

Remote homologue identification of *Drosophila* GAGA factor in mouse

Suresh Kumar

Department of Bioinformatics, School of Biotechnology and Health Sciences, Karunya University, Coimbatore -641114, Tamil Nadu, India; Suresh Kumar - Email: sureshbio@gmail.com; *Corresponding author

Received August 13, 2011; Accepted August 13, 2011; Published August 20, 2011

Abstract:

GAGA factor (GAF) is involved in both gene activation and gene repression and plays a role in the modulation of chromatin structure. In *Drosophila*, Trithorax like (Trl) gene encodes the DNA binding protein called GAGA factor (GAF). Trl-GAF binds to GAGA sites through its C2H2 zinc finger domain and has an N-terminal BTB/POZ domain. Identification of Trl-GAF homologue in mouse helps in deeper understanding of the mechanism and function. Conventional alignment tools such as BLAST and FASTA cannot identify homologues in mouse genome as their sequence identity is below 30%. In the present study, various sequence and structure analyses were followed for the detection of remote homologues of *Drosophila* GAGA FACTOR in mouse to identify as Zbtb3. Through homology modeling and docking approach, the zinc finger region of mouse Zbtb3 showed conserved residues and favorable DNA binding sites with GAGA sites similar to that of *Drosophila* GAGA FACTOR.

Keywords: Remote homologue, GAGA factor, BTB, Zinc finger, Zbtb3, molecular modeling

Background:

In *Drosophila*, chromatin factors such as Polycomb group (PcG) and trithorax group (trxG) proteins play a crucial role in the development of gene regulation. One of the trxG group gene is Trithorax like (Trl), also known as GAGA factor (GAF). It is identified as a sequence specific DNA binding protein, which could stimulate transcriptional activity [1]. GAGA promotes transcription by blocking repressive effects of histones. GAGA is required for proper expression of variety of development loci that contain GAGA binding sites in their upstream regulatory regions. Trl-GAF binds to GAGA sites through its C2H2 zinc finger domain and has an N-terminal BTB/PZ (broad complex tramtrack bric a brac/poxvirus zinc finger) protein-protein interaction domain. At N terminus of GAGA there is BTB/PZ domain required for oligomerization which is highly conserved and involves in protein-protein interaction. The C-terminal glutamine-rich region is required for transcriptional activation. Members of GAGA factor family bind to a 5'-GAGAG-3' DNA consensus binding site, and contain a Cys2-His2 zinc finger core as well as an N-terminal extension containing two highly basic regions. The zinc finger core binds in the DNA major groove and recognizes the first three GAG bases of the consensus in a

manner similar to that seen in other classical zinc finger-DNA complexes [2]. It has been shown that GAGA factor has multiple roles in regulating genes during *Drosophila* development. A convenient way to study the function, structure and mechanism of a gene is to identify homologues (evolutionary relationships) in model organisms. It is interesting to identify homologues of GAGA factor in mouse. Dynamic programming based alignment tools such as BLAST and FASTA have been widely used to provide evidence for homology by matching a new sequence against a database of previously annotated sequences. However, these approaches can only detect homologous proteins that exhibit significant sequence similarity. But if sequence identity is below 30% it may be very difficult to detect homologue and they are called as remote homologue which share the same evolutionary ancestry. Accurate detection of homologue at low levels of sequence identity still remains a challenging problem. Recently, the vertebrate homologue of *Drosophila* GAGA factor was identified in mouse as c-Krox/Th-POK gene encoded by zbtb7b [3]. In this present study through insilico analysis another probable homologue of *Drosophila* GAGA factor in mouse is reported as Zbtb3 (Zinc finger and BTB domain-containing protein 3).

Materials and Methodology:

Sequence retrieval:

The sequence of *Drosophila* GAGA factor was retrieved from the UniprotKB (Q08605) and the domain boundaries were taken from Uniprot annotations. *Drosophila* GAGA factor contains BTB domain at N-terminus and zinc finger domain at C-terminus.

