Abstract

Transposable elements (TEs) are DNA segments that can move from one position in the genome to another, irrespective of any sequence homology between the excision and the integration sites. The movement of TEs is mediated by transposase proteins (maybe encoded by TE). TEs can provide raw material for generation of new protein coding genes in the host genome by the process of exaptation. The host often conserves the ancestral or slightly modified function of the TE-encoded protein while the ability to autonomously transpose is lost during evolution (molecular domestication). Studying active as well as domesticated transposases is of utmost importance for the basic understanding of how the conserved molecular protein architecture is adapted for diverse cellular functions over the course of evolution. For my doctoral dissertation, I have studied the conserved and divergent modular protein architectures of the well-studied *Drosophila* P element transposase (DmTNP; active transposase) and human THAP9 (hTHAP9; a DmTNP homolog of unknown function).

DmTNP is extensively characterized whereas hTHAP9 is only recently discovered. Based on sequence and structural homology, the domain organisation of hTHAP9 was found to be similar to DmTNP. An amino terminal zinc coordinating DNA binding domain (THAP domain) is conserved between both DmTNP and hTHAP9. Interestingly, the THAP domain is also conserved amongst eleven other human proteins, which are together grouped into the human THAP protein family. Interestingly, DmTNP as well as some human THAP proteins are known to form functional oligomers.

In the **first chapter**, I have adopted bioinformatic to explore the possibility of oligomerisation of human THAP proteins via a coiled coil region. This study identified a 40-

residue long leucine rich region, which had an amphipathic pattern similar to leucine zipper regions (Fig. A), in all the twelve human THAP proteins including hTHAP9(Fig. B).

The **second chapter** is an experimental continuation of the first chapter wherein I have established that hTHAP9 undergoes homo-oligomerisation using a combination of *in silico* prediction tools as well as biochemical techniques. Surprisingly, it was demonstrated that the amino-terminal regions of both hTHAP9 and DmTNP (after deletion of the predicted coiled coil region or mutating the leucines, either individually or together), retained the ability to form higher order oligomeric states (Fig. C). Moreover, this study also revealed that the oligomerisation of both hTHAP9 and DmTNP is partially mediated by DNA (Fig. D).

DmTNP binds GTP via an identified GTP binding region (G domain) as a prerequisite to forming the tetrameric synaptic paired end complex (PEC) bound to both ends of transposon DNA. However, there are no reports of GTP hydrolysis by DmTNP. Thus, it is speculated that GTP binding leads to DmTNP oligomerisation. It was interesting to speculate that hTHAP may also bind GTP as a prerequisite of oligomerisation. In the **third chapter**, I have predicted a GTP binding region (G-domain) in hTHAP9 by writing a customized Spacers and Mismatch Algorithm (SMA) (Fig. E) after failing to do the same using available *in silico* tools as well as an *in vivo* GFP solubility screen. The predicted G domains of hTHAP9 and DmTNP are now being used for experimental studies. **The fourth chapter** describes the extension of SMA such that it can be used to predict putative G boxes in any protein.
The DNA binding site of hTHAP9 is hitherto unknown. The DNA binding domain of hTHAP9 is the THAP domain. Despite having a similar structural fold, the THAP domains of different human THAP proteins have different DNA binding sites. Thus, in the **fifth chapter**, I have studied the synapomorphic (sequence and structural) variations in the THAP domains

of human THAP family proteins (Fig. F) and their homologs using *in silico* motif prediction and secondary structure prediction tools. This study independently identified THAP domain size- and sequence-based sub-groups amongst the homologs of each human THAP protein.

Overall, the results of my thesis will direct future research towards better understanding of the evolution and cellular function of hTHAP9. The **first chapter** can be continued to study the possible protein interaction partners of human THAP proteins. This will also help advance the understanding of hTHAP9 hetero-oligomerisation which may be mediated either by one short motif or many motifs across the length of the protein as is suggested in the **second chapter**. The identification of a putative G domain in hTHAP9 in the **third and fourth chapter** is already being used to study possible GTP binding to hTHAP9 *in vitro*. The novel findings from the **fifth chapter** will not only help the ongoing studies to identify the DNA binding site of hTHAP9 but it can also be used to understand the evolution of the THAP domain in animals.