IJPSR (2014), Vol. 5, Issue 3 **(Research Article)**

INTERNATIONAL JOURNAL OF PHARMACEUTICAL SCIENCES AND RESEARCH

Received on 24 October, 2013; received in revised form, 22 November, 2013; accepted, 13 February, 2014; published 01 March, 2014

HOMOLOGY MODELING AND ANALYSIS OF STRUCTURE PREDICTIONS OF KLF8 PROTEIN FROM *HOMO SAPIENS*

Vinod Kumar Jatav*, Lovekesh Parasher, Ankit Raje and Samraddhi Tiwari

Department of Biotechnology, Madhav Institute of Technology and Science, Gwalior, Madhya Pradesh, India

Keywords:

KLF8, NCBI, BLAST, Homology modeling, I- TASSER, SWISS-MODEL

Correspondence to Author:

Vinod Kumar Jatav

Assistant Professor, Department of Biotechnology, Madhav Institute of Technology and Science, Gwalior, Madhya Pradesh, India

E-mail: vinod.mits@gmail.com

ABSTRACT: KLF proteins are highly conserved among mammals from human to mouse, with many KLFs also having homologs in *Gallus gallus* (Chicken), *Danio rerio* (Zebra fish), and *Xenopus laevis* (Frog). On the basis of functional characteristics, KLF proteins can be divided into three distinct groups. KLFs in group 1 (KLFs 3, 8, and 12) serve as transcriptional repressors through their interaction with the carboxy-terminal binding protein (CtBP). Family members in group 2 (KLFs 1, 2, 4, 5, 6, and 7) function predominantly as transcriptional activators. KLFs in group 3 (KLFs 9, 10, 11, 13, 14, and 16) have repressor activity through their interaction with the common transcriptional corepressor Sin3A. The encoded protein is thought to play an important role in the regulation of epithelial to mesenchymal transition, a process which occurs normally during development but also during metastasis. A homology modelling method was used for the prediction of the structure and other various physico-chemical properties of protein were obtained using ProtParam. 3D structure was constructed for the target protein using I- TASSER and SWISS-MODEL. The comparison between the structures generated from above mentioned tools indicates greater acceptability of the structure generated from I-TASSER and validation of structure was performed by using Ramachandran plot. The present study gave an outlook on KLF8 protein and further research to be carried out in preventing the cancer in human.

INTRODUCTION: The Kruppel type zinc finger transcription factors comprise a conserved family of DNA binding proteins that are important in developmental regulation. The Kruppel zinc finger transcription factor was initially identified in *Drosophila* as a segmentation gene. Kruppel like factor 8 (KLF8), also called basic Kruppel like factor 3 and zinc finger protein 741, is a 359 amino acid transcriptional repressor that binds CACCC elements in DNA and activates or represses their target genes in a context dependent manner [1].

QUICK RESPONSE CODE

DOI: 10.13040/IJPSR.0975-8232.5(3).1045-50

Article can be accessed online on: www.ijpsr.com

DOI link: http://dx.doi.org/10.13040/IJPSR.0975-8232.5(3).1045-50

Kruppel like factor 8 (KLF8) is a GT-box (CACCC) binding dual-transcription factor that has a critical role in the regulation of cell cycle progression, transformation EMT and invasion [2, 10].

KLF8 expression is promoted by Src and phosphatidylinositol 3 kinase signaling[3] and its transcription is activated by Sp1 or inhibited by KLF3[4]. KLF8 is also regulated by posttranslational sumoylation and localization [5]. FAK is a critical regulator of both EMT and MMPs during breast cancer invasion and metastasis [6]. Like FAK, KLF8 is highly overexpressed in invasive human cancers including breast cancer [7] and promotes breast cancer cell invasion.

A novel role of KLF8 in regulating oncogenic transformation.

Our finding that KLF8 is overexpressed in human cancer combined with considerable evidence that link elevated expression and/or activity of Src, FAK and cyclin D1 with the development of human cancer [8] strongly suggests a potentially significant role for KLF8 in tumor progression of human cancer.