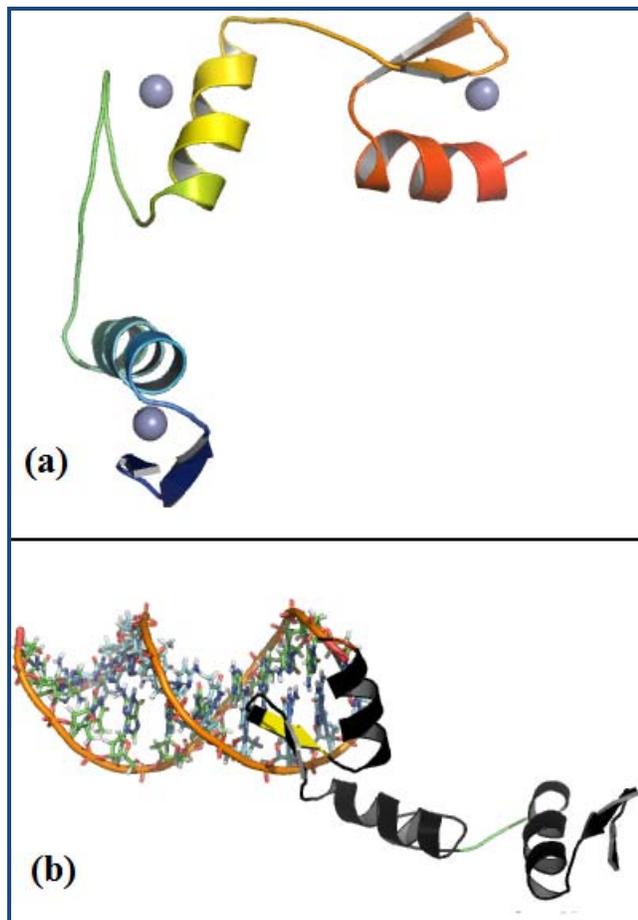


Figure 1: (a) Homology Model of zinc finger domain of Zbtb3 consisting of N and C terminus in the presence of zinc ions shown as spheres in cyan and (b) Docked zinc finger with DNA showing conserved TGEKP residues at linker region in green, glycine residue in red and AGC sequence is shown in yellow.

Remote Homology Detection Methodology:

Similarity search:

The retrieved sequence of GAGA factor (GAF) was searched against mouse genome database through Protein-Protein Blast (BlastP) of NCBI with default parameters.

PSI-Blast search:

PSI-Blast makes use of iterated BLAST searches in order to extend the number of evolutionary relationships detected. Related proteins were searched using PSI-BLAST [4] with five iterations using the top twenty maximum hits.

Hidden Markow model and protein family search:

The *Drosophila* GAGA factor sequence was used for searching databases of protein sequence using profile Hidden Markow

Models (HMMs) for detection of remote homologs using the server HHpred [5]. Through this search about eight probable homologue clusters were obtained. The entire eight probable clusters were crosschecked with pfam database of protein families [http://pfam.sanger.ac.uk/].

Multiple sequence alignment and phylogenetic analysis:

From the eight probable homologue clusters, protein sequences were obtained and multiple alignment was constructed using ClustalW [6]. The phylogenetic tree was constructed with distance matrix method.

Homology modeling:

The sequence of mouse zbtb3 was obtained from the Uniprot database (Q91X45) and the sequence was queried against the Protein DataBank (PDB) using PSI-BLAST. Through the PSI-BLAST results, template, which has high sequence identity and fewer gaps were selected. Comparative modeling of both BTB domain and Zinc finger domain of Zbtb3 was performed by using Modeller software version 9v5 [7]. The resulting model was evaluated using Verify3D [8] and Procheck. The quality of models was assessed by Ramachandran plot analysis [9]. The structures were visualized using pymol.

Docking of Zbtb3 and GAGA-DNA complex:

The homology modeled zinc finger domain of Zbtb3 was further studied to know how they interact with GAGA-DNA complex through docking approach. NMR structure of GAGA factor/DNA complex (PDB: 1YUI), which binds to the single zinc finger, was used. Zinc finger was removed and only GAGA factor/DNA complex was used for docking with homology modeled zinc finger of Zbtb3. The modeled zinc finger domain was docked with GAGA-DNA complex (1YUI) using Hex software [10].

