7) and had similar CD spectra (Figure 5.8) compared to wild-type translin. However, the translinB3 mutant was completely inactive in DNA-binding assays with radiolabeled DNA probe as well at higher concentrations (100-fold higher compared to radiolabeled gel) of protein and DNA in ethidium bromide stained agarose gel electrophoresis (Figure 5.10 A&B). Similar results were reported for the *S. pombe* translin where substitution in Arg-210 and Arg-211 (equivalent to Arg-192 and Lys-193 of human translinB3 motif) dramatically reduced DNA-binding activity (Eliahoo et al., 2010). In comparison, translinB2 mutant displayed detectable binding at higher concentrations of protein and DNA (lanes 7 & 8 Figure 5.10B). Surprisingly, the activity could not be restored in heteromeric complex of translinB3 with wild-type TRAX (translinB3-TRAX; lane 4, Figure 5.10A) as well as with mutant TRAX (translinB3-TRAXB2, lane 5; translinB3-TRAXB3, lane 6; Figure 5.10A). The results suggest strongly that B3 region residues play a dominant role in the DNA-binding activity of homomeric translin as well as heteromeric translin-TRAX complex.



(A) **(B)**

Figure 5.10 DNA-binding activity of proteins/complexes. A) [γ-^{32}P]-labeled Bcl-CL1 24-mer ssDNA (200 nM) was incubated with human translin (100 nM of octameric translin) and translin-TRAX complexes (100 nM of translin-TRAX complex). The mixtures were resolved on the 4.5% native-PAGE. Lane 1, human translin; lane 2, human translin-TRAX complex; lanes 3, 4 and 5, complexes of translinB3 mutant with wild-type human TRAX, with B2 mutant of human TRAX and with B3 mutant of human TRAX, respectively. **B)** Bcl-CL1 24-mer ssDNA (200 pmol) was incubated with two different concentrations of human translin (50 pmol and 100 pmol of translin octamer). The mixtures were resolved on the 1.5% agarose gel. Lanes 1 & 2, wild-type human translin (tsn); lanes 4 & 5, basic-3 region mutant of human translin (translinB3); lanes 7 & 8, basic-2 region mutant of human translin (translinB2). Lane 3, only DNA; lane 6, blank.

Structural analysis of translin-DNA complex

The crystals of translin-DNA complex were found to be nearly isomorphous to the crystals of native-translin, for which the structure has been reported at 2.2 Å resolution (PDB code 1J1J; Sugiura et al., 2004). The structure was refined with REFMAC5 (Vagin et al., 2004) and model was subsequently fitted to the electron density maps using COOT (Emsley et al., 2004). The structure was refined to R_{work}/R_{free}, values of 0.19/0.26, and the stereochemistry of the model was excellent. The electron density for B2 region residues was well defined in the electron density maps. But the electron density for Arg-192 and Lys-193 of B3 motif was not clearly defined. A few electron density blobs, too big to be accounted for by water molecules, were observed sand-witched between Arg-153/B (Arg-153 residue of the B-chain) and Lys-203/A as well as Arg-153/A and Lys-203/D. The electron density for these was observed at 5σ contour levels in difference Fourier maps. However, a contiguous electron density was not observed to fit DNA molecule. Refinement in crystallization strategy, say using symmetrical DNA sequence may help to elucidate DNA-binding in translin.

DISCUSSION

The nuclease activity of translin-trax heteromeric complex has been in recent focus due to its role in siRNA processing (Liu et al., 2009; Ye et al., 2011). A number of studies also provided evidence that trax and translin play important role in DNA damage recovery (Jendling et al., 2010; Erdemir et al., 2002a). The functional role of these proteins vis-à-vis their DNA-binding ability has also been argued recently (Lluis et al., 2010). Binding of DNA to the translin-trax complex has been thought to be mediated

through translin, and trax was not expected to bind the nucleic acid substrates (Chennathukuzhi et al., 2001a; Lluis et al., 2010; Ye et al., 2011). In the present study, we show direct binding of ssDNA with the poly-histidine tagged trax (TRAX) protein by crosslinking radiolabeled ssDNA with translin-TRAX complex using UV-laser. The DNA-binding studies were carried out with the heteromeric complexes as TRAX alone is not stable by itself (Yang et al., 2004; Gupta et al., 2005; Tian et al., 2011). The heteromeric complexes used in DNA-binding studies were purified using size exclusion column chromatography and were of molecular masses ~300 kDa. From the relative intensity of translin or TRAX bands on the SDS-PAGE equimolar ratio of the two proteins in the heteromeric complexes could be taken. Although we expect heteromeric complexes as $translin_4$-$trax_4$ multimers, minor contamination due conformational oligomers of compositions $translin_5$-$trax_3$ or $translin_6$-$trax_2$, given the resolution of Superdex 200 gel filtration column, cannot be ruled out. Post-irradiation, the heteromeric complex was disrupted using chaotropic agent, urea, under boiling condition. The his-tagged TRAX/TRAX-DNA crosslinked complex was purified by metal chelating affinity matrix under denaturing conditions, while untagged translin was recovered in extensive wash steps. The TRAX-DNA complex was clearly detected on autoradiograph of a SDS-PAGE analysis of the eluted fractions. Also, the translin-DNA covalent complex was detected in the irradiated translin:DNA mixture. The TRAX-DNA complex migrated corresponding to molecular mass of about 44 kDa, compared to 36 kDa translin-DNA covalent complex (Figure 5.1).

Since the UV light is known to induce protein–DNA crosslinking predominantly at their contact points, we tested the hypothesis that the B2 and B3 motifs in TRAX play a critical role in the nucleic acid binding activity of the

heteromeric complex. The B2 and B3 motifs of TRAX are juxtaposed to the DNA-binding B2/B3 motifs on translin in the recently resolved crystal structures of translin-TRAX complexes (Ye et al., 2011; Tian et al., 2011) (Figure 5.4). We mutated residues of these motifs and examined the DNA-binding activity of mutant complexes. These mutations did not cause any perturbation in structure or their oligomeric status as wild-type and mutant complexes showed similar CD spectra and elution profiles on gel-filtration analysis. We found that mutations in the TRAX B3 site drastically reduce the binding activity of the heteromeric complex formed with wild type translin, while mutations in the TRAX B2 site do not. However, the TRAX B2 mutations do impair the ability of TRAX to form fully active heteromeric complexes with the translinB2 mutant. Also, drastic loss of DNA-binding activity was observed in heteromeric complex translinB2-TRAXB3 (Figure 5.9). Lastly, homomeric or heteromeric complexes with substitutions in the B3 motifs of translin are totally inactive in binding assays (Figure 5.10).

The DNA-binding domain of human translin is formed by combination of its basic motifs in a multimeric structure and loss of multimeric structure results in abrogation of its DNA-binding abilities (Aoki et al., 1999). The DNA-binding competent octameric or decameric translin complexes were earlier shown to be constituted by energetically stable and evolutionarily conserved dimers in up-down configuration (Gupta et al., 2008). In contrast, trax has been shown not to interact with itself using yeast two hybrid system (Chennathukuzhi et al., 2001a). Additionally, heteromeric complexes have been observed in a variety of trax/translin compositions; $translin_4$-$trax_4$ (Lluis et al., 2010), $translin_5$-$trax_3$ and $translin_6$-$trax_2$ (Ye et al., 2011) and $translin_4$-$trax_2$ (Tian et al., 2011). The complexes are constituted by translin-trax

and translin-translin up-down dimers. Importantly, trax-trax homodimers are not observed. These observations together hint that trax alone cannot oligomerize. It can thus be rationalized that DNA-binding activity of trax could not be detected in earlier studies since it alone could not achieve nucleic acid binding-competent oligomeric status.

The conformation of translin-trax heterodimer closely resembles that of translin-translin dimer in crystal structures (Ye et al., 2011; Sugiura et al., 2004) (Figure 5.4). Since mutations in both of the B2 and B3 sites abolish the nucleic acid binding activity of the complex, it can be concluded that B2 and B3 motifs of both translin and trax contributes to DNA-binding activity. The identified nucleic acid binding motifs reside on the equatorial region of the translin-trax octamer. The shortest distance between DNA-binding motif and trax catalytic centre responsible for RNase activity of translin-trax complex is about 9Å (Liu et al., 2009; Ye et al., 2011). We anticipate that knowledge of DNA-binding domains of trax will help guide future studies aimed at elucidating relationship between its high affinity nucleic acid binding and RNase activities of the translin-trax complexes.

In summary, we isolated the covalent complex of trax with the probe ssDNA after disrupting the UV-irradiated heteromeric translin-trax complex. We have identified the B2 and B3 sites in trax as mediating its role in nucleic acid binding activity of the heteromeric complex. We suggest that the complex is assembled from translin-trax or translin-translin dimer subunits, and B2/B3 motifs of both translin and trax primarily contribute to DNA-binding activity.

Table 5.1 Details of the constructs and expression systems used, and ssDNA-binding characteristics of the purified proteins

Protein/Complex (abbreviation)	**Expression system**	**Activity**[1]
Human translin (translin)	*p*QE9-*tsn* transformed into *E. coli* BL21 (DE3)	+ + +
Mutant of basic-2 motif of human translin (translinB2)	*p*QE9-*tsnB2* transformed into *E. coli* BL21 (DE3)	ND[2]
Mutant of basic-3 motif of human translin (translinB3)	*p*QE9-*tsnB3* transformed into *E. coli* BL21 (DE3)	ND
Complex of human translin and TRAX proteins (translin-TRAX)	*p*ET21a-*tsn* and *p*ET28a-*trx* co-transformed into *E. coli* BL21(DE3)	+ + + +
Complex of human translin with human B2 mutant TRAX (translin-TRAXB2)	*p*ET21a-*tsn* and *p*ET28a-*trxB2* co-transformed into *E. coli* BL21(DE3)	+ + + +
Complex of human translin with human B3 mutant TRAX (translin-TRAXB3)	*p*ET21a-*tsn* and *p*ET28a-*trxB3* co-transformed into *E. coli* BL21(DE3)	+
Complex of human B2 mutant translin with human TRAX (translinB2-TRAX)	*p*ET21a-*tsnB2* and *p*ET28a-*trx* co-transformed into *E. coli* BL21(DE3)	+ + + +
Complex of human B2 mutant Translin with human B2 mutant TRAX (translinB2-TRAXB2)	*p*ET21a-*tsnB2* and *p*ET28a-*trxB2* co-transformed into *E. coli* BL21(DE3)	+ + +
Complex of human B2 mutant translin with human B3 mutant TRAX (translinB2-TRAXB3)	*p*ET21a-*tsnB2* and *p*ET28a-*trxB3* co-transformed into *E. coli* BL21(DE3)	ND[2]
Complex of human B3 mutant translin with human TRAX (translinB3-TRAX)	*p*ET21a-*tsnB3* and *p*ET28a-*trx* co-transformed into *E. coli* BL21(DE3)	ND
Complex of human B3 mutant Translin with human B2 mutant TRAX (translinB3-TRAXB2)	*p*ET21a-*tsnB3* and *p*ET28a-*trxB2* co-transformed into *E. coli* BL21(DE3)	ND
Complex of human B3 mutant translin with human B3 mutant TRAX (translinB3-TRAXB3)	*p*ET21a-*tsnB3* and *p*ET28a-*trxB3* co-transformed into *E. coli* BL21(DE3)	ND

ND-Activity was not detected under experimental conditions using radiolabeled ssDNA

1. The relative DNA-binding activity represented as four bins of 25% each.
2. Weak activity detected with higher concentrations of proteins and DNA

Chapter 6

Evolution of translin-like proteins and concluding remarks

INTRODUCTION

Orthologs are genes that evolve from a single ancestral gene by vertical descent. Such genes have often retained their biological role and the structure, essential for the activity, in the present day species. Paralogs, evolve after the gene duplication event, diverge leading to emerging of new specificities and functions due to weaker evolutionary pressure (Jensen, 2001; Koonin, 2005). Orthology and paralogy are key concepts of evolutionary genomics introduced by Fitch (1970). Translin and its interacting partner protein trax are very well conserved in the evolution. Translin binds with ssDNA sequences found at the chromosomal breakpoint junction and is also involved in regulation of mRNA transport and translation in brain and testis (Jaendling et al., 2010). Trax protein also harbours DNA-binding residues (chapter 5) and in addition, harbours RNase activity in its sequence (Liu et al., 2009; Ye et al., 2011). Trax in eukaryotic species shares nearly 22% sequence identity with translin from the same species. Trax harbours nuclear localization signal while translin harbours nuclear export signal. Crystal structures of heteromeric translin-trax complexes of human and drosophila have been reported recently (Ye et al., 2011; Tian et al., 2011). The analysis showed that monomeric structure of translin and trax are very similar with an RMSD of ~1.6 Å over 177 equivalent C^{α} atoms. The structural similarity and presence of similar DNA-binding motifs suggest that translin and trax proteins share a common ancestor. This prompted us to study the evolutionary relationship of both the proteins as well as to determine ancestry of their genes. We elucidated characteristic sequence motifs of both the proteins, distinguished homologs of these proteins in a variety of organisms, including prokaryotes, and studied their relationship by phylogenetic analysis.