Recent investigation revealed that KLF8 induces tumor cell epithelial to mesenchymal transition (EMT) and maintains the invasive potential of cancer, which seemingly plays a crucial role in metastatic progression of human carcinoma [9]. KLF8 protein and mRNA expression was not only significantly higher in renal cell carcinoma (RCC) than in non-neoplastic renal tissue, inhibition of KLF8 via siRNA could also induce cell apoptosis in vitro and reduced tumor growth *in vivo* [10]**.** KLF8 overexpression was demonstrated in highly metastatic or recurrent hepatocellular carcinomas (HCC).

In this study, KLF8 could even be linked with clinical data and revealed to be a significant predictor for overall and progression-free survival[11]. Only recently additional data was presented, which identified KLF8 as an activator for matrix metalloproteinase 9 (MMP9), leading to enhanced tumor invasion of human breast cancer [12].

Knowing the three-dimensional structure of a protein can give invaluable information about its functional and evolutionary features and in addition, the structural knowledge is useful in drug design efforts. Homology modeling makes structure predictions based primarily on its sequence similarity to one or more proteins of known structures.

Comparative modeling comprise of mainly five stages.

(a) Identifying evolutionary related sequences of already known structure.

(b) Aligning of the query sequence to the template structures.

(c) Model building of structurally conserved regions using known templates.

(d) Modeling side chains and loops which are different than the templates.

(e) Refinement and evaluation of the quality of the model through conformational sampling [13].

MATERIALS AND METHODOLOGY:

Retrieval of target sequence: The FASTA sequence of the KLF8 protein from Homo sapiens was retrieved from the NCBI database [http://www.ncbi.nlm.nih.gov] that has 359 amino acids (gi-85567555, gb- AAI12110.1). It was found that the three-dimensional structure of the protein was not available in Protein Data Bank.

Sequence analysis: Protein sequence was compared for detecting homologous sequences found in databases using Basic Local Alignment Search Tool-Protein (BLAST-P) [14, 15]. Using the sequence, the physicochemical properties of the protein were calculated with the ProtParam [16].

Sequence alignment: Protein sequences alignments of the target were obtained in FASTA [17] format and Clustal W [18] was used for the multiple sequence alignment and tree was obtained. Default parameters were applied and the aligned sequences were inspected for number of gaps and insertions.

Secondary structure prediction: Secondary structure has been predicted using SOPMA and PHYRE 2 [19] software where the FASTA format of the sequence was given as input. It provides the structural information of the protein sequence in form of coils, helices and strands with template information.

Homology modelling and structure refinement: 3-D structure prediction was done by using online tools SWISS-MODEL and I- TASSER .The modeling involves four basic steps, first searching structure showing homology with target, then selecting a best template having maximum identity with the target sequence which follows its alignment with the target and modeling the structure using I- TASSER. The modelled structure is then evaluated using PROCHECK [20].

RESULTS & DISCUSSION: The similarity search for the sequence was carried out with the help of BLAST-P tool. The BLAST- P analysis gives the alignment with 22 protein sequences.

All alignments of proteins were chosen and tabulated above (similarity ranging between 90% to 100%). AAI12110.1 aligned with NP_009181.2, AAH31355.1, CCO02795.1 and AFE71444.1 each are (100%) identity.

Phylogenetic analysis of KLF8 protein:

Phylogenetic Tree:

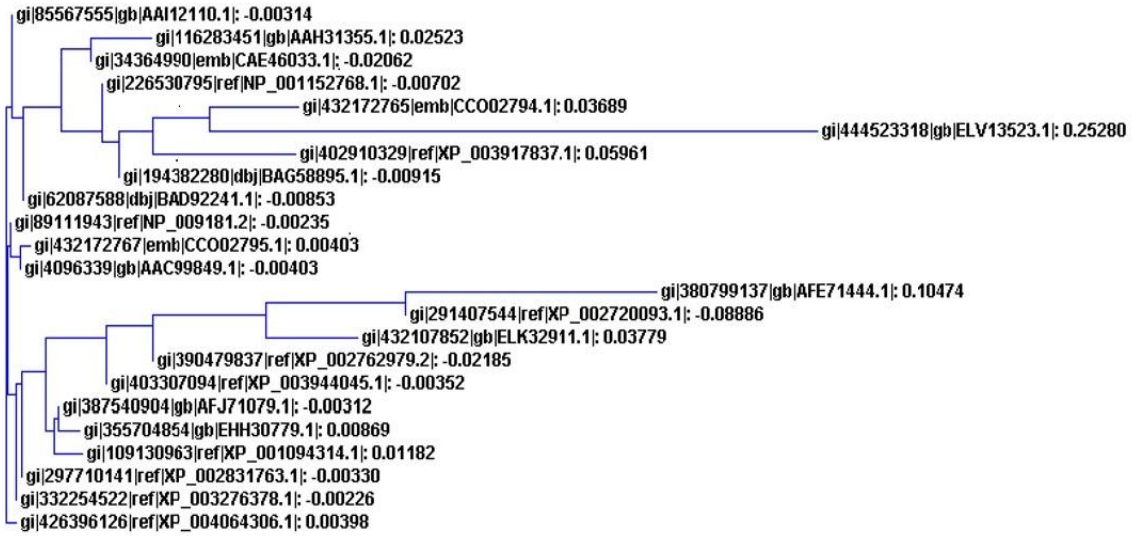

FIGURE 1: SHOWS THE PHYLOGENETIC RELATIONSHIP AMONG THE KLF8 PROTEINS

A multiple sequence alignment of the chosen proteins from the BLAST- P result was performed with the help of Clustal W. A phylogenetic tree of the alignment of KLF8 protein was also generated. The Phylogram showed that the proteins with accession no. AAH31355.1, CAE46033.1, NP 001152768.1, CCO02794.1, ELV13523.1, XP 003917837.1 and BAG58895.1 show evolutionary relationship with **AAI12110.1** they share common ancestor but has deviated due to evolutionary changes (**Figure 1**)

Important physiochemical properties of polyprotein determined using ProtParam: Table 1 represents Physiochemical Parameters calculations using Expasy's ProtParam tool. A protein with instability index smaller than 40 is predicted to be stable, whether a value above 40 predicts that the protein may be unstable[21]. Instability index of 69.47 indicates the unstable nature of protein. Low extinction coefficient (17335) shows the presence of Cys, Trp and Tyr residues in low concentration. The aliphatic index is assumed as a positive factor for the increase of thermal stability. High aliphatic index (74.40) of query protein suggests that the protein may be stable for a wide temperature range. The Grand Average hydropathy (GRAVY) value is low (-0.469) and indicates the possibility of interaction with water.

TABLE 1: SHOWS THE PHYSICOCHEMICAL PROPERTIES OF PROTEIN.

S. No.	Property	Value
1	Number of amino acids	359
2	Molecular weight	39313.6
3	Theoretical Pi	7.19
4	Total number of negatively charged residues (Asp + Glu)	37
5	Total number of positively charged residues (Arg + Lys)	37
6	Extinction coefficient	17335
7	Extinction coefficient*	16960
8	Instability index	69.47
9	Aliphatic index	74.40
10	Grand average of hydropathicity (GRAVY)	-0.469

SOPMA results: The prediction of secondary structure of KLF8 protein from *Homo sapiens* was performed by two software namely SOPMA (Self Optimized Prediction Method with Alignment) and phyre 2. The accuracy of SOPMA to predict amino acids correctly is considered to be 69.5% to describe secondary structure prediction. The results of SOPMA are presented in **Table 2**. These results indicate greater number of random coils in comparison to other secondary structure elements (alpha helix, extended strand and beta turns). Secondary structure prediction by SOPMA was done by taking default parameters default parameters (Window width: 17, similarity threshold: 8 and number of states: 4). Secondary structure and disorder prediction was performed using phyre 2.

TABLE 2: SECONDARY STRUCTURE COMPOSITION WITH DIFFERENT COLORS

Alpha helix	(Hh)	**53 is 14.76%**
3_{10} helix	(Gg)	**0 is 0.00%**
Pi helix	(Ii)	**0 is 0.00%**
Beta bridge	(Bb)	**0 is 0.00%**
Extended strand	(Ee)	**67 is 18.66%**
Beta turn	(Tt)	**20 is 5.57%**
Bend region	(Ss)	**0 is 0.00%**
Random coil	(Cc)	**219 is 61.00%**
Ambiguous states	(?)	**0 is 0.00%**
Other states		**0 is 0.00%**

SOPMA analyzes the stability of AAI12110.1 on the basis of no. of coil (219), helix (53), and strand (67).