Results and Discussion:

Sequence analysis:

Through BlastP search it was found that *Drosophila* GAGA factor had very less identity (below 20%) with mouse. Since sequence identity of the *Drosophila* GAGA factor with mouse was below the twilight zone, various rigorous methodologies were followed for the detection of remote homologues. PSI-Blast search method groups the closely related protein sequences and reflects the information from a protein family better than pair wise sequence comparison method like Blast. PSI-Blast based search was carried out with default parameters against mouse genome with five iterations. Careful selection of the proteins containing both BTB and Zinc finger domain and removal of redundant proteins resulted in about twenty probable proteins. Profile Hidden Markov Model (HMMs) is one of the most powerful approaches for remote homolog identification. The mouse protein database was searched by profile Hidden Markow Models (HMMs) with HHPred and eight probable remote homologues were selected (**Table 1, see supplementary material**). Crosschecking the entire eight clusters with pfam database of protein families [11] confirmed that they belong to the family of BTB and Zf-C2H2. Through pair wise sequence alignment, comparison of GAGA factor with the eight probable remote homologue clusters was carried out by EMBOSS using the Needleman-Wunsch alignment algorithm to find out the optimum alignment of two sequences along their entire length [12]. Zbtb3 was found to have

maximum sequence identity in both domain wise and in whole sequence alignment (**Table 1**, see **Supplementary material**). Phylogenetic analysis with distance analysis suggests that *zbtb3* was close to GAGA factor. Recently, it is reported that large number of transcription factors are mostly intrinsically disordered [13]. From the Disprot database [14], it is inferred that disordered residues were between regions 137-519 for *Drosophila* GAGA. The *Zbtb3* predicted with consensus disorder prediction method [15] resulted in the disordered residues placed between regions 134-498 which agreed well with *Drosophila* GAGA factor. Intrinsic plasticity enables a disorder region to recognize and bind many biological targets with high specificity.

Structural analysis:

Comparative modeling of *Zbtb3* for both BTB domain and Zinc finger domain was carried out using Modeller program. For modeling of BTB domain, template of crystal structure of the Btb domain of human myoneurin from the PDB ID 2VPK was used. For modeling of Zinc finger domain, template of crystal structure of a designed zinc finger protein from the PDB ID 1Mey was used. The sequence alignment of target and template for both BTB domain and Zinc finger domain was about 40-42%. Individual alignments were given as input to Modeller to build 3D structures, and the resulting models were evaluated using Verify 3D and Procheck. This analysis has led to the conclusion that the models were more reliable. Homology model of *zbtb3* zinc finger consists of three finger coordinate zinc ions with a combination of cysteine and histidine residues (**Figure 1a**). As *Drosophila* GAGA factor binds well with DNA, it is important to analyze and characterize the predicted homologue for affinity to DNA-binding sites. The *Zbtb3* was predicted with DBS-PRED [16] and found that their binding affinity was about 34.7%. Solution and solid-state structural studies have revealed that alpha helix of Zinc finger makes contact with the major groove of DNA [17]. The homology modeled zinc finger domain of *Zbtb3* was docked to GAGA-DNA, keeping DNA as a fixed molecule, using Hex. Hex employs spherical polar Fourier correlation and considers both shapes complementarily and with electrostatic effects. The binding energy score obtained was -3.14 kcal/mol and RMS value was about 1.00. From the docked model, it is inferred that it contains favorable binding sites like AGC sequence at N terminus, the presence of highly conserved glycine residues within alpha helix and TGEKP residues at the linker region necessary for promoting zinc finger to fit into the DNA major groove and also to strengthen the DNA binding (**Figure 1b**). This result agrees with the previous findings [18]. *Zbtb3* was also found to be the homolog of *lolla* gene in *Drosophila*. *Lolla* gene functions with Trl for maintaining the repressed state of

target genes and also binds to a DNA polycomb response element (PRE) at the bithroax complex [19]. As it is known that Trl binds with GAGA factor in *Drosophila*, it supports the hypothesis that *Zbtb3* might play a similar function like Trl, binding with GAGA factor.