METHODS

Identification of specific sequence motifs for translin and trax proteins

Protein sequences of translin and trax were obtained from UniProt database of EBI (European Bioinformatics Institute) and GenBank database of NCBI (National Center for Biotechnology Information). All the sequences were taken in the FASTA format. Initially, the candidate eukaryotic orthologs of translin and trax were searched independently in the UniProt databases with Blast search engine using human proteins as query sequences. The candidate orthologs of each protein were selected with E-value cut-offs of $<10^{-20}$. In order to reduce redundancy and bias in the sequence set, the sequences were clustered and sequences of distinct but diverse eukaryotic species were retained in two independent sets of translin and trax proteins. These candidate 'true' sequences of translin and trax were aligned independently using multiple sequence alignment tool ClustalW (Thompson et al., 1994) and by using structure based alignment tool (3DCoffee; www.tcoffee.org; Armougom et al., 2006). The ClustalW alignment was manually edited using BioEdit version 5.0.9 (Hall, 1999); the large insertion/deletion sequences (indels) observed in loop regions (compared to template structure) with poor homology were deleted from the alignment. This edited multiple sequence alignments were used to predict the conserved motifs for translin and trax independently. The graphical representation of patterns in the alignment was generated using WebLogo server (http://weblogo.berkeley.edu/logo.cgi) (Crooks et al., 2004). The strictly conserved motifs of translin or trax family proteins were identified from the sequence logo. Subsequently, more diverged sequences observed in species from amoebozoa, algae, archaea and eubacteria were mined from the UniProt and GenBank databases. These sequences were grouped into translin or trax sets based on the presence

of conserved motifs specific to each family. These newly found sequences were aligned with translin and trax 'true' sequences using ClustalW.

Secondary structure and fold prediction for bacterial translin-like protein

The secondary structure for prokaryotic (*Methanococcus vannielii*) translin-like protein was predicted with two independent servers, PSIPRED server, (http://bioinf.cs.ucl.ac.uk/psipred/) (McGuffin et al., 2000) and PROF server (http://www.aber.ac.uk/~phiwww/prof/). The fold for the Methanococcus protein was predicted with GenTHREADER (http://bioinf.cs.ucl.ac.uk/psipred/) (Jones, 1999).

Phylogenetic analysis

Phylogenetic and molecular evolutionary analyses were conducted using MEGA version 5.0 (Tamura et al., 2011). The evolutionary history was inferred using the Neighbor-Joining method (Saitou & Nei, 1987**)**. The bootstrap consensus tree inferred from 500 replicates is taken to represent the evolutionary history (Felsenstein, 1985). All positions with less than 50% site coverage were eliminated. That is, fewer than 50% alignment gaps, missing data, and ambiguous residues were allowed at any position. The evolutionary distances were computed using the Poisson correction method (Zuckerkandl & Pauling, 1965).

RESULTS AND DISCUSSION

Translin and trax proteins are highly conserved within eukaryotes. The least pair-wise sequence identity between the most diverse eukaryotes for translin, for instance, human

and S. pombe (Fungi) is observed to be 37%. Translin and trax proteins of the same species also share significant sequence and structural homology. For instance, human translin and trax proteins share 22% sequence identity and very similar monomeric structure (RMSD ~ 1.6 Å over 177 C^{α} atoms). This suggested that both the proteins might have evolved from a common ancestor gene. We searched for the orthologs of translin and trax from the UniProt database using human translin and human trax as query sequences, respectively, with an E-value cut-off of 10^{-20}. All translin and trax sequences of eukaryotes were aligned together using multiple sequence alignment tool ClustalW. The guide tree, constructed using multiple sequence alignment of translin and trax sequences, clearly demonstrated independent clustering of translin and trax proteins (Figure 6.1), which suggested that translin and trax have evolved independently in the eukaryotic evolution (within the studied evolutionary distances). It is likely that translin and trax divergence occurred prior to evolution of eukaryotes. The analysis also hinted that translin and trax sequences may harbour the sequence motifs specific to each family. To find the conserved signature motifs of translin and trax, respectively, a subset of 'true' (high-confidence) translin or trax sequences were selected from eukaryotes and were assembled in independent datasets for translin or trax proteins. Each subset had sufficient representation of eukaryote kingdoms with reasonable divergence to reduce the bias. These true translin or trax sequences were independently aligned using sequence and structure based alignment methods, ClustalW and 3DCoffee, respectively. The sequences used in the multiple sequence alignments for translin and trax are listed in Tables 6.1 & 6.2.

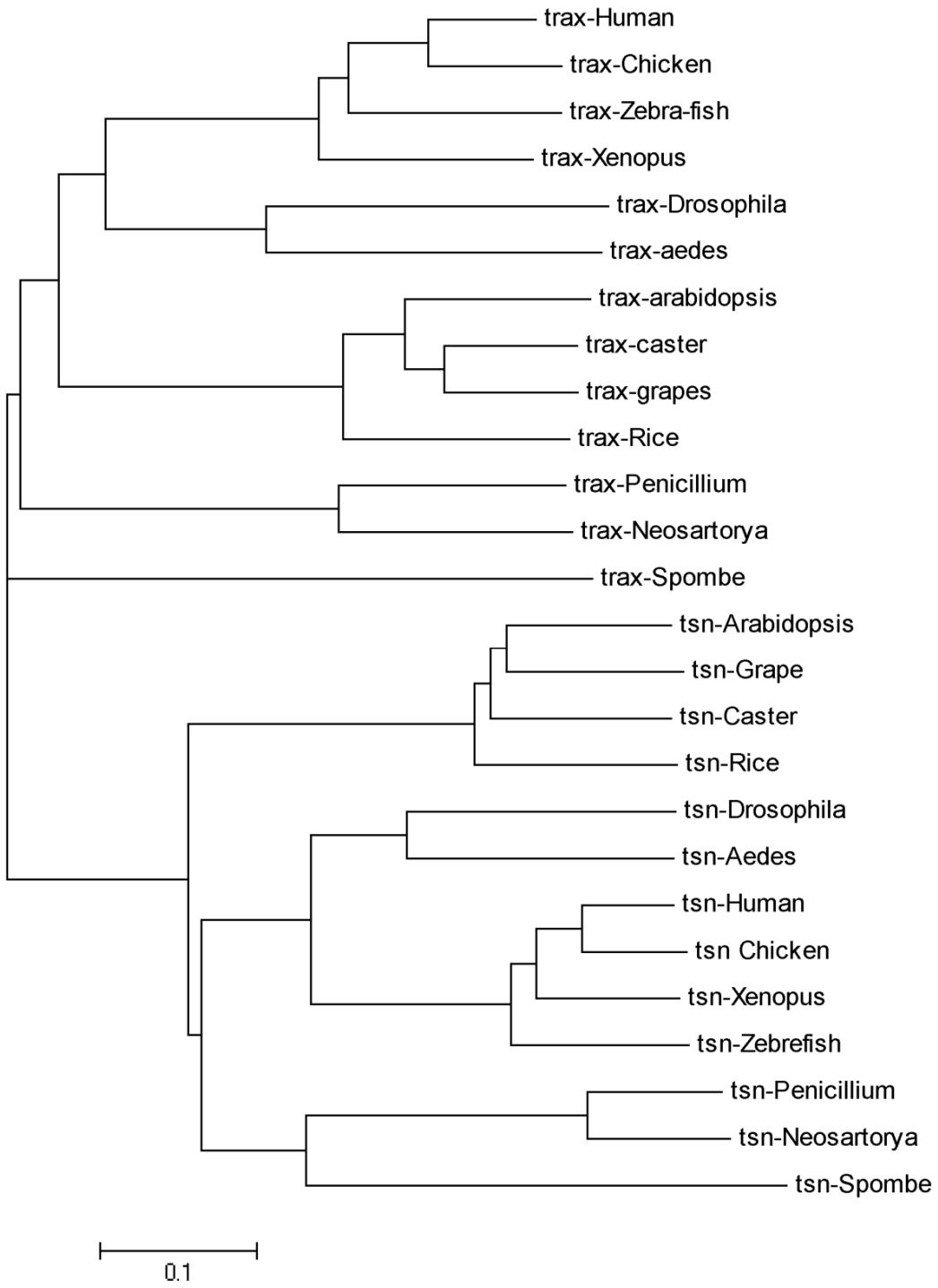


Figure 6.1 Dendrogram of translin and trax proteins. The protein sequences of eukaryotic translin and trax (listed in Table 6.1 & Table 6.2) were aligned together using ClustalW. The sequence alignment was used as input to construct the guide tree using MEGA5 software suite.

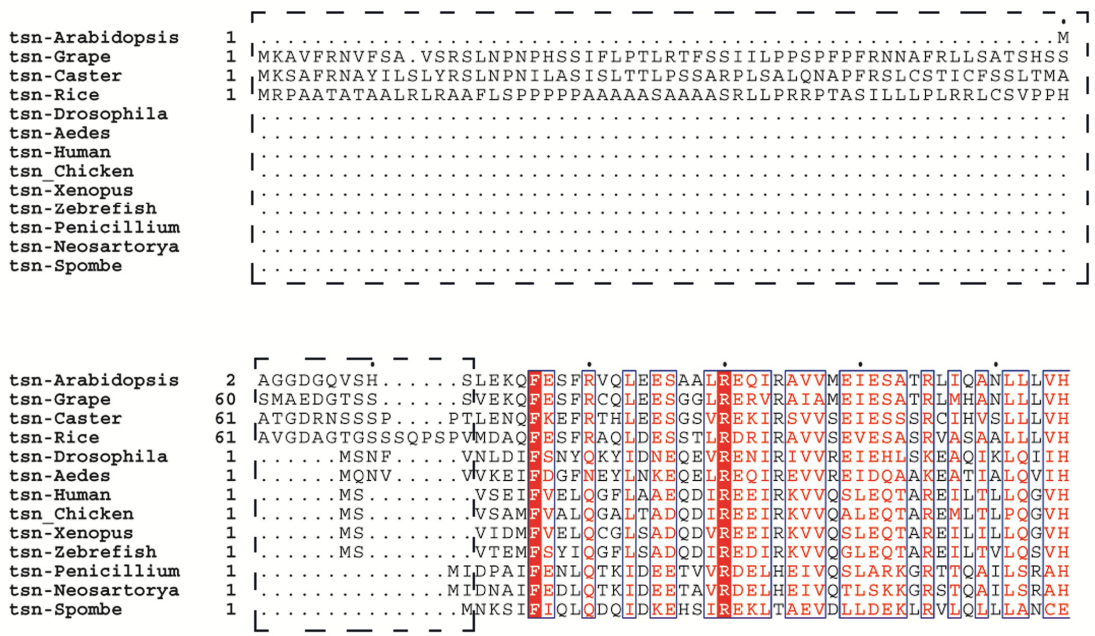
Specific sequence motifs for translin

Translin sequences were found to be highly conserved in all the eukaryotes and have a length of 220–235 amino acids, except in plants (Figure 6.2). The plant translin sequences have additional 60–70 residues at its N-terminus, compared to vertebrate translins. Translin from fungi species were found to have an insertion of 10–12 residues, compared to other translin orthologs. The alignment of translin sequences was edited manually; the N-terminal regions of plant translins, 12 residue insertions of fungi translins and a few C-terminal residues not conserved in translin orthologs, were removed from the alignment. The removed regions are enclosed in dashed line box in the alignments (Figure 6.2). The manually edited alignment was used as input for the construction of sequence logo. The nuclear export signal (NES) sequence (LxxxLxxLxL) was not conserved in translin orthologs. It was earlier proposed that NES present on mouse translin sequence might be responsible for shuttling of translin and trax between nucleus and cytoplasm (Chennathukuzhi et al., 2001a).

The C-terminus region (181–215 of human translin) was observed to be highly conserved in translin proteins. Three motifs identified, specific to translin, from the sequence pattern of translin orthologs are: **motif-1**, V-[TI]-x-G-[DN]-Y-x-x-P; **motif-2**, L-[NS]-L-[KR]-N-D-x-L-R-[KR] and **motif-3**, [KR]-[KR]-x-E-[ED]-[VI]-[VY]-Y-D. These motifs 1–3 have been marked in the multiple sequence alignment of translin proteins (Figure 6.2). The residues of motif-1 and motif-2 of translin are known to be important for the function of translin proteins. The VTAGD residues of motif-1 have been reported as GTP binding site in mouse translin. GTP binding reduces the affinity of translin towards RNA, not towards DNA. The mutation (VTAGD to VTNSD) resulted in loss of GTP binding activity as well as RNA binding activity of mouse

translin (Chennathukuzhi et al., 2001b). However, drosophila ortholog of translin harbors A162M substitution and does not bind with GTP (results presented in chapter 4). The conserved proline in motif-1 has also been observed to play important role in the stability of translin fold (Gupta et al., 2008; Chapter 4 of the presented work).