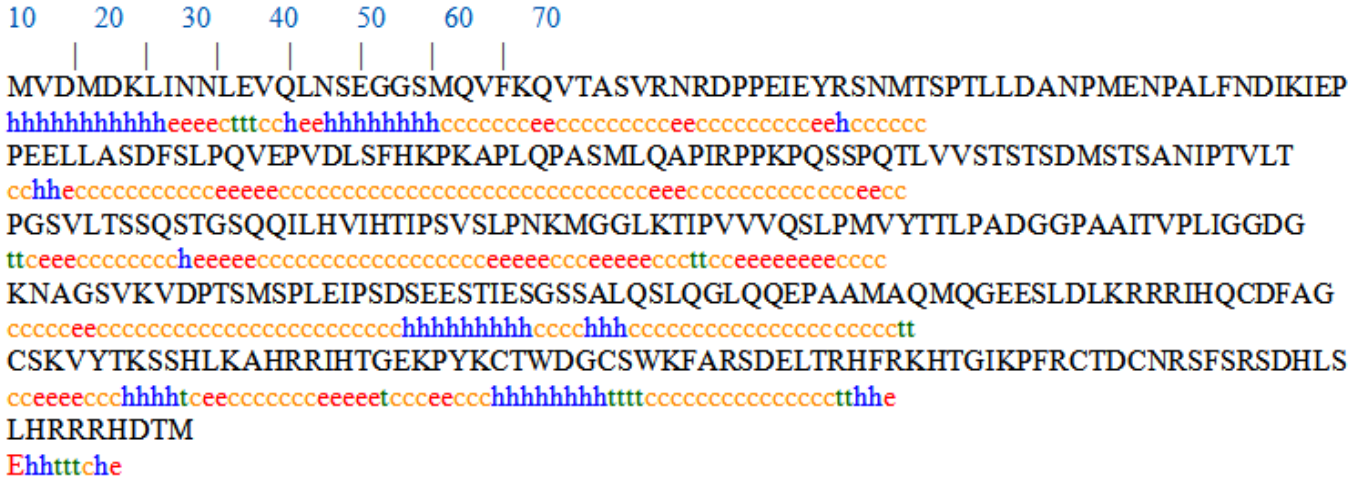

FIGURE 2: SHOWS THE SECONDARY STRUCTURE OF KLF8 PROTEIN

I- TASSER result:

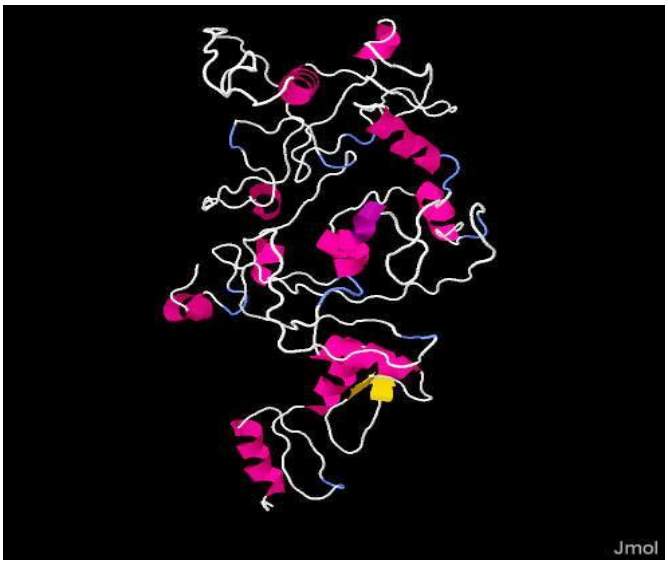

FIGURE 3: SHOWS THE PREDICTED STRUCTURE OF KLF8 PROTEIN

Model shows the structure of protein (KLF8).following color shows below

- Helices: magenta
- Sheets: yellow
- Turns: pale blue
- All other residues: white

There is no experimental structure available for the protein considered. KLF8 Protein from *Homo sapiens* (human) was structured via two homology modeling programs namely SWISS-MODEL and I-TASSER. The ϕ and ψ of Ramachandran plot produced by of non-glycine, non proline residues were summarized in **table 3**.

