**Tropical Agricultural Research Vol. 19: 52 - 58 (2007)**

# Comparative Structure Modeling of Metal Regulatory Transcriptional Factor; MTF-1

K.H. Rajapaksha, H.A.M. Wickramasinghe[1], J.M.R.S. Bandara[1], N.U. Jayawardana[1]
D.M.A.N. Senavirathne[1] and T.C. Bamunuarachchige[1]

Postgraduate Institute of Agriculture
University of Peradeniya
Peradeniya, Sri Lanka

***ABSTRACT.*** *Metallothionein is a heavy metal binding cystein rich protein, of which is mediated by the binding of heavy metal transcription factor (MTF-1), to its metal response elements (MREs). But its exact signaling pathway and the structures of interacting molecules are not very clear. In order to predict the structure of MTF-1, modeller 9v1 mediated alignment of National Center for Biotechnology Information (NCBI) sequence CAC14279, containing 791 amino acids, with non-redundant (Protein Data Bank) PDB sequences was performed. Since the result was not supportive for the full length modeling of the sequence, best aligning region falling from 111$^{th}$ residue to 284$^{th}$ residue was selected and modeled. The rest was submitted for Protein Specific Iterative Basic Local Alignment Search Tool (PSI-BLAST) and the GeneThreader. Based on the results, 284$^{th}$ residue to 360$^{th}$ residue region was modeled and the two structures were matched by chimera to produce a single structure accounting for 31.47% of the total sequence length of MTF-1. The rest of the sequence accounting for 68.53% requires more information on its folding pattern. Thus experimental approaches such as X-ray crystallography and Nuclear Magnetic Resonance (NMR) need to be employed.*

## INTRODUCTION

Metallothioneins (MTs) constitute a conserved family of cysteine-rich heavy metal binding proteins. MT is found ubiquitously in the tissues of many animal species and plays a role in the homeostasis of essential metals such as zinc and copper (Basu and Lazo, 1990). From insects to mammals, metallothionein genes are induced in response to heavy metals transcription factor (MTF-1), a zinc finger transcription factor, which binds to short DNA sequence motifs, termed metal response elements (MREs) by direct zinc binding to the MTF-1 zinc fingers (Bittel *et al*., 2000; Chen *et al*., 1999). The activation of MTF-1 is mediated by the presence of heavy metals, thus the heavy metal detoxification is controlled to a large extent by the MTF-1. In searching data bases for the sequence of MTF-1, Expressed sequence Tags (EST) sequences CAC14279 derived from adult fly heads and from schneider cells were found and cDNA encodes a protein of 791 amino acids, which is slightly larger than vertebrate MTF-1. It has 78% similarity in the zinc finger region and 27% outside it, *i.e*., 39% in the total protein with the vertebrates (Dalton *et al*., 2000; Maur *et al*., 1999). Experiments have shown that MTF-1 is strongly expressed in the gut of adult *Drosophila* and larvae (Bonneton and Wegnez, 1995) but the exact signaling system and the

[1] Department of Agricultural Biology, Faculty of Agriculture, University of Peradeniya, Peradeniya, SriLanka.

structures of the interacting molecules which is the limiting factor for the predicting of the signaling pathway, remains to be discovered. In this study prediction of the structure of MTF-1 was attempted using evolutionary information.