Conclusion:

Drosophila GAGA factor plays a major role in developmental gene regulation. In this study, through computational analysis it is shown that mouse *Zbtb3* is a remote homolog of *Drosophila* GAGA factor. The reported homolog exhibits strong similarity of highly conserved BTB domain as well as the zinc finger domain to the corresponding regions of *Drosophila* GAGA factor. The homology modeling and molecular docking was applied in this study to explore the binding mechanism of GAGA factor and *Zbtb3*. Through molecular modeling studies, it contains favorable DNA binding sites, which are conserved. These predictions confirm that *Zbtb3* can bind to GAGA repeats.

References:

- [1] Farkas G *et al.* *Nature* 1994 **371**: 806 [PMID:7935842]
- [2] Omichinski JG *et al.* *Nat Struct Biol.* 1997 **4**: 122 [PMID: 9033593]
- [3] Matharu NK *et al.* *J Mol Biol.* 2010 **400**: 434 [PMID: 20471984]
- [4] Schäffer AA *et al.* *Nucleic Acids Res.* 2001 **29**: 2994 [PMID: 11452024]
- [5] Söding J. *Bioinformatics* 2005 **21**: 951 [PMID: 15531603]
- [6] Chenna R *et al.* *Nucleic Acids Res.* 2003 **31**: 3497 [PMID: 12824352]
- [7] Sali A *et al.* *Proteins* 1995 **23**: 318 [PMID: 8710825]
- [8] Bowie JU *et al.* *Science* 1991 **253**: 164 [PMID: 1853201]
- [9] Ramachandran GN *et al.* *J Mol Biol.* 1963 **7**: 95 [PMID: 13990617]
- [10] Ritchie DW *et al.* *Bioinformatics* 2008 **24**: 1865 [PMID: 18591193]
- [11] <http://pfam.sanger.ac.uk/>
- [12] http://www.ebi.ac.uk/Tools/psa/emboss_needle/
- [13] Liu J *et al.* *Biochemistry* 2006 **45**: 6873 [PMID:16734424]
- [14] Vucetic S *et al.* *Bioinformatics* 2005 **21**: 137 [PMID: 15310560]
- [15] Kumar S & Carugo O. *Open Biochem J.* 2008 **2**: 1 [PMID: 18949069]
- [16] Ahmad S *et al.* *Bioinformatics* 2004 **20**: 477 [PMID:14990443]
- [17] Lander ES *et al.* *Nature* 2001 **409**: 860 [PMID:11237011]
- [18] Wolfe SA *et al.* *Annu Rev Biophys Biomol Struct.* 2000 **29**: 183 [PMID:10940247]
- [19] Tracy L *et al.* *Apoptosis* 2009 **14**: 969 [PMID: 19533361]

Edited by P Kanguane

Citation: Kumar. *Bioinformation* 7(1): 29-32 (2011)

License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited.

Supplementary material:

Table 1: Eight probable mouse remote homologue clusters of *Drosophila* GAGA factor. The table shows name of gene with synonyms, size of amino acids (AA) and uniprot ID. The table also shows the pair wise identity of mouse BTB, ZNF and both domains with *Drosophila* GAGA factor by the Needleman-Wunsch alignment algorithm method using EMBOSS programme.

S.No	Name	Size (AA)	Uniprot ID	BTB (%)	ZNF (%)	Both (%)
1	Zbtb3	518	Q91X45	31.1	17	14.4
2	Zbtb32 Rog Tzfp	465	Q9JKD9	27.3	8.1	14.5
3	Zbtb44 Btbd15	453	Q8R0A2	29.7	17	14.3
4	Zbtb7b Thpok	544	Q64321	25	7.4	14.4
5	Zbtb43 Kiaa0414	467	Q9DAI4	33.8	5.9	13.4
6	Zbtb8b	484	Q8CII0	27.1	14.9	13.9
7	Znf238 Rp58 Zfp238	522	Q9WUK6	32.4	9.7	15.7
8	Bcl6b Bazf	474	O88282	29.4	5.2%	14.8