TABLE 3: RAMACHANDRAN PLOT CALCULATION AND COMPARATIVE ANALYSIS OF THE MODELS FROM SWISS-MODEL AND I- TASSER COMPUTED WITH THE PROCHECK PROGRAM

Server	Parameters	Value (%)
SWISS-MODEL	Total number of residues	**56**
	Residues in the most Favoured Region	**94.2**
	Residues in additionally allowed region	**4.0**
	Residues in generously allowed region	**2.0**
	Residues in disallowed region	**0.0**
I- TASSER	Total number of residues	**359**
	Residues in the most Favoured Region	**79.4**
	Residues in additionally allowed region	**14.7**
	Residues in generously allowed region	**3.9**
	Residues in disallowed region	**2.0**

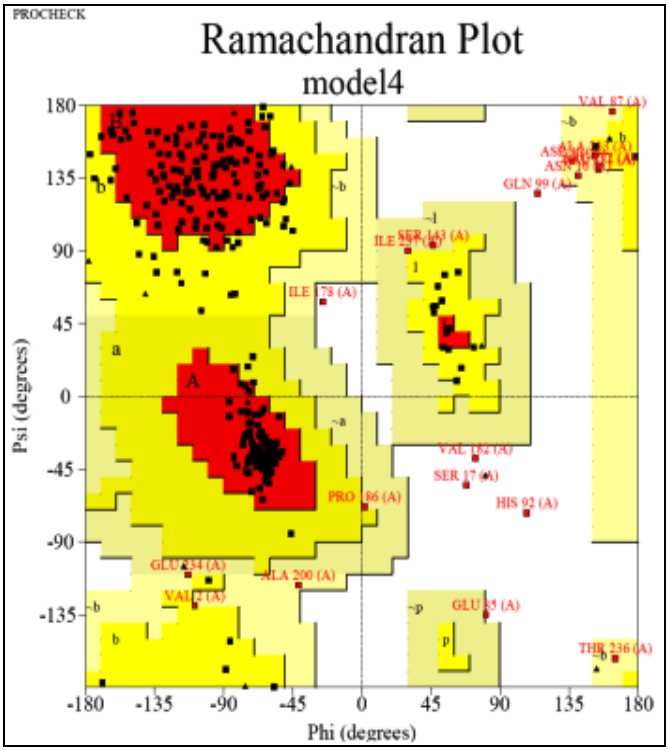

FIGURE 4: RAMACHANDRAN PLOT OF THE PREDICTED STRUCTURE BY I- TASSER

On the basis of the results from two different programs SWISS-MODEL and I- TASSER compared in table 3, it was observed that the model obtained from I- TASSER was more satisfactory as compared to that by SWISS-MODEL. The final model was seen in J-MOL and is shown in figure 3.

To find out the accuracy of predicted models using Ramachandran plot calculations and to check the stereo chemical quality PROCHECK program was used. The results obtained from Ramachandran plot shows (**fig. 4**) that the Residues in most favored regions [A, B, L] are 79.4%, Residues in additional allowed regions [a, b, l, p] are 14.7% and Residues in generously allowed regions [~a, ~b, ~l, ~p] are 3.9%, Residues in disallowed regions 2.0%.

The distribution of the main chain bond lengths and bond angles was within the limits for these proteins. These figures suggest a good quality of the predicted model.

CONCLUSION: In the modeling procedure of KLF8 protein from *Homo sapiens* (human), two online tools were used and the predicted model verification program PROCHECK was found that the model generated from I- TASSER was well conformed to the stereochemistry suggesting reasonably good quality. Computational methods help to predict protein structure more quickly and economically. Present study suggests that the predicted structure is ready to be verified in vitro and will help researchers for further experimentation in disease prevention.

ACKNOWLEDGEMENTS: The authors are grateful Department of Biotechnology, MITS, Gwalior, providing us such kind of opportunity to work.