## MATERIALS AND METHODS

For the prediction of a possible structure for the MTF-1, comparative structure modeling was attempted. Initially the sequence CAC14279, containing 791 amino acids obtained from National Center for Biotechnology Information (NCBI) was used to find a matching profile, using "pdb_95.pir" text format file containing non-redundant Protein Data Bank (PDB) sequences at 95% sequence identity in the database. Due to the template sequence and PDB sequence length mismatch, a multiple sequence alignment between the MTF-1 sequence and PDB sequences was performed using CLC Free Workshop with an open gap penalty of 10 and gap extension penalty of 1. In order to find a suitable known three-dimensional structure homologous to MTF-1, PSI-BLAST (Protein Specific Iterative Basic Local Alignment Search Tool) (Altschul *et al*., 1997) was performed using the engine located at NCBI and gene threading was done using the mGeneThreder (McGuffin *et al*., 2003; Jones, 1999). Considering the aligning region of non-redundant PDB sequences with MTF-1, section ranging from $111^{th}$ residue to $284^{th}$ residue was modeled using the modeler 9V1. Remaining sequences were again aligned with non-redundant PDB sequences followed by a PSI-BLAST and gene threading (Jones, 1999; McGuffin *et al*., 2003). Considering the results another section ranging from $260^{th}$ to $340^{th}$ residue was modeled using the modeler 9V1. Since there were not enough information available for modeling the rest of the sections, completed models were matched using University of California San Francisco (UCSF) Chimera (Pettersen *et al*., 2004; Pei *et al*., 2001) with the Needleman-Wunsch algorithm with a secondary structure score of 30% and a gap extension penalty of 1 to create a single structure.

## RESULTS AND DISCUSSION

NCBI sequence CAC14279, containing 791 amino acids, alignment with "pdb_95.pir" text format file containing non-redundant PDB sequences resulted six PDB structures having the PDB identities as 1g2dC, 1g2fC, 2gliA, 1meyC, 1tf3A, and 1tf6A. (Wolfe *et al*., 2001; Nolte *et al*., 1998; Foster *et al*., 1997; Kim and Burg, 1996; Pavletich *et al*., 1993;) (Figure 1).

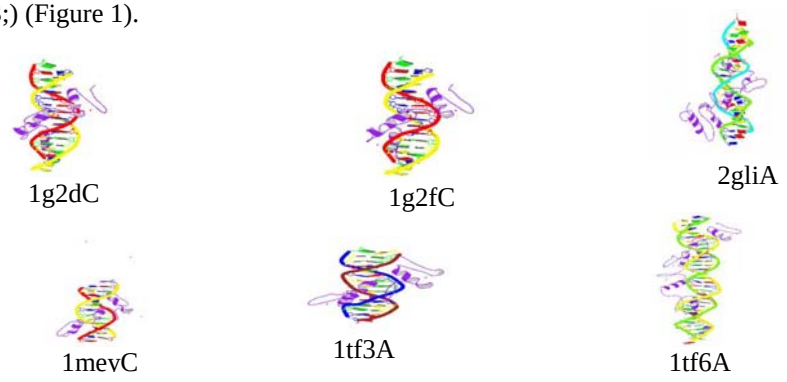


**Figure 1. Caption of evolutionary related PDB (Protein Data Bank) structures.**

Since the sequence length of each of the above models was about 35% of the template length, a multiple sequence alignment with the MTF-1 sequence using CLC Free Workshop, resulted an alignment from 111$^{th}$ residue to 284$^{th}$ residue of the template with the PDB structures (Figure 2).