Conserved leucine residues (Leu-184 and Leu-191 of human translin) of motif-2 were reported crucial for multimeric assembly of translin such that substitution of leucine residues resulted in loss of octameric state and thus loss of nucleic acid binding activity (Aoki et al., 1999). The R-[RK] residues of motif-2 have been identified as part of the basic-3 region of translin proteins and were found to be crucial for nucleic acid binding activity of translin proteins (discussed in chapter 5). The motif-3 residues were found to be highly specific to translin orthologs and were useful for distinguishing translin and trax proteins from highly diverse genomes in the present analysis. The functional role of motif-3 residues (206–211 of human translin) of translin proteins is not yet clear. However, C-terminal residues (205–228) of translin were observed crucial for its multimeric sate as well as for its interaction with trax (Wu et al., 1998).



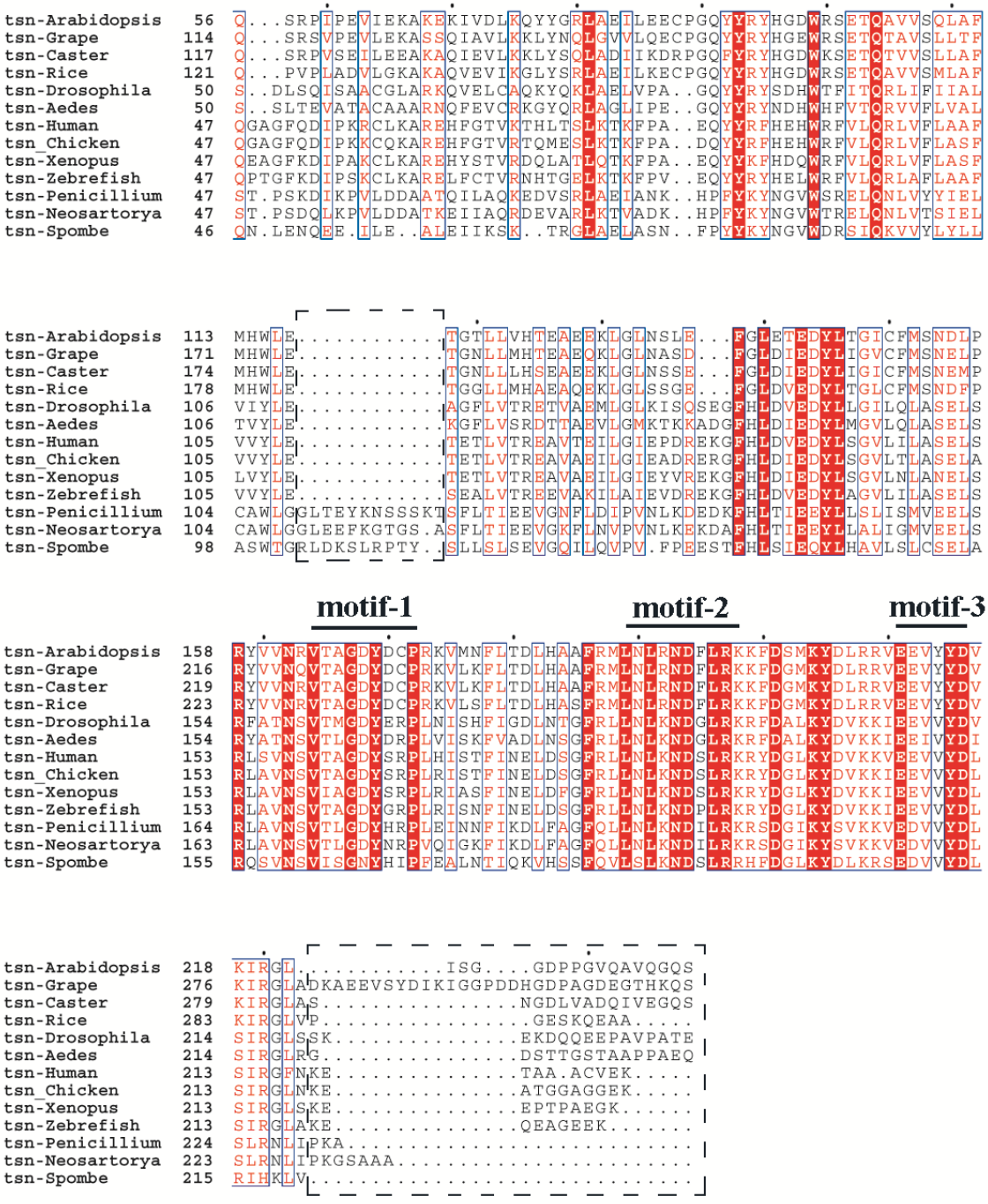


Figure 6.2A Identification of sequence motifs specific to translin.

The 'true' translin sequences with high confidence were selected as candidate sequences. These candidate sequences were aligned with ClustalW. The sequence motifs are mentioned as **motif-1 to 3**. The regions with large indels, enclosed within dashed lines were deleted from the alignment

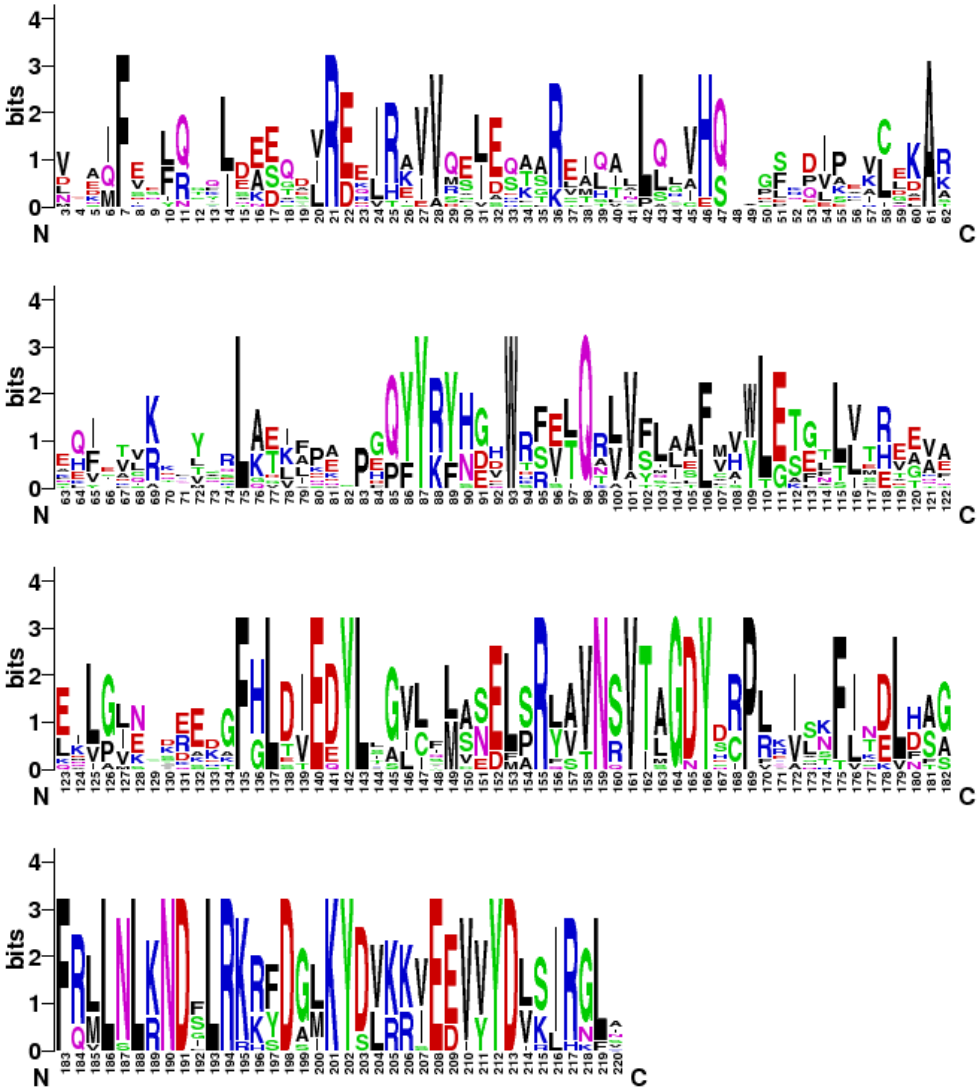


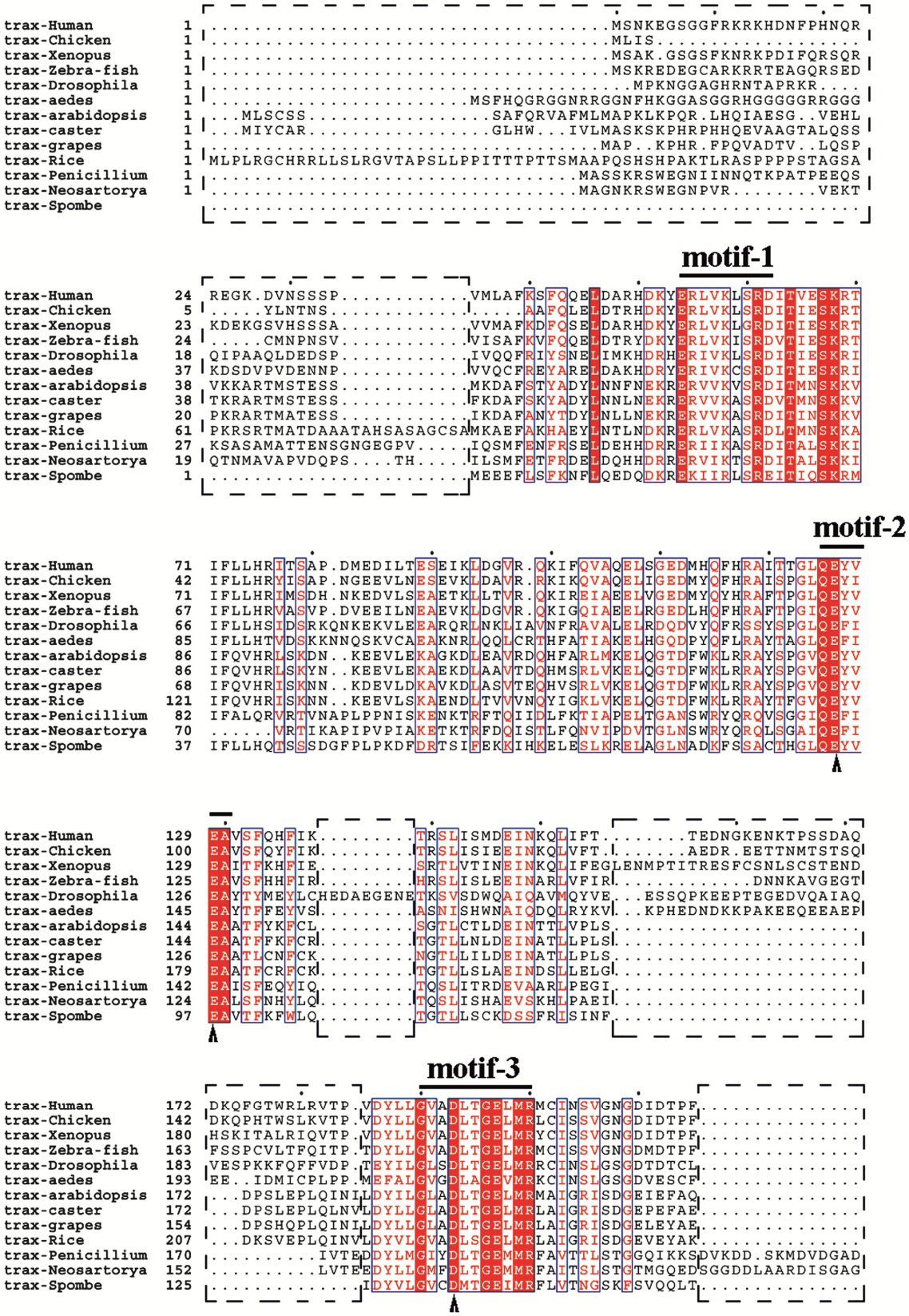
Figure 6.2B

Figure 6.2B Identification of sequence motifs specific to translin.

Logo-plots of translin designed using WebLogo server. These Logo-plots were used to find the sequence motifs specific to translin family proteins.