REFERENCES:

1. Santa Cruz biotechnology, inc. (http://www.scbt.com /datasheet-69294-klf8-d-16-antibody.html).
2. Zhao J, Bian ZC, Yee K, Chen BP, Chien S, Guan JL: Identification of transcription factor KLF8 as a downstream target of focal adhesion kinase in its regulation of cyclin D1 and cell cycle progression. Mol Cell 2003; 1503–1515.
3. Wang X, Urvalek AM, Liu J and Zhao J: Activation of KLF8 transcription by FAK in human ovarian epithelial and cancer cells. J Biol Chem 2008; 283: 13934-13942.
4. Eaton SA, Funnell AP, Sue N, Nicholas H, Pearson RC: A network of Kruppel-like Factors (Klfs). Klf8 is repressed by Klf3 and activated by Klf1 in vivo. J Biol Chem 2008; 283: 26937–26947.
5. Mehta TS, Lu H, Wang X, Urvalek AM, Nguyen KH, Monzur F, Hammond JD, Ma JQ and Zhao J: A unique sequence in the N-terminal regulatory region controls the nuclear localization of KLF8 by cooperating with the C-terminal zinc-fingers. Cell Res 2009; 19: 1098-1109.

6. Zhao J. & Guan JL: Signal transduction by focal adhesion kinase in cancer. Cancer Metastasis Rev 2009; 28: 35-49.
7. Wang X, Zhao J: KLF8 transformation factor participation in oncogenic transformation. Oncogene 2007; 36: 456-461.
8. Gabarra-Niecko V, Schaller MD, Dunty JM: FAK regulates biological processes important for the pathogenesis of cancer. Cancer Metastasis Rev.2003; 22:359-374.
9. Wang X, Zheng M, Liu G: Kruppel-like factor 8 induces epithelial to mesenchymal transition and epithelial cell invasion. Cancer Res 2007; 67:7184–7193.
10. Fu WJ, Li JC, Wu XY, Yang ZB, Mo ZN: Small interference RNA targeting Kruppel-like factor 8 inhibits the renal carcinoma 786-0 cells growth in vitro and in vivo. J Cancer Res Clin Oncol 2010; 1255–1265.
11. Li JC, Yang XR, Sun HX, Xu Y, and Zhou J: Up-regulation of Kruppel-like factor 8 promotes tumor invasion and indicates poor prognosis for hepatocellular carcinoma. Gastroenterology 2010; 139: 2146–2157.
12. Wang X, Lu H, Urvalek AM, Li T, Yu L, et al. KLF8 promotes human breast cancer cell invasion and metastasis by transcriptional activation of MMP9. Oncogene 2011; 30:1901–1911.
13. Abhilash M, Nandhini K: International Journal of Pharma Sciences and Research 2010; Vol.1 (1):40-50.
14. Altschul S, Madden T, Schaffer A, Zhang J,Zhang Z, Miller W, Lipman D Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 1997; 25, 3389–3402.
15. Gasteiger E: Protein Identification and Analysis Tools on the ExPASy Server. In: John M. Walker ed, The Proteomics Protocols Handbook, Humana Press.2005; 571-607.
16. Berman HM, Battistuz T, Bhat TN, Bluhm WF, Bourne PE, Burkhardt K, Feng Z, Gilliland GL, Iype L, Jain S: The Protein Data Bank. Acta Crystallogr. D. Biol. Crystallogr. 2002; 58, 899–907.
17. Geourjon C, Deléage G SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiplealignments. Comput Appl Biosci. (1995)11: 681-684.
18. Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A.;Protein Identification and Analysis Tools on the ExPASy Server;(In) John M. Walker (ed): The Proteomics Protocols Handbook, Humana Press 2005; 571-607.
19. Lawrence Kelley, Benjamin Jefferys. Phyre2: Protein Homology/analogY Recognition Engine V 2.0. Structural Bioinformatics Group, Imperial College, London. Retrieved 22 April 2011
20. Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. J Biomol NMR. (1996) 8: 477-486.
21. Guruprasad K, Reddy BVP, Pandit MW: Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. Prot Eng 1990; 4: 155-164.

How to cite this article:
Jatav VK, Parasher L, Singh A and Tiwari S: Homology modeling and analysis of structure predictions of KLF8 protein from *Homo sapiens*. *Int J Pharm Sci Res* 2014; 5(3): 1045-50.doi: 10.13040/IJPSR.0975-8232.5(3).1045-50