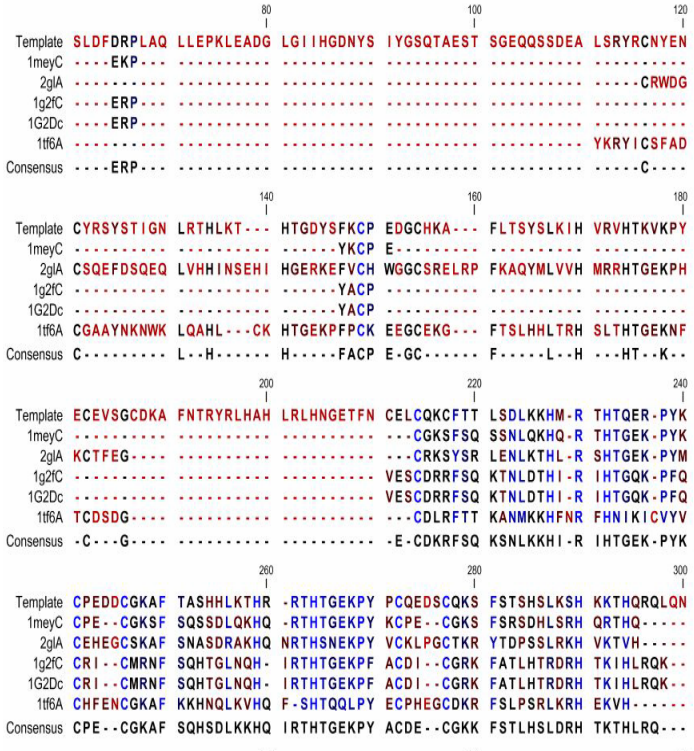


**Figure 2. Multiple alignments between Metal Transcription Factor-1 sequence and non-redundant PDB (Protein Data Bank) sequences.**

PSI-BLAST which was used to find a suitable known 3D structure homologous to MTF-1 produced no suitable hit for the MTF-1 within 500 hits. In search of a better model, the template sequence was submitted to the mGeneThreder (Jones, 1999; McGuffin *et al*., 2003) to find out any homologous folding pattern for the MTF-1 sequence. The sever returned with 13 hits. Out of them only 2GHF, Solution structure of the complete zinc-finger region of human zinc-fingers and homeoboxes 1 (ZHX1) containing 102 residues (Wienk *et al*., 2007) and 1UN6, the crystal structure of a zinc finger- RNA complex containing 61 residues (Lu *et al*., 2003) were selected considering the precision of the search.

Since the sequence length is still around 102 to 190 residues, creating a modeling remains difficult. To overcome the problem of the sequence length, aligning region of the template with the PDB structures which is having 173 residues ranging from 111$^{th}$ residue to 284$^{th}$ residue, was selected. Out of resulted PDB structures, 1tf6A, the co-crystal structure of Xenopus TFIIIA zinc finger domain bound to the 5s ribosomal RNA gene internal control region (having 188 residues) was selected for comparative modeling based on the alignment

and the sequence length. The selected region from 111$^{th}$ to 284$^{th}$ was realigned using the Modeler-9V1 and the template_1tf6A.ali, and the template_1tf6A.pap was created. Those alignments were used in the auto modeling using Modeler-9V1 to produce 30 structures and they were evaluated based on the Discrete Optimized Protein Energy (DOPE) score to select the best structure (Figure 3).

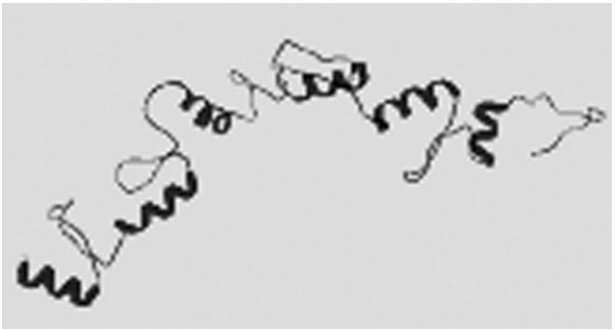

**Figure 3. Chimera rendering of Metal Transcription Factor-1 from 111$^{th}$ residue to 284$^{th}$ residue.**

Remaining MTF-1 templates, each section having residues from 1 to 115 and 280 to 791, was used to find a matching profile within "pdb_95.pir" which ended up with no matching profile. And the same sequences were used for the PSI-BLAST (Altschul *et al*., 1997) also found no suitable protein structure.

Then those sequences were again submitted to the mGeneThreder (Jones, 1999; McGuffin *et al*., 2003) separately. This step resulted no acceptable folding pattern for the segment from 1$^{st}$ to 115$^{th}$ residue and a single molecule having the PDB identity of 2CQF, the Solution Structure of the Zinc-finger domain in LIN-28 (Someya *et al*., 2007) for the segment from 280$^{th}$ to 791$^{st}$ residue, which is also of low quality. In the reanalysis of the mGeneThreder (McGuffin *et al*., 2003; Jones, 1999) resulted, for the submitted sequence having 791 residues, a significant hit having the PDB identity of 1UN6 was detected to have a better alignment with a region of the template ranging from 260$^{th}$ to 340$^{th}$ residue (Figure 4).