Specific sequence motifs for trax

Trax sequences were observed to be less conserved with a large number of indels and variations in the sequence length (Figure 6.3A). However, the observed indels were localized in the loop regions when the trax sequences were mapped on the known structure of human or drosophila trax. Trax sequences were also manually edited and indels, marked in the alignment, were deleted. This edited alignment was used to find the conserved sequence motifs specific to trax family proteins. Three motifs mainly, **motif-1**, E-[RK]-[LIV]-[VI]-[KR]-x-[SG]-R-[DE]; **motif-2**, Q-**E-**[YF]-[VI]-**E-**A; and **motif-3**, G-[VLIM]-x-**D-**[LM]-[TAS]-G-E-[LMIV]-M-R, were identified from the Logo plots (Figure 6.3B). Notably, the glutamate (E) and aspartate (D) residues, marked as bold letters in the motif-2 and motif-3, have recently been reported crucial for RNase activity of trax observed in the translin-trax complex of drosophila and human (Liu et al., 2009; Ye et al., 2011). The acidic amino acid triad (two glutamate and one aspartate), disposed in close spatial proximity in trax structure, is expected to be the Mg^{2+} coordinating site that is required for RNase activity (Liu et al., 2009; Tian et al., 2011). The functional role of motif-1 residues is not yet clear. DP-bind server, used to identify nucleic acid binding site, did not predict direct involvement of motif-1 residues in nucleic acid binding activity of trax proteins. We further explored the role of these residues in shuttling of trax between cytoplasm and nucleus. The nuclear export signal (NES) prediction server (http://www.cbs.dtu.dk/services/NetNES/) identified motif-1 region as nuclear export signal in Chicken and S. pombe orthologs of trax (score 50–60%). However, weak signatures were also present in other orthologs like Xenopus and drosophila (score ~40%). The results have not been validated further.


motif-1
motif-2
motif-3

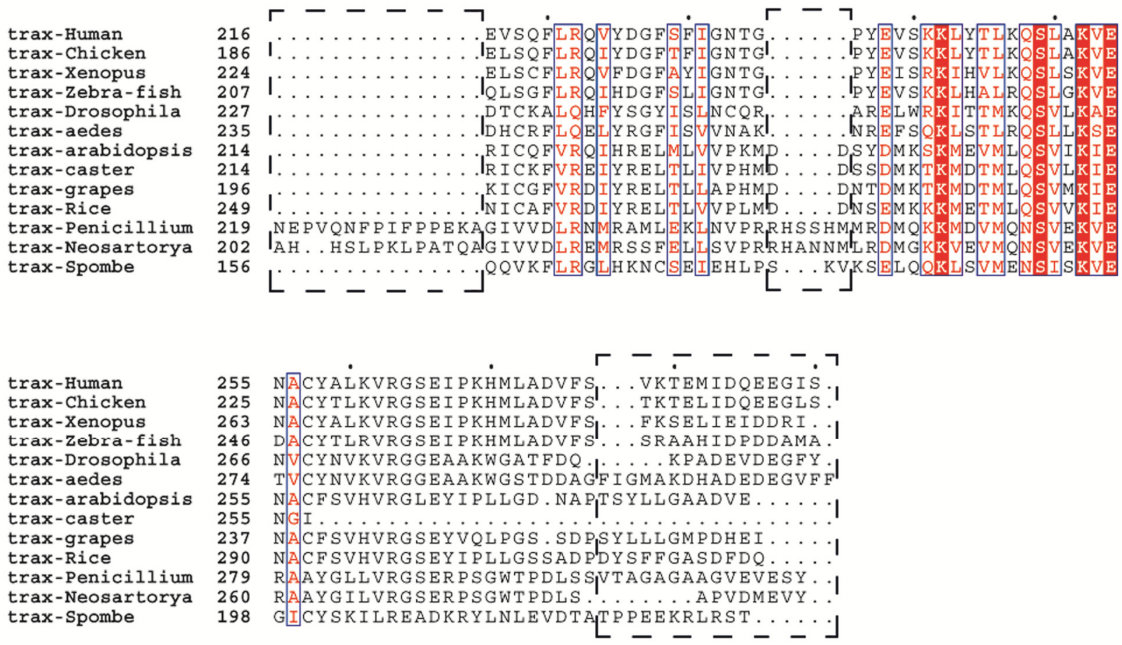


Figure 6.3A Identification of sequence motifs specific to trax.

The 'true' trax sequences with high confidence were selected as candidate sequences. These candidate sequences were aligned with ClustalW. The sequence motifs are mentioned as motif-1 to 3. The regions with large indels, enclosed within dashed lines were deleted from the alignment.

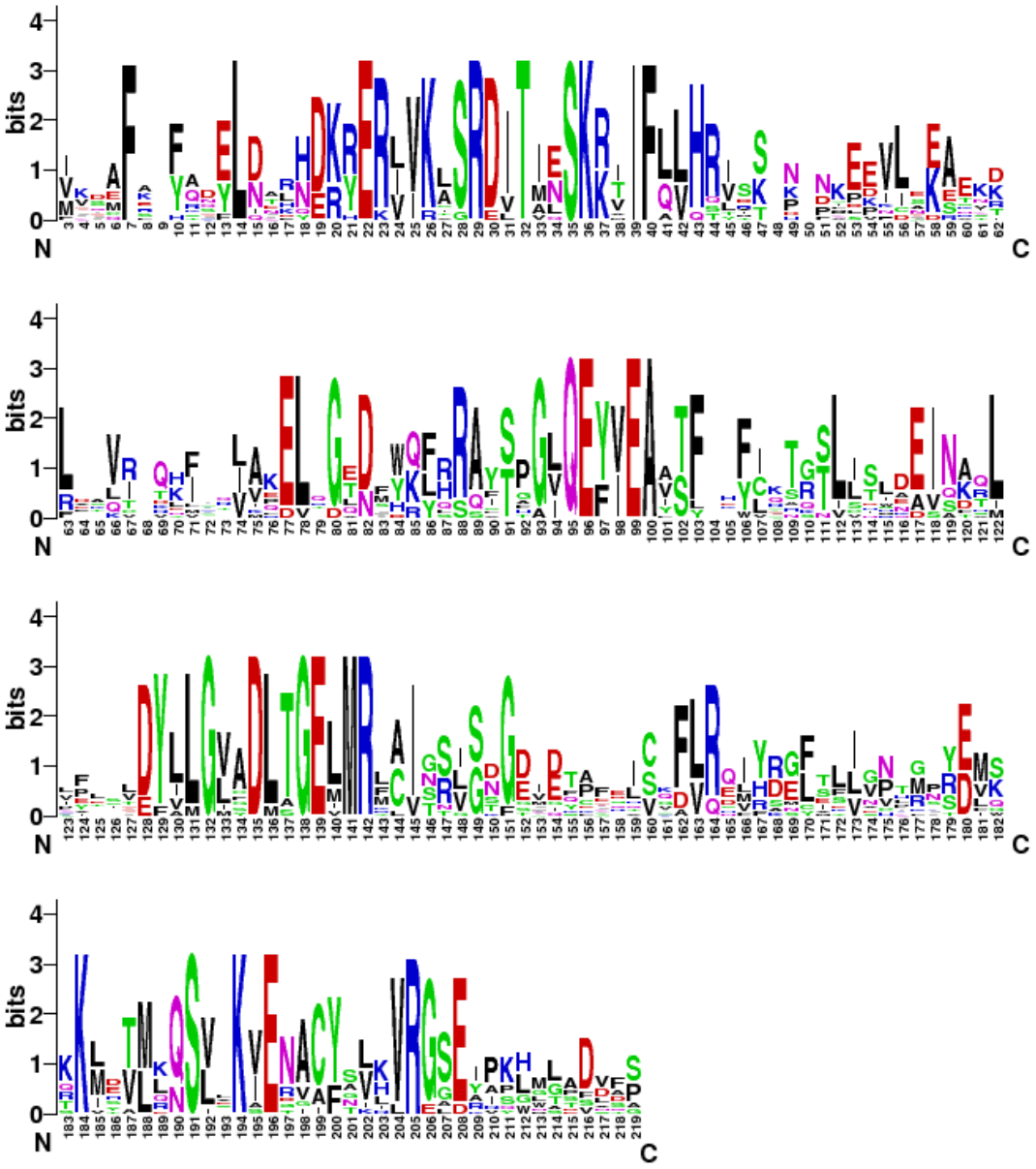


Figure 6.3B

Figure 6.3B Identification of sequence motifs specific to trax.

Logo-plots of trax designed using WebLogo server. These Logo-plots were used to find the sequence motifs specific to trax family proteins.

Evolution of eukaryotic translin and trax

Subsequently, search for the presence of translin and/or trax in the genomes of distantly related species was carried out. These sequences were characterized as translin or trax based on the elucidated protein-specific sequence motifs. Most of the newly identified sequences had been annotated as member of translin superfamily (pfam01997). We found two translin-like proteins in single cellular eukaryote amoeba. These two translin-like amoebozoa proteins could be characterized as translin and trax, respectively.

Further search into the prokaryotes revealed interesting results. Only one copy of translin-like protein was observed in archaea and Chlorobacteria (Chloroflexi phylum of eubacteria, also known as green non-sulphur bacteria) (Table 6.3). However, translin-like proteins were not observed in other eubacterial phylums. The evolutionary history of eukaryotic translin and trax proteins together with bacterial translin-like protein was inferred with phylogenetic analysis. The phylogenetic tree, constructed using the Neighbor-Joining method (Figure 6.4), demonstrated that bacterial protein is the ancestor of eukaryotic trax and translin proteins. Interestingly, the prokaryotic translin-like protein possessed all the trax-specific motifs. Particularly, residues required for RNase activity, that is specific to trax proteins, were strictly conserved in prokaryote sequences (Figure 6.5). Also, translin specific motifs were not present in the prokaryotic protein. An extensive search (with low E-values of <0.01) for proteins having translin-specific motifs further did not reveal orthologs of translin in the genomes of prokaryote species (NCBI database). The analysis thus suggested that annotated translin-like protein in prokaryotes is actually closely related to eukaryotic trax.

To be assured that the prokaryotic protein is the ancestral protein of trax and translin, we predicted the secondary structure and fold of the bacterial protein. The

secondary structure of translin-like protein of *Methanococcus vannielii* (prokaryote) was predicted with two independent servers; PSIPRED and PROF servers. Both the servers predicted the Methanococcus protein to be all helical with high confidence values and the predicted helices matched well with the helices observed in the crystal structure of trax or translin. The predicted secondary structure of Methanococcus protein has been marked on the multiple sequence alignment of prokaryotic translin-like protein and eukaryotic trax (Figure 6.5). The fold prediction server, GenTHREADER, also predicted the fold of Methanococcus protein to be similar to human trax (K-chain of 3QB5) with high confidence (probability of false positive, p-value, $\sim 10^{-7}$). The data taken together suggests that prokaryotic translin-like protein is the ancestral protein of trax and translin. Since bacterial protein possesses the signature motifs conserved for trax family proteins, it can be suggested that eukaryotic trax and bacterial translin-like proteins are orthologs and bacterial translin-like proteins may possess RNase activity. Translin, a paralog of trax, might have evolved later in the evolution. However, the results raised interesting questions that if trax is the ancestral protein of translin then why trax itself is not stable in eukaryotes and requires translin for its proper fold? Did trax evolve dependence on translin, later in the evolution? Also, archaea are known to be closer to eukaryotes in the evolution but Chlorobacteria has no direct evolutionary relation with archaea or eukaryotes. This study may be useful in investigations related to eukaryotic evolution.

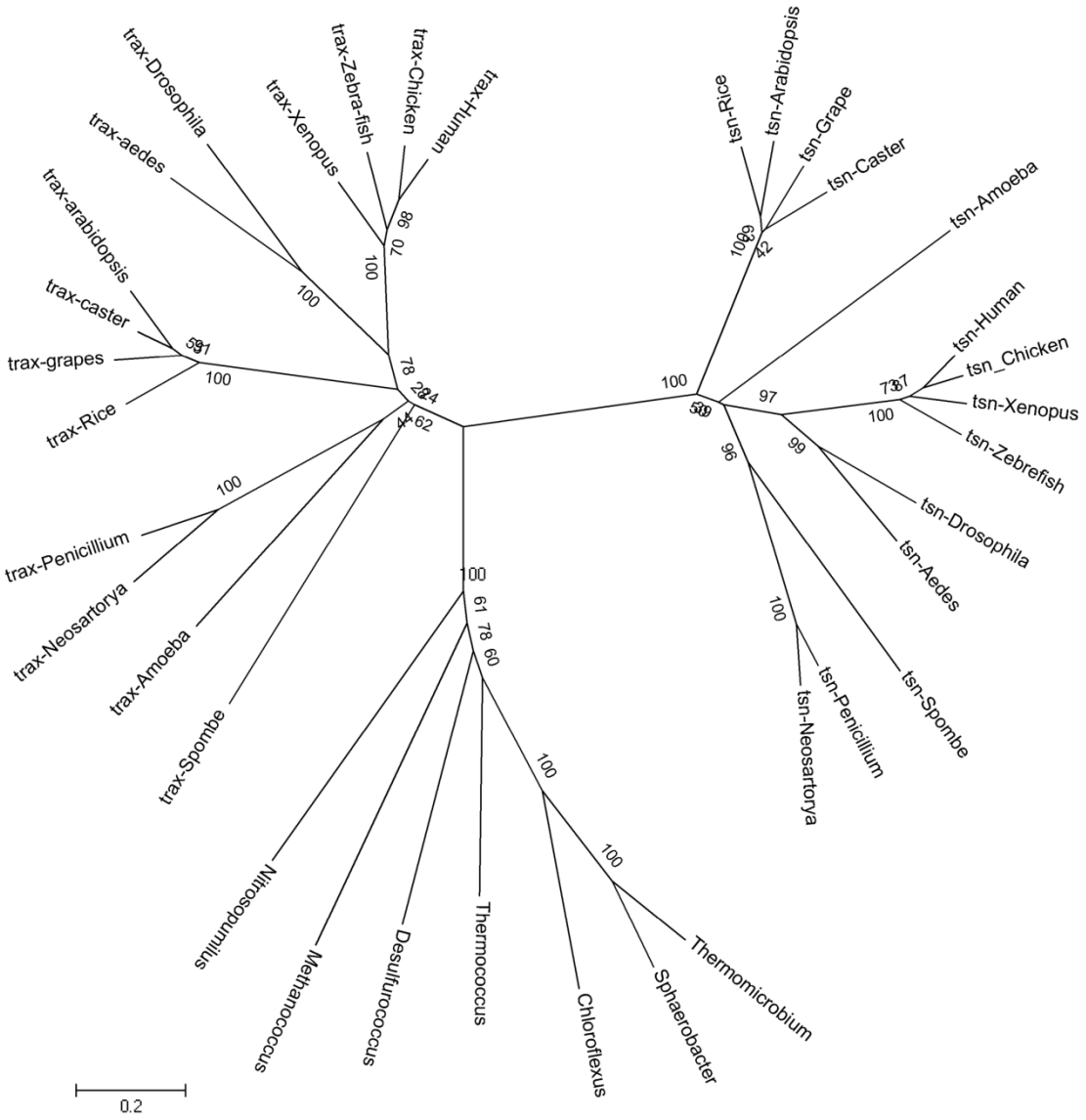


Figure 6.4 Evolutionary relationships of eukaryote translin and trax proteins with bacterial translin-like protein.