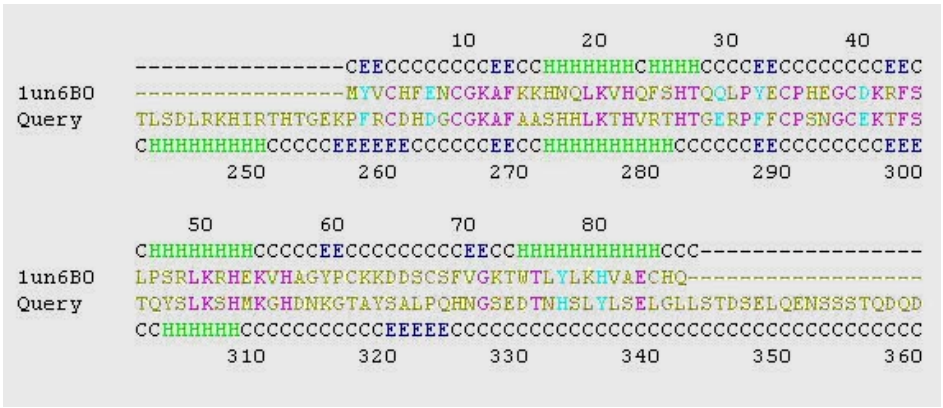


**Figure 4. mGeneThreder result of MTF-1 full length sequence.**

This segment was again selected from the template and aligned with 1UN6 using the Modeler-9V1. The alignment result was used for the construction of 30 models and the evaluation was done based on the DOPE scores. (Figure 5).

**Figure 5. Chimera rendering of MTF-1 from 260$^{th}$ residue to 340$^{th}$ residue**

Comparatively modeled structures ranging from 111$^{th}$ to 284$^{th}$ and 260$^{th}$ to 340$^{th}$ residues were matched by UCSF Chimera (Pettersen *et al*., 2004; Pei *et al*., 2001) using the Needleman-Wunsch algorithm with a secondary structure score of 30% and a gap extension penalty of 1.0 to create a single structure (Figure 6).

**Figure 6. Chimera rendering of MTF-1 from 111$^{th}$ residue to 340$^{th}$ residue.**

During evolution, variations are created due to recombination and mutation which are commonly observed in nature. Those are accounted for homologous regions of a protein. BLAST and alignment programs are designed to identify such regions with 95% precision. By combining those structurally and functionally related domains, homology modeling of proteins is conducted. MTF-1 had its evolutionary relationship with 7 proteins. They had their homologies at different locations depending on the evolutionary background. Homology search with PSI-BLAST and genethreading result those domains as an alignment which later was used for threading the template sequence along those known aligning domains of PDB structures.

## CONCLUSIONS

Comparative structure modeling could model only the region ranging from 111$^{th}$ residue to 340$^{th}$ residue out of 791 residues of MTF-1. The rest of the structure identification requires an experimental technique such as X-ray crystallography and NMR (Nuclear magnetic resonance).

## ACKNOWLEDGEMENT

Authors wish to thank the organizers of the 6th European school of bioinformatics especially Dr. Mark Martie Renome for the guidance provided to us.

## REFERENCES

Altschul, Stephen, F., Thomas, L., Madden, Alejandro, A., Schäffer, Jinghui, Z., Zheng, Z., Webb, M. and David, J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, Nucleic Acids Res. 25: 3389 - 3402.

Basu, A. and Lazo, J.S. (1990). A hypothesis regarding the protective role of metallothioneins against the toxicity of DNA interactive anticancer drugs. Toxicol. Lett. 50: 123 - 135.

Bittel, D.C., Smirnova, I.V. and. Andrews, G.K. (2000). Functional heterogeneity in the zinc fingers of the metalloregulatory transcription factor, MTF-1. J. Biol. Chem. 275: 37194 - 37201.