The evolutionary history was inferred using the Neighbor-Joining method. The bootstrap consensus tree inferred from 500 replicates is taken to represent the evolutionary history of the taxa analyzed. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method and are in the units of the number of amino acid substitutions per site. The analysis involved 35 sequences and all positions with less than 50% site coverage were eliminated. That is, fewer than 50% alignment gaps, missing data, and ambiguous residues were allowed at any position. There were a total of 222 positions in the final dataset. Evolutionary analyses were conducted using MEGA5.

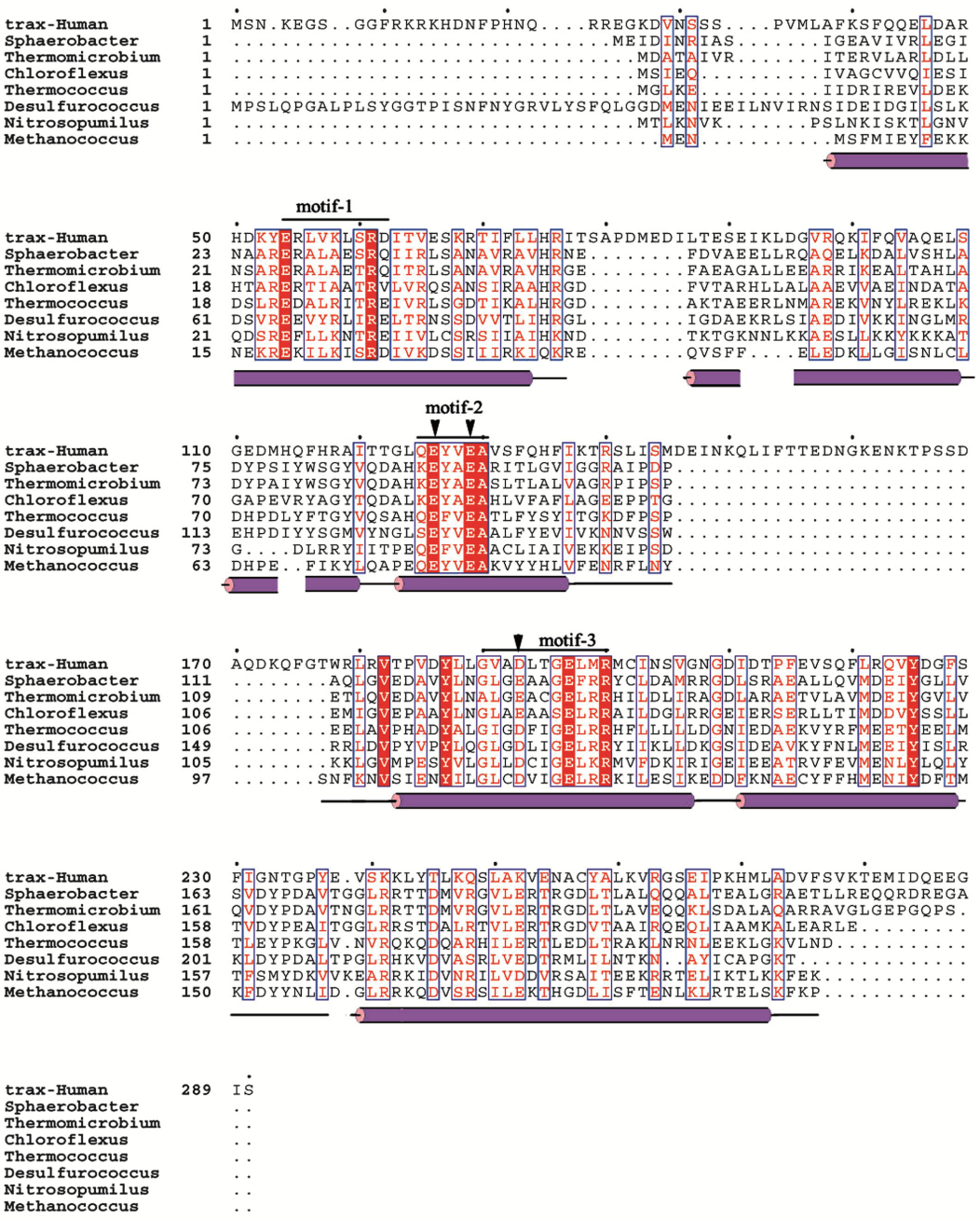


Figure 6.5 Alignment of prokaryotic translin-like proteins with human trax.

The prokaryotic translin-like protein of archaea (Methanococcus, Thermococcus, Nitrosopumilus and Desulphurococcus) and chlorobacteria (Chloroflexus, Thermomicrobium and Sphaerobacter) were aligned with human trax using ClustalW. The trax specific motifs (motifs-1 to 3) are shown here and found to be conserved in prokaryotic translin-like proteins. Residues crucial for RNase activity of eukaryotic trax are marked with ▼. The secondary structure of Methanococcus protein, predicted with PSIPRED server, is also shown. The filled cylinders represent the helices and coil regions are shown as a line.

Further to find the roots of translin, we identified at least three species in marine algae (diatom) with genes encoding single polypeptide having two translin-like domains. Intriguingly, independent analysis of each domain revealed that the N-terminal domain (domain-1) possesses the motif signature (motif-2 and motif-3) of translin proteins while C-terminal domain (domain-2) possesses trax-specific sequence signatures (Figures 6.6 and 6.7). One species of algae (*Micromonas pusilla*) was found to have a single polypeptide of length 180 amino acid (uniprot ID, C1MLX6) such that trax-specific motifs were conserved, though, protein also carried weak signatures of C-terminal residues of the translin protein (data not shown). Can this gene be precursor to the gene duplication event resulting in independent evolution of translin and trax family?

Summarily, we have identified sequence specific motifs for translin and trax family proteins. It was observed that prokaryotic translin-like protein harbors the trax specific sequence motifs and probably possesses RNase activity like eukaryotic trax. This prokaryotic protein (trax) is suggested to be the ancestor of eukaryotic trax and translin.

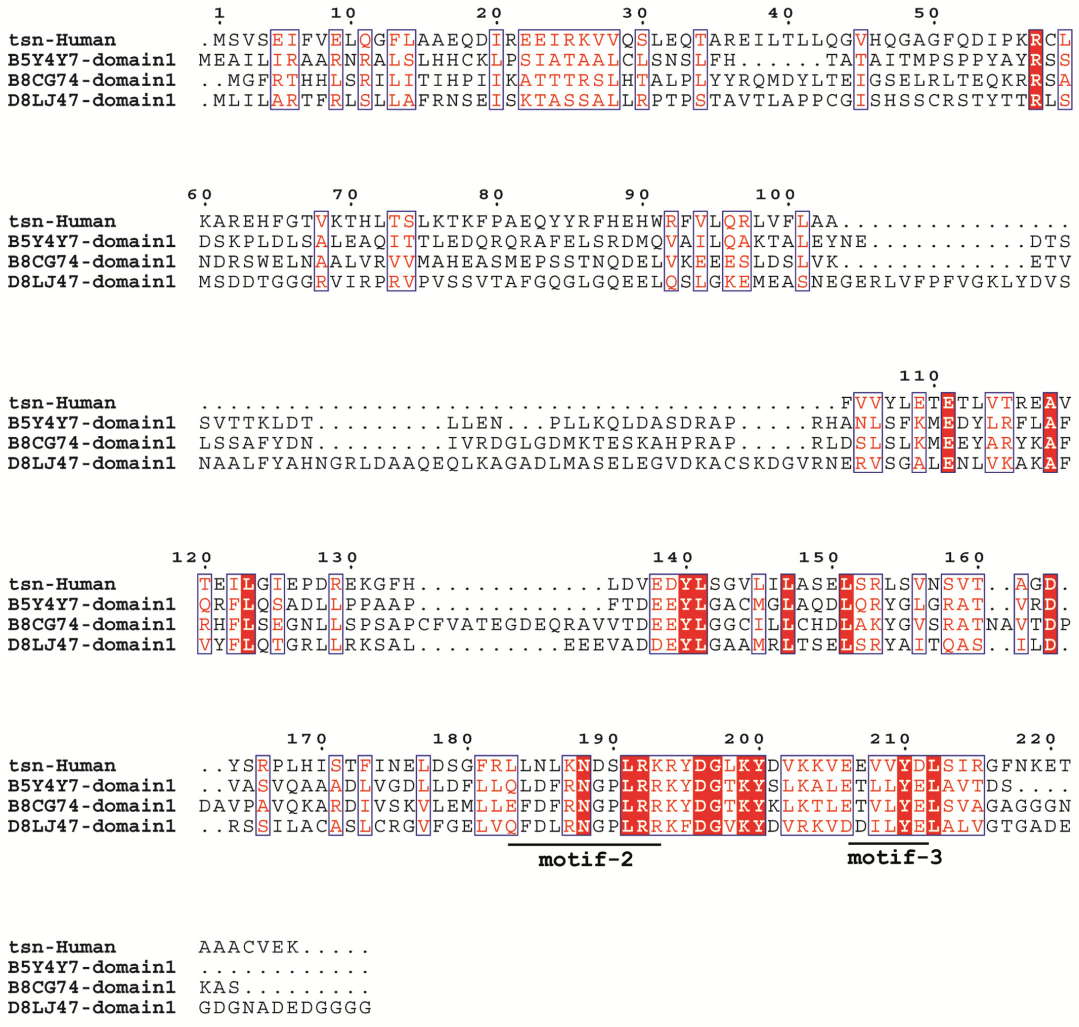


Figure 6.6 Multiple Sequence alignment of N-terminal domain of diatom proteins with human translin (tsn-Human). The uniprot ID of each diatom sequence has been used as sequence identifier in the alignment.

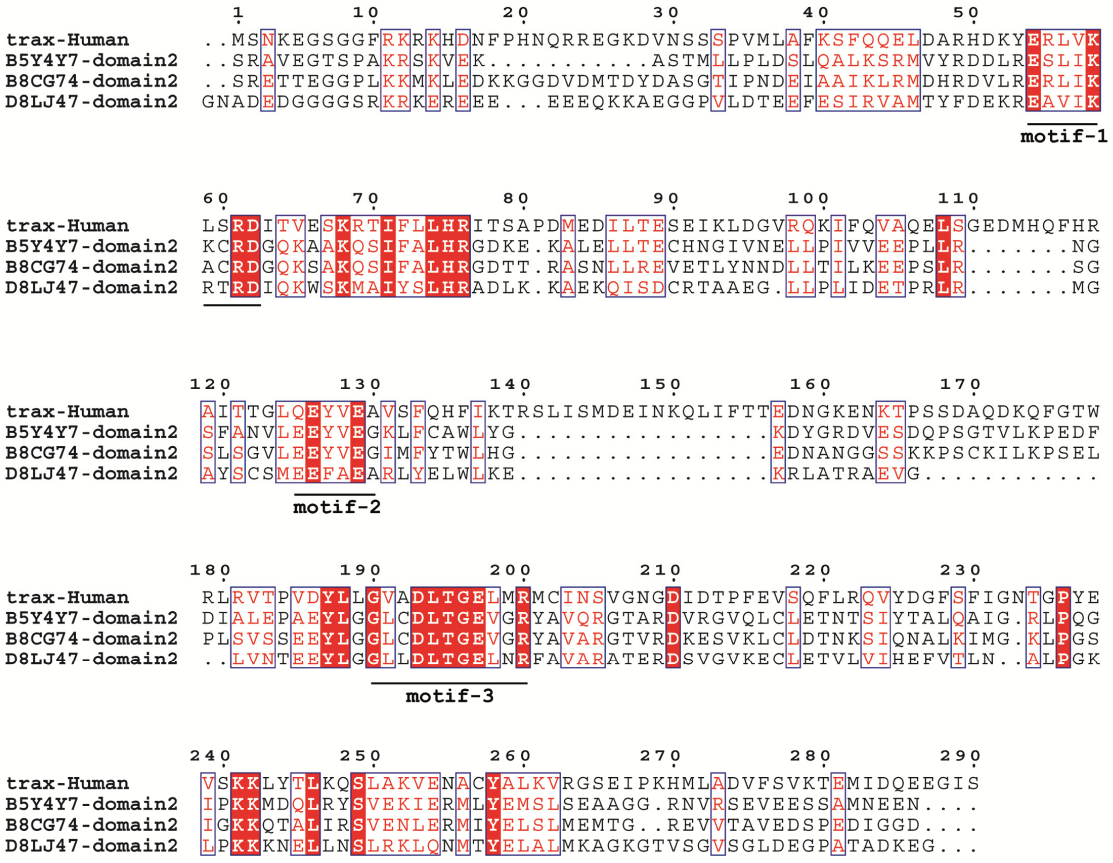


Figure 6.7 Multiple Sequence alignment of C-terminal domain of diatom proteins with human trax (trax-Human). The uniprot ID of diatom sequence has been used as sequence identifier in the alignment.

Table 6.1: Candidate sequences used to detect conserved motifs among translin proteins

Symbol-used	UniProt ID	species	% identity with human translin
tsn-Human	Q15631	*Homo sapiens* (Human)	100%
tsn-Chicken	P79769	*Gallus gallus* (Chicken)	86%
tsn-Xenopus	Q9IAM5	*Xenopus laevis* (frog)	81%
tsn-Zebrafish	Q4V9N1	*Danio rerio* (Zebrafish)	79%
tsn-Drosophila	Q7JVK6	*Drosophila melanogaster* (Fruit fly)	52%
tsn-Aedes	Q17D78	*Aedes aegypti* (yellow fever Mosquito)	52%
tsn-Grape	F6HR38	*Vitis vinifera* (Grape)	37%
tsn-Arabidopsis	Q9SJK5	*Arabidopsis thaliana*	39%
tsn-Rice	Q337S9	*Oryza sativa* (Rice)	38%
tsn-Caster	B9S0J5	*Ricinus communis* (Caster beans)	39%
tsn- Penicillium	B6QTZ6	*Penicillium marneffei*	39%
tsn-Neosartorya	Q4WY98	*Neosartorya fumigata*	39%
tsn-Spombe	Q9P7V3	*Schizosaccharomyces pombe*	37%
tsn-Amoeba*	F0ZR80	*Dictyostelium purpureum*	29%

*-not used to deduce sequence motifs

Table 6.2: Candidate sequences used to detect conserved motifs among trax proteins

Symbol-used	UniProt ID	species	% identity with human trax
trax-Human	Q99598	*Homo sapiens* (Human)	100%
trax-Chicken	E1BQ94	*Gallus gallus* (Chicken)	80%
trax-Xenopus	Q7SZ15	*Xenopus laevis* (frog)	64%
trax-Zebrafish	Q4V8Z1	*Danio rerio* (Zebrafish)	63%
trax-Drosophila	Q8I9E1	*Drosophila melanogaster* (Fruit fly)	36%
trax-Aedes	Q16T33	*Aedes aegypti* (yellow fever Mosquito)	37%
trax-Grape	D7TA79	*Vitis vinifera* (Grape)	34%
trax-Arabidopsis	Q8GZ75	*Arabidopsis thaliana*	34%
trax-Rice	B9EV41	*Oryza sativa* (Rice)	37%
trax-Caster	B9RE89	*Ricinus communis* (Caster beans)	33%
trax- Penicillium	B6Q7Y0	*Penicillium marneffei*	29%
trax-Neosartorya	B0YBV0	*Neosartorya fumigata*	28%
trax-Spombe	O74955	*Schizosaccharomyces pombe*	29%
trax-Amoeba*	F0ZNG8	*Dictyostelium purpureum*	27%

*-not used to deduce sequence motifs

Table 6.3: Prokaryotic translin-like protein

Symbol-used	Accession number	Species
Methanococcus	A6UQI9*	*Methanococcus vannielii*
Nitrosopumilus	A9A431*	*Nitrosopumilus maritimus*
Thermococcus	B6YUL0*	*Thermococcus onnurineus*
Desulfurococcus	B8D386*	*Desulfurococcus kamchatkensis*
Chloroflexus	YP_002462708**	*Chloroflexus aggregans*
Sphaerobacter	YP_003319152**	*Sphaerobacter thermophilus*
Thermomicrobium	YP_002521417**	*Thermomicrobium roseum*

*- UniProt ID

**-GenBank Accession number

Concluding remarks

In the forgoing chapters we have discussed various aspects of structure-function of translin and trax proteins, including crystal structure elucidation of translin protein (P168S mutant of drosophila translin), biochemical characterization of wild-type drosophila translin, elucidation of new DNA-binding motifs of translin, and demonstration of DNA binding activity of the trax protein. We also discussed evolution of translin and trax proteins, and deciphered unique sequence-specific motifs of each of these families.

Here I will take the opportunity to draw the attention on the conclusions drawn from the whole work presented in this thesis.

- Up-down dimer is the minimum structural unit, energetically stable, in translin proteins and different oligomeric states observed in translin orthologs can be achieved by repetition of these up-down dimers.
- The quaternary structure of translin proteins play important role in DNA binding activity and change in quaternary structure leads to loss of DNA-binding activity.
- The C-terminal residues are critical for oligomeric status of the translin proteins.
- Translin-associated factor-X, trax protein, is DNA binding competent. Importantly also, two DNA binding motifs (B2 and B3) have been identified on trax sequence using mutational studies.
- The B3 domains of both translin and trax contribute dominantly to DNA-binding activity of biologically important heteromeric translin-trax complex.
- Bacterial translin-like protein is suggested to be the ancestor of translin and trax proteins. The bacterial protein possesses trax specific sequence motifs and thus is an ortholog of eukaryotic trax. Eukaryotic translin, a paralog of trax, might have evolved after gene duplication event during the evolution of eukaryotes.

Future directions

Present studies suggested that DNA binding residues are present at the internal cavity of translin/translin-trax octameric oligomers. Also, majority of the interacting partners of translin/trax proteins are involved in DNA metabolic pathways. The work presented here may lead to insight the role trax and translin proteins play in DNA metabolism. Identification of nucleic acid binding motifs on trax may accelerate research in understanding the mechanism of RNase activity observed for translin-trax complex. However, structure of nucleic acid bound to translin or translin-trax complex will resolve the interesting questions arising due to differences observed in crystal structures and electron microscopic studies.

Functional characterization of bacterial translin-like proteins will resolve the questions of translin-independent role of the trax proteins. Study aimed to analyze the evolution of nuclear localization and nuclear export signal sequences in trax and/or translin may reveal some interesting facts about the evolution of eukaryotes and evolution of nuclear envelope. In prokaryotes, translin-like protein was observed only in Chlorobacteria and archaea, thus suggesting an evolutionary relationship. Further studies are required to find a linkage between these two subkingdoms.

Bibliography

1. Aharoni A, Baran N, Manor H (1993) Characterization of a multi-subunit human protein which selectively binds single stranded d(GA)n and d(GT)n sequence repeats in DNA. Nucleic Acids Res 21:5221–8.
2. Aisiku OR, Runnels LW, Scarlata S (2010) Identification of a novel binding partner of phospholipase cβ1: translin-associated factor X. PLoS One 5:e15001
3. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–402.
4. Aoki K, Suzuki K, Sugano T, Tasaka T, Nakahara K, Kuge O, Omori A, Kasai M (1995) A novel gene, Translin, encodes a recombination hotspot binding protein associated with chromosomal translocations. Nat Genet 10:167–74.
5. Aoki K, Ishida R, Kasai M (1997) Isolation and characterization of a cDNA encoding a Translin-like protein, TRAX. FEBS Lett 401:109–12.
6. Aoki K, Suzuki K, Ishida R, Kasai M (1999) The DNA binding activity of Translin is mediated by a basic region in the ring-shaped structure conserved in evolution. FEBS Lett 443:363–6.
7. Armougom F, Moretti S, Poirot O, Audic S, Durnas P, Schaeli B, Keduas V, Notredame C (2006) Expresso: automatic incorporation of structural information in multiple sequence alignments using 3D-Coffee. Nucleic Acids Res 34:W604–8.
8. Atlas M, Head D, Behm F, Schmidt E, Zeleznik-Le NH, Roe BA, Burian D, Domer PH (1998) Cloning and sequence analysis of four t(9;11) therapy-related leukemia breakpoints. Leukemia 12:1895–902.

9. Bagasra O, Prilliman KR (2004) RNA interference: the molecular immune system. J Mol Histol 35:545–53.
10. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The Protein Data Bank. Nucleic Acids Res 28:235–42.
11. Bhat TN (1988) Calculation of an OMIT map. J Appl Cryst 21:279–81.
12. Blow DM, Crick FHC (1959) The treatment of errors in the isomorphous replacement method. Acta Crystallogr 12:794–802.
13. Blundell TL, Johnson LN (1976) Protein Crystallography. Academic press, London.
14. Bray JD, Chennathukuzhi VM, Hecht NB (2002) Identification and characterization of cDNAs encoding four novel proteins that interact with translin associated factor-X. Genomics 79:799–808.
15. Bray JD, Chennathukuzhi VM, Hecht NB (2004) KIF2Abeta: A kinesin family member enriched in mouse male germ cells, interacts with translin associated factor-X (TRAX). Mol Reprod Dev 69:387–96.
16. Brunger AT (1997) Free R value: Cross validation in crystallography. Methods Enzymol 277:366–96.
17. Cannon TD, Hennah W, van Erp TG, Thompson PM, Lonnqvist J, Huttunen M, Gasperoni T, Tuulio-Henriksson A, Pirkola T, Toga AW, Kaprio J, Mazziotta J, Peltonen L (2005) Association of DISC1/TRAX haplotypes with schizophrenia, reduced prefrontal gray matter, and impaired short- and long-term memory. Arch Gen Psychiatry 62:1205–13.
18. Castro A, Peter M, Magnaghi-Jaulin L, Vigneron S, Loyaux D, Lorca T, Labbé JC (2000) Part of Xenopus translin is localized in the centrosomes during mitosis. Biochem Biophys Res Commun 276:515–23.

19. CCP4 (1994) The CCP4 suite: Programs for Protein Crystallography. Acta Crystallogr D50:760–3.
20. Chalk JG, Barr FG, Mitchell CD (1997) Translin recognition site sequences flank chromosome translocation breakpoints in alveolar rhabdomyosarcoma cell lines. Oncogene 15:1199–205.
21. Chayen NE, Saridakis E (2008) Protein crystallization: from purified protein to diffraction-quality crystal. Nature Methods 5:147–53.
22. Chen VB, Arendall WB 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC (2010) MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D66:12–21.
23. Chennathukuzhi VM, Kurihara Y, Bray JD, Hecht NB (2001a) Trax (translin-associated factor X), a primarily cytoplasmic protein, inhibits the binding of TB-RBP (translin) to RNA. J Biol Chem 276:13256–63.
24. Chennathukuzhi VM, Kurihara Y, Bray JD, Yang J, Hecht NB (2001b) Altering the GTP binding site of the DNA/RNA-binding protein, Translin/TB-RBP, decreases RNA binding and may create a dominant negative phenotype. Nucleic Acids Res 29:4433–40.
25. Chennathukuzhi VM, Stein JM, Abel T, Donlon S, Yang S, Miller JP, Allman DM, Simmons RA, Hecht NB (2003a) Mice deficient for testis-brain RNA-binding protein exhibit a coordinate loss of TRAX, reduced fertility, altered gene expression in the brain, and behavioral changes. Mol Cell Biol 23:6419–34.
26. Chennathukuzhi VM, Morales CR, El-Alfy M, Hecht NB (2003b) The kinesin KIF17b and RNA-binding protein TB-RBP transport specific cAMP-responsive

element modulator-regulated mRNAs in male germ cells. Proc Natl Acad Sci USA 100:15566–71.

27. Chiaruttini C, Vicario A, Li Z, Baj G, Braiuca P, Wu Y, Lee FS, Gardossi L, Baraban JM, Tongiorgi E (2009) Dendritic trafficking of BDNF mRNA is mediated by translin and blocked by the G196A (Val66Met) mutation. Proc Natl Acad Sci USA 106:16481–6.
28. Cho YS, Chennathukuzhi VM, Handel MA, Eppig J, Hecht NB (2004) The relative levels of translin-associated factor X (TRAX) and testis brain RNA-binding protein determine their nucleocytoplasmic distribution in male germ cells. J Biol Chem 279:31514–23.
29. Cho YS, Iguchi N, Yang J, Handel MA, Hecht NB (2005) Meiotic messenger RNA and noncoding RNA targets of the RNA-binding protein Translin (TSN) in mouse testis. Biol Reprod 73:840–7.
30. Claussen M, Koch R, Jin ZY, Suter B (2006) Functional characterization of Drosophila Translin and Trax. Genetics 174:1337–47.
31. Cour T, Kiemer L, Mølgaard A, Gupta R, Skriver K, Brunak S (2004) Analysis and prediction of leucine-rich nuclear export signals. Protein Eng 17:527–36.
32. Cowtan KD (1994) DM program. In Joint CCP4 ESF-EACBM Newsletter 31:34–8.
33. Cowtan KD (1998) Modified phased translation functions and their application to molecular fragment location Acta Crystallogr D54:750–6.
34. Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: A sequence logo generator. Genome Research 14:1188–90.
35. Das R, Baker D (2009) Prospects for de novo phasing with de novo protein models. Acta Crystallogr D65:169–75.

36. Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, Arendall WB 3rd, Snoeyink J, Richardson JS, Richardson DC (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res 35:W375–83.
37. Delano WL (2002) The PyMOL Molecular Graphics System. DeLano Scientific, San Carlos, CA.
38. Devon RS, Taylor MS, Millar JK, Porteous DJ (2000) Isolation and characterization of the mouse translin-associated protein X (Trax) gene. Mamm Genome 11:395–8.
39. Diller DJ, Redinbo MR, Pohl E, Hol WG (1999) A database method for automated map interpretation in protein crystallography. Proteins 36:526–41.
40. Dodson E (2003) Is it jolly SAD? Acta Crystallogr D59:1958–65.
41. Drenth J (1994) Principles of Protein X-ray Crystallography. Springer-Verlag, New York.
42. Eliahoo E, Ben Yosef R, Pérez-Cano L, Fernández-Recio J, Glaser F, Manor H (2010) Mapping of interaction sites of the *Schizosaccharomyces pombe* protein Translin with nucleic acids and proteins: a combined molecular genetics and bioinformatics study. Nucleic Acids Res 38:2975–89.
43. Emsley P, Cowtan K (2004) Coot: Model-Building Tools for Molecular Graphics. Acta Crystallogr D60:2126–32.
44. Erdemir T, Bilican B, Oncel D, Goding CR, Yavuzer U (2002a) DNA damage-dependent interaction of the nuclear matrix protein C1D with Translin-associated factor X (TRAX). J Cell Sci 115:207–16.

45. Erdemir T, Bilican B, Cagatay T, Goding CR, Yavuzer U (2002b) *Saccharomyces cerevisiae* C1D is implicated in both non-homologous DNA end joining and homologous recombination. Mol Microbiol 46:947–57.
46. Felsenstein J (1985) Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39:783–91.
47. Finkenstadt PM, Kang WS, Jeon M, Taira E, Tang W, Baraban JM (2000) Somatodendritic localization of Translin, a component of the Translin/Trax RNA binding complex. J Neurochem 75:1754–62.
48. Finkenstadt PM, Jeon M, Baraban JM (2001) Masking of the Translin/Trax complex by endogenous RNA. FEBS Lett 498:6–10.
49. Finkenstadt PM, Jeon M, Baraban JM (2002) Trax is a component of the Translin-containing RNA binding complex. J Neurochem 83:202–10.
50. Fitch WM (1970) Distinguishing homologous from analogous proteins. Syst Zool 19:99–113.
51. Fukuda Y, Ishida R, Aoki K, Nakahara K, Takashi T, Mochida K, Suzuki O, Matsuda J, Kasai M (2008) Contribution of Translin to hematopoietic regeneration after sublethal ionizing irradiation. Biol Pharm Bull 31:207–11.
52. Gouet P, Robert X, Courcelle E (2003) ESPript/ENDscript: extracting and rendering sequence and 3D information from atomic structures of proteins. Nucleic Acids Res 31:3320–3.
53. Gu W, Wu XQ, Meng XH, Morales C, Alfy M, Hecht NB (1998) The RNA- and DNA-binding protein TB-RBP is spatially and developmentally regulated during spermatogenesis. Mol Reprod Dev 49:219–28.
54. Guex N, Peitsch MC (1997) SWISS-MODEL and the Swiss-Pdb Viewer: An environment for comparative protein modeling. Electrophoresis 18:2714–23.

55. Gupta GD, Makde RD, Kamdar RP, D'Souza JS, Kulkarni MG, Kumar V, Rao BJ (2005) Co-expressed recombinant human Translin-Trax complex binds DNA. FEBS Lett 579:3141–6.

56. Gupta GD, Makde RD, Rao BJ, Kumar V (2008) Crystal structures of Drosophila mutant translin and characterization of translin variants reveal the structural plasticity of translin proteins. FEBS J 275:4235–49.

57. Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser 41:95–8.

58. Han JR, Gu W, Hecht NB (1995a) Testis-brain RNA-binding protein, a testicular translational regulatory RNA-binding protein, is present in the brain and binds to the 3' untranslated regions of transported brain mRNAs. Biol Reprod 53:707–17.

59. Han JR, Yiu GK, Hecht NB (1995b) Testis/brain RNA-binding protein attaches translationally repressed and transported mRNAs to microtubules. Proc Natl Acad Sci USA 92:9550–4.

60. Han MK, Lin P, Paek D, Harvey JJ, Fuior E, Knutson JR (2002) Fluorescence studies of pyrene maleimide-labeled translin: excimer fluorescence indicates subunits associate in a tail-to-tail configuration to form octamer. Biochemistry 41:3468–76.

61. Hasegawa T, Isobe K (1999) Evidence for the interaction between Translin and GADD34 in mammalian cells. Biochim Biophys Acta 1428:161–8.

62. Hecht NB (2000) Intracellular and intercellular transport of many germ cell mRNAs is mediated by the DNA- and RNA-binding protein, testis-brain-RNA-binding protein (TB-RBP). Mol Reprod Dev 56:252–3.

63. Hendrickson WA, Lattman EE (1970) Representation of phase probability distributions for simplified combination of independent phase information. Acta Crystallogr B26:136–43.

64. Hendrickson WA, Horton JR, LeMaster DM (1990) Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. EMBO J 9:1665–72.

65. Hennah W, Tuulio-Henriksson A, Paunio T, Ekelund J, Varilo T, Partonen T, Cannon TD, Lönnqvist J, Peltonen L (2005) A haplotype within the DISC1 gene is associated with visual memory functions in families with a high density of schizophrenia. Mol Psychiatry 10:1097–103.

66. Heras B, Martin JL (2005) Post-crystallization treatments for improving diffraction quality of protein crystals. Acta Crystallogr D61:1173–80.

67. Holm L, Park J (2000) DaliLite workbench for protein structure comparison. Bioinformatics 16:566–7.

68. Hooft RW, Vriend G, Sander C, Abola EE (1996) Errors in protein structures. Nature 381:272.

69. Hosaka T, Kanoe H, Nakayama T, Murakami H, Yamamoto H, Nakamata T, Tsuboyama T, Oka M, Kasai M, Sasaki MS, Nakamura T, Toguchida J (2000) Translin binds to the sequences adjacent to the breakpoints of the TLS and CHOP genes in liposarcomas with translocation t(12;6). Oncogene 19:5821–5.

70. Hwang S, Gou Z, Kuznetsov IB (2007) DP-Bind: a web server for sequence-based prediction of DNA-binding residues in DNA-binding proteins. Bioinformatics 23:634–6.

71. Ishida R, Okado H, Sato H, Shionoiri C, Aoki K, Kasai M (2002) A role for the octameric ring protein, Translin, in mitotic cell division. FEBS Lett 525:105–10.
72. Jacob E, Pucshansky L, Zeruya E, Baran N, Manor H (2004) The human protein translin specifically binds single-stranded microsatellite repeats, d(GT)n, and G-strand telomeric repeats, d(TTAGGG)n: a study of the binding parameters. J Mol Biol 344:939–50.
73. Jaendling A, Ramayah S, Pryce DW, McFarlane RJ (2008) Functional characterization of the Schizosaccharomyces pombe homologue of the leukaemia-associated translocation breakpoint binding protein translin and its binding partner, TRAX. Biochim Biophys Acta 1783:203–13.
74. Jaendling A, McFarlane RJ (2010) Biological roles of translin and translin-associated factor-X: RNA metabolism comes to the fore. Biochem J 429:225–34.
75. Jancarik J, Kim SH (1991) Sparse matrix sampling: a screening method for crystallization of proteins. J Appl Cryst 24:409–11.
76. Jeffs AR, Benjes SM, Smith TL, Sowerby SJ, Morris CM (1998) The BCR gene recombines preferentially with Alu elements in complex BCR-ABL translocations of chronic myeloid leukaemia. Hum Mol Genet 7:767–76.
77. Jensen RA (2001) Orthologs and paralogs -we need to get it right. Genome Boil 2:Interactions 1002.
78. Jones DT (1999) GenTHREADER: an efficient and reliable protein fold recognition method for genomic sequences. J Mol Biol 287:797–815.
79. Kaluzhny D, Laufman O, Timofeev E, Borisova O, Manor H, Shchyolkina A (2005) Conformational changes induced in the human protein translin and in the single-stranded oligodeoxynucleotides d(GT)(12) and d(TTAGGG)(5) upon

binding of these oligodeoxynucleotides by translin. J Biomol Struct Dyn 23:257–65.

80. Kanoe H, Nakayama T, Hosaka T, Murakami H, Yamamoto H, Nakashima Y, Tsuboyama T, Nakamura T, Ron D, Sasaki MS, Toguchida J (1999) Characteristics of genomic breakpoints in TLS-CHOP translocations in liposarcomas suggest the involvement of Translin and topoisomerase II in the process of translocation. Oncogene 18:721–9.
81. Kasai M, Matsuzaki T, Katayanagi K, Omori A, Maziarz RT, Strominger JL, Aoki K, Suzuki K (1997) The translin ring specifically recognizes DNA ends at recombination hot spots in the human genome. J Biol Chem 272:11402–7.
82. Kendrew JC, Bodo G, Dintzis HM, Parrish RG, Wyckoff H, Phillips DC (1958) A three-dimensional model of the myoglobin molecule obtained by X-ray analysis. Nature 181:662–6.
83. Keuerleber S, Gsandtner I, Freissmuth M (2011) From cradle to twilight: the.carboxyl terminus directs the fate of the A(2A)-adenosine receptor. Biochim Biophys Acta 1808:1350–7.
84. Kishi T, Okochi T, Kitajima T, Ujike H, Inada T, Yamada M, Uchimura N, Sora I, Iyo M, Ozaki N, Correll CU, Iwata N (2011) Lack of association between translin-associated factor X gene (TSNAX) and methamphetamine dependence in the Japanese population. Prog Neuropsychopharmacol Biol Psychiatry 35:1618–22.
85. Kwon YK, Hecht NB (1991) Cytoplasmic protein binding to highly conserved sequences in the 3' untranslated region of mouse protamine 2 mRNA, a translationally regulated transcript of male germ cells. Proc Natl Acad Sci USA 88:3584–8.

86. Kobayashi S, Takashima A, Anzai K (1998) The dendritic translocation of translin protein in the form of BC1 RNA protein particles in developing rat hippocampal neurons in primary culture. Biochem Biophys Res Commun 253:448–53.

87. Koonin EV (2005) Orthologs, paralogs, and evolutionary genomics. Annu Rev Genet 39:309–38.

88. Krissinel E, Henrick K (2007) Inference of macromolecular assemblies from crystalline state. J Mol Biol 372:774–97.

89. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: A program to check the stereochemical quality of protein structures. J Appl Cryst 26:283–91.

90. Laufman O, Ben Yosef R, Adir N, Manor H (2005) Cloning and characterization of the Schizosaccharomyces pombe homologs of the human protein Translin and the Translin-associated protein TRAX. Nucleic Acids Res 33:4128–39.

91. Lee SP, Fuior E, Lewis MS, Han MK (2001) Analytical ultracentrifugation studies of translin: analysis of protein-DNA interactions using a single-stranded fluorogenic oligonucleotide. Biochemistry 40:14081–8.

92. Lewin B (2004) Genes VIII. Pearson Prentice Hall, New Jersey.

93. Li Z, Baraban JM (2004) High affinity binding of the Translin/Trax complex to RNA does not require the presence of Y or H elements. Brain Res Mol Brain Res 120:123–9.

94. Li Z, Wu Y, Baraban JM (2008) The Translin/Trax RNA binding complex: clues to function in the nervous system. Biochim Biophys Acta 1779:479–85.

95. Liu Y, Ye X, Jiang F, Liang C, Chen D, Peng J, Kinch LN, Grishin NV, Liu Q (2009) C3PO, an endoribonuclease that promotes RNAi by facilitating RISC activation. Science 325:750–3.
96. Lluis M, Hoe W, Schleit J, Robertus J (2010) Analysis of nucleic acid binding by a recombinant translin-trax complex. Biochem Biophys Res Commun 396:709–13.
97. McGuffin LJ, Bryson K, Jones DT (2000) The PSIPRED protein structure prediction server. Bioinformatics 16:404–5.
98. Matthews BW (1968) Solvent content of protein crystals. J Mol Biol 33:491–7.
99. Mellon SH, Bair SR, Depoix C, Vigne JL, Hecht NB, Brake PB (2006) Translin coactivates steroidogenic factor-1–stimulated transcription. Mol Endocrinol 21:89–105.
100. Miller GL (1959) Protein determination for large number of samples. Anal Chem 31:964.
101. Miller R, Gallo SM, Khalak HG, Weeks CM (1994) SnB: crystal structure determination via shake-and-bake J Appl Cryst 27:613–21.
102. Millar JK, Christie S, Semple CA, Porteous DJ (2000) Chromosomal location and genomic structure of the human translin-associated factor X gene (TRAX; TSNAX) revealed by intergenic splicing to DISC1, a gene disrupted by a translocation segregating with schizophrenia. Genomics 67:69–77.
103. Morales CR, Wu XQ, Hecht NB (1998) The DNA/RNA-binding protein, TB-RBP, moves from the nucleus to the cytoplasm and through intercellular bridges in male germ cells. Dev Biol 201:113–23.
104. Morris RJ, Bricogne G (2003) Sheldrick's 1.2 Å rule and beyond. Acta Crystallogr D59:615–7.