Bonneton, F. and Wegnez, M. (1995). Developmental variability of metallothionein *Mtn* gene expression in the species of the *Drosophila melanogaster* subgroup. Dev. Genet. 16: 253 - 263.

Chen, X., Chu, M. and Giedroc, D. P. (1999). MRE-binding transcription factor-1: weak zinc-binding finger domains 5 and 6 modulate the structure, affinity, and specificity of the metal-response element complex. Biochemistry. 38: 12915 - 12925.

Dalton, T. P., Solis, W. A., Nebert, D. W. and Carvan, M. J. (2000). Characterization of the MTF-1 transcription factor from zebrafish and trout cells. Biochem. Physiol. 126: 325 - 335.

Foster, M.P., Wuttke, D.S., Radhakrishnan, I., Case, D.A., Gottesfeld, J.M., Wright, P.E. (1997). Domain packing and dynamics in the DNA complex of the N-terminal zinc fingers of TFIIIA. Nature Structural Biology.
URL: http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1TF3 (Accessed on 25th May 2007).

Jones, D.T. (1999). GenTHREADER: an efficient and reliable protein fold recognition method for genomic sequences. J. Mol. Biol. 287: 797 - 815.

Kim, C.A., Berg, J.M. (1996). Resolution crystal structure of a designed Zinc finger protein bound to DNA, Nature Structural Biology. 3: 940 - 945.

Lu, D., Searles, M.A., Klug, A. (2003). Crystal structure of a Zinc-finger-RNA complex reveals two modes of molecular recognition. Nature. 426: 96 - 100.

McGuffin, L.J. and Jones, D.T. (2003). Improvement of the genthreader method for genomic fold recognition. Bioinformatics, 19: 874 - 881.

Maur, A., Belser, T., Elgar, G., Georgiev, O. and Schaffner, W. (1999). Characterization of the transcription factor MTF-1 from the Japanese pufferfish (*Fugu rubripes*) reveals evolutionary conservation of heavy metal stress response. Biol. Chem. 380: 175 - 185.

Nolte, R.T., Conlin, R.M., Harrison, S.C., Brown, R.S. (1998). Differing roles for zinc fingers in DNA recognition: structure of a six-finger transcription factor IIIA complex. Proc. Natl.Acad.Sci.USA (Accessed on 25th May 2007).
URL: http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=1TF6

Pavletich, N.P. and Pabo, C.O. (1993). Crystal structure of a five-finger GLI-DNA complex: new perspectives on zinc fingers. Science
URL: http://www.rcsb.org/pdb/explore/pubmed.do?structureId=2GLI (Accessed on 25th May 2007).

Pei, J. and Grishin, N.V. (2001). AL2CO: Calculation of positional conservation in a protein sequence alignment. Bioinformatics. 17 (8): 700 - 712.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera - A visualization system for exploratory research and analysis J. Comput. Chem. 25 (13): 1605 - 1612.

Someya, T., Muto, Y., Inoue, M., Kigawa, T., Terada, T., Shirouzu, M. and Yokoyama, (2007). Solution structure of the Zinc-finger domain in LIN - 28 2007.
URL:http://www.rcsb.org/pdb/static.do?p=education_discussion/molecule_of_the_month/pdb87_report.html (Accessed on 25th May, 2007).

Wienk, H., Lammers, I., Wu, J., Hotze, A., Wechselberger, R., Kaptein, R., Folkers, G. (2007). Solution structure of the zinc-finger region of human zinc-fingers and homeoboxes 1 (ZHX1).
URL:http://www.rcsb.org/pdb/static.do?p=education_discussion/molecule_of_the_month/pdb87_report.html (Accessed on 25th May, 2007).

Wolfe, S.A., Grant, R.A., Elrod-Erickson, M. and Pabo, C.O. (2001). Beyond the recognition code: structures of two cys2His2 zinc finger/TATA box complexes. (Accessed on 25th May 2007).
URL: http://www.rcsb.org/pdb/explore/pubmed.do?structureId=1G2D