105. Morris RJ, Zwart PH, Cohen S, Fernandez FJ, Kakaris M, Kirillova O, Vonrhein C, Perrakis A, Lamzin VS (2004) Breaking good resolutions with ARP/wARP. J Synchrotron Rad 11:56–9

106. Muramatsu T, Ohmae A, Anzai K (1998) BC1 RNA protein particles in mouse brain contain two y-,h-element-binding proteins, translin and a 37 kDa protein. Biochem Biophys Res Commun 247:7–11.

107. Navaza J (1994) AMoRe: an automated package for molecular replacement. Acta Crystallogr A50:157–63.

108. Nelson LD, Cox MM (2005) Lehninger: Principles of Biochemistry, 4th ed. Freeman WH & Co, New York.

109. Okuda A, Kishi T, Okochi T, Ikeda M, Kitajima T, Tsunoka T, Okumukura T, Fukuo Y, Kinoshita Y, Kawashima K, Yamanouchi Y, Inada T, Ozaki N, Iwata N (2009) Translin-associated factor X gene (TSNAX) may be associated with female major depressive disorder in the Japanese population. Neuromolecular Med 12:78–85.

110. Otwinowski Z, Minor W **(**1997) *In* Processing of X-ray Diffraction Data Collected in Oscillation Mode, Methods in Enzymology (Carter CW & Sweet R.M, eds.) 276:307–26.

111. Palo OM, Antila M, Silander K, Hennah W, Kilpinen H, Soronen P, Tuulio-Henriksson A, Kieseppä T, Partonen T, Lönnqvist J, Peltonen L, Paunio T (2007) Association of distinct allelic haplotypes of DISC1 with psychotic and bipolar spectrum disorders and with underlying cognitive impairments. Hum Mol Genet 16:2517–28.

112. Pascal JM, Hart PJ, Hecht NB, Robertus JD (2002) Crystal Structure of TB-RBP, a novel RNA-binding and regulating protein. J Mol Biol 319:1049–57.

113. Perez-Iratxeta C, Andrade-Navarro MA (2008) K2D2: estimation of protein secondary structure from circular dichroism spectra. BMC Struct Biol 8:25.

114. Perrakis A, Morris R, Lazmin VS (1999) Automated protein model building combined with iterative structure refinement. Nat Struct Biol 6:458–63.

115. Perutz MF (1956) Isomorphous replacement and phase determination in non-centrosymmetric space groups. Acta Crystallogr 9:867–73.

116. Powell HR (1999) The Rossmann Fourier autoindexing algorithm in MOSFLM. Acta Crystallogr D55:1690–5.

117. Read RJ (1986) Improved Fourier coefficients for maps using phases from partial structure with errors. Acta Crystallogr A42:140–9.

118. Remm M, Storm CE, Sonnhammer EL (2001) Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. J Mol Biol 314:1041–52.

119. Rossmann MG, Blow DM (1962) The detection of sub-units within the crystallographic asymmetric unit. Acta Crystallogr 15:24–31.

120. Saitou N, Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–25.

121. Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 234:779–815.

122. Sambrook J, Russell DW (2001) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, NY.

123. Schneider TD, Stephens RM (1990) Sequence Logos: A New Way to Display Consensus Sequences. Nucleic Acids Res 18:6097–100

124. Schneider TR, Sheldrick (2002) GM Substructure solution with SHELXD. Acta Crystallogr D58:1772–9.

125. Schröder GF, Levitt M, Brunger AT (2010) Super-resolution biomolecular crystallography with low-resolution data. Nature 464:1218–22.

126. Schröer U, Volk GF, Liedtke T, Thanos S (2007) Translin-associated factor-X (Trax) is a molecular switch of growth-associated protein (GAP)-43 that controls axonal regeneration. Eur J Neurosci 26:2169–78.

127. Schwede T, Kopp J, Guex N, Peitsch MC (2003) SWISS-MODEL: An automated protein homology-modeling server. Nucleic Acids Res 31:3381–5.

128. Sengupta K, Kamdar RP, D'Souza JS, Mustafi SM, Rao BJ (2006) GTP-induced conformational changes in translin: a comparison between human and Drosophila proteins. Biochemistry 45:861–70.

129. Sengupta K, Rao BJ (2002) Translin binding to DNA: recruitment through DNA ends and consequent conformational transitions. Biochemistry 41:15315–26.

130. Sheldrick GM (1990) Phase annealing in SHELX-90: direct methods for larger structures. Acta Crystallogr A46:467–73.

131. Stein JM, Bergman W, Fang Y, Davison L, Brensinger C, Robinson MB, Hecht NB, Abel T (2006) Behavioral and neurochemical alterations in mice lacking the RNA-binding protein translin. J Neurosci 26:2184–96.

132. Sugiura I, Sasaki C, Hasegawa T, Kohno T, Sugio S, Moriyama H, Kasai M, Matsuzaki T (2004) Structure of human translin at 2.2 A resolution. Acta Crystallogr D60:674–9.

133. Sun CN, Cheng HC, Chou JL, Lee SY, Lin YW, Lai HL, Chen HM, Chern Y (2006) Rescue of p53 blockage by the A(2A) adenosine receptor via a novel interacting protein, translin-associated protein X. Mol Pharmacol 70:454–66.

134. Sun CN, Chuang HC, Wang JY, Chen SY, Cheng YY, Lee CF, Chern Y (2010) The A2A adenosine receptor rescues neuritogenesis impaired by p53 blockage via KIF2A, a kinesin family member. Dev Neurobiol 70:604–21.

135. Suseendranathan K, Sengupta K, Rikhy R, D'Souza JS, Kokkanti M, Kulkarni MG, Kamdar R, Changede R, Sinha R, Subramanian L, Singh K, Rodrigues V, Rao BJ (2007) Expression pattern of Drosophila translin and behavioral analyses of the mutant. Eur J Cell Biol 86:173–86.

136. Taira E, Finkenstadt PM, Baraban JM (1998) Identification of translin and trax as components of the GS1 strand-specific DNA binding complex enriched in brain. J Neurochem 71:471–7.

137. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Mol Biol Evol 28:2731–9.

138. Taylor G (2003) The phase problem. Acta Crystallogr D59:1881–90.

139. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–80.

140. Thomson PA, Wray NR, Millar JK, Evans KL, Hellard SL, Condie A, Muir WJ, Blackwood DH, Porteous DJ (2005) Association between the TRAX/DISC locus and both bipolar disorder and schizophrenia in the Scottish population. Mol Psychiatry 10:657–68.

141. Tian Y, Simanshu DK, Ascano M, Diaz-Avalos R, Park AY, Juranek SA, Rice WJ, Yin Q, Robinson CV, Tuschl T, Patel DJ (2011) Multimeric assembly and

biochemical characterization of the Trax-translin endonuclease complex. Nat Struct Mol Biol 18:658–64.

142. Trabuco LG, Villa E, Mitra K, Frank J, Schulten K (2008) Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. Structure16:673–83.
143. Tronrud DE (2004) Introduction to macromolecular refinement. Acta Crystallogr D60:2156–68.
144. Usón I, Sheldrick GM (1999) Advances in direct methods for protein crystallography. Curr Opin Struct Biol 9:643–8.
145. Vagin AA, Teplyakov A (1997) MOLREP: an automated program for molecular replacement. J Appl Cryst 30:1022–5.
146. Vagin AA, Steiner RA, Lebedev AA, Potterton L, McNicholas S, Long F, Murshudov GN (2004) REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. Acta Crystallogr D60:2184–95.
147. VanLoock MS, Yu X, Kasai M, Egelman EH (2001) Electron microscopic studies of the translin octameric ring. J Struct Biol 135:58–66.
148. Wang J, Boja ES, Oubrahim H, Chock PB (2004) Testis brain ribonucleic acid-binding protein/translin possesses both single-stranded and double-stranded ribonuclease activities. Biochemistry 43:13424–31.
149. Wei Y, Sun M, Nilsson G, Dwight T, Xie Y, Wang J, Hou Y, Larsson O, Larsson C, Zhu X (2003) Characteristic sequence motifs located at the genomic breakpoints of the translocation t(X;18) in synovial sarcomas. Oncogene 22:2215–22.

150. Wilmanns M, Eisenberg D (1993) Three-dimensional profiles from residue-pair preferences: identification of sequences with beta/alpha-barrel fold. Proc Natl Acad Sci USA 90:1379–83.

151. Wu RF, Osatomi K, Terada LS, Uyeda K (2003) Identification of Translin/Trax complex as a glucose response element binding protein in liver. Biochim Biophys Acta 1624:29–35.

152. Wu XQ, Gu W, Meng X, Hecht NB (1997) The RNA-binding protein, TB-RBP, is the mouse homologue of translin, a recombination protein associated with chromosomal translocations. Proc Natl Acad Sci USA 94:5640–5.

153. Wu XQ, Xu L, Hecht NB (1998) Dimerization of the testis brain RNA-binding protein (translin) is mediated through its C-terminus and is required for DNA- and RNA-binding. Nucleic Acids Res 26:1675–80.

154. Wu XQ, Lefrancois S, Morales CR, Hecht NB (1999a) Protein-protein interactions between the testis brain RNA-binding protein and the transitional endoplasmic reticulum ATPase, a cytoskeletal gamma actin and Trax in male germ cells and the brain. Biochemistry 38:11261–70.

155. Wu XQ, Petrusz P, Hecht NB (1999b) Testis-brain RNA-binding protein (Translin) is primarily expressed in neurons of the mouse brain. Brain Res 819:174–8.

156. Wu XQ, Hecht NB (2000) Mouse testis brain ribonucleic acid-binding protein/translin colocalizes with microtubules and is immunoprecipitated with messenger ribonucleic acids encoding myelin basic protein, alpha calmodulin kinase II, and protamines 1 and 2. Biol Reprod 62:720–5.

157. Wu YC, Williamson R, Li Z, Vicario A, Xu J, Kasai M, Chern Y, Tongiorgi E, Baraban JM (2011) Dendritic trafficking of brain-derived neurotrophic factor

mRNA: regulation by translin-dependent and -independent mechanisms. J Neurochem 116:1112–21.

158. Xiang H, Wang J, Hisaoka M, Zhu X (2008) Characteristic sequence motifs located at the genomic breakpoints of the translocation t(12;16) and t(12;22) in myxoid liposarcoma. Pathology 40:547–52.
159. Yang J, Chennathukuzhi V, Miki K, O'Brien DA, Hecht NB (2003) Mouse testis brain RNA-binding protein/translin selectively binds to the messenger RNA of the fibrous sheath protein glyceraldehyde 3–phosphate dehydrogenase-S and suppresses its translation in vitro. Biol Reprod 68:853–9.
160. Yang S, Cho YS, Chennathukuzhi VM, Underkoffler LA, Loomes K, Hecht NB (2004) Translin-associated factor X is post-transcriptionally regulated by its partner protein TB-RBP, and both are essential for normal cell proliferation. J Biol Chem 279:12605–14.
161. Ye X, Huang N, Liu Y, Paroo Z, Huerta C, Li P, Chen S, Liu Q, Zhang H (2011) Structure of C3PO and mechanism of human RISC activation. Nat Struct Mol Biol 18:650–7.
162. Yu Z, Hecht NB (2008) The DNA/RNA-binding protein, translin, binds microRNA122a and increases its in vivo stability. J Androl 29:572–9.
163. Zezula J, Freissmuth M (2008) The A(2A)-adenosine receptor: a GPCR with unique features? Br J Pharmacol 153:S184–90.
164. Zuckerkandl E, Pauling L (1965) Evolutionary divergence and convergence in proteins. Edited in Evolving Genes and Proteins by Bryson V, Vogel HJ, pp. 97–166. Academic Press, New York.

Publications

Publications in international journals

The following manuscripts, relevant to work presented in the thesis, have been published /communicated.

- **Gupta GD**, Makde RD, Rao BJ, Kumar V (2008) Crystal structures of *drosophila* mutant translin and characterization of translin variants reveal the structural plasticity of translin proteins. FEBS J 275:4235-49.
- **Gupta GD**, Kumar V (2012) Identification of nucleic acid binding sites on Translin-associated factor X (TRAX) Protein. PLoS ONE 7:e33035.
- Kumar V, **Gupta GD** (2012) Low-resolution structure of drosophila translin. FEBS Openbio 2:37-46.
- **Gupta GD**, Kale A, Kumar V (2012) Molecular evolution of translin superfamily proteins within the genomes of eubacteria, archaea and eukaryotes. Submitted**.**

Entries of protein structures in the Protein Data Bank (PDB)

- **Gupta GD,** Makde RD, Kumar V (2007) Crystal structure of Drosophila melanogaster translin protein. **PDB-ID 2QRX**.
- **Gupta GD,** Makde RD, Kumar V (2007) Crystal structure of Drosophila melanogaster translin protein. **PDB-ID 2QVA**.
- Kumar V, **Gupta GD** (2012) Crystal structure of Drosophila melanogaster translin protein. **PDB-ID 4DG7**.