International Journal of Systems Biology, ISSN: 0975–2900, Volume 2, Issue 1, 2010, pp-12-19

Conformation analysis of homology model of human merlin

Sivakumar K.C. and Vidhya Ramaswamy
International Centre for intellectual Training and Empowerment (INCITE)
Sri Rams, N.H.Road, Kazhakuttom, Trivandrum, Kerala, India, 695582, vidhya.ramaswamy@gmail.com

Abstract - Merlin shares sequence similarity with the 4.1 super family of proteins (ezrin, radixin, and moesin) that link cell surface glycoproteins to the actin cytoskeleton. We modeled the structure of human merlin using the structure of moesin from *Spodoptera frugiperda* as the template. The present model of merlin structure suggests an interaction of its extreme C- terminal region with the subdomains B and C of FERM domain, masking the binding site of beta II spectrin. Our model suggests that FERM domain is masked in a closed conformation of merlin preventing the interaction of other proteins with it. Modeling the complete structure of merlin revealed a novel central alpha helical domain with a helix-coil-helix. The actin binding site in the carboxy terminal is absent in merlin. For merlin (closed conformation), the indirect actin binding site in the FERM domain is also not available for interaction with other proteins.
Keywords- Ezrin, moesin and radixin proteins, merlin, neurofibromatosis, betaII spectrin, modelling, tumour suppressor

1. Introduction

Neurofibromatosis (NF2) is an autosomal dominant inherited tumor predisposition syndrome and affected individuals are prone to the development of neuronal tumors [1]. The Neurofibromatosis 2 gene, identified by positional cloning in 1993, encodes a protein termed merlin or schwannomin [2,3]. Merlin contains three predicted structurally important regions, including an amino terminal FERM domain (residues 1–302), a central alpha helical region (residues 303–479) and a unique carboxyl terminus (residues 480–595) [4,5]. These predictions are based on its similarity to the 4.1 super family of proteins (ezrin, radixin, and moesin) since merlin shares 45% overall sequence identity with them and 65% identity over the first 300 residues [6]. Ezrin, radixin, and moesin proteins function as membrane-organizing proteins, linking the plasma membrane to the cytoskeleton [7,8]. Over expression of merlin results in suppression of cell growth whereas over expression of ezrin is correlated with cell proliferation, despite their sequence similarity and ability to associate with each other [9]. *Nf2* knock out mice are embryonic lethal as they do not develop proper extra embryonic tissues and fail to implant and *Nf2* +/– heterozygous mice develop a wide range of metastatic tumors [10]. Loss of merlin in *Drosophila melanogaster* results in over proliferation of cells [11]. Merlin interacts with a number of effectors, including a sodium-hydrogen exchange regulatory factor, an actin binding protein (beta II spectrin), schwannomin interacting protein (syntenin), the CD44 transmembrane hyaluronic acid binding protein, and hepatocyte growth factor-regulated tyrosine kinase substrate (HRS/HGS) [2,12]. Like the 4.1 super family of proteins, merlin exists in two conformations (open and closed), but in contrast to them, merlin is active (growth suppressive) in its closed conformation [13]. Under cell growth permissive conditions, merlin is phosphorylated by RAC1-dependent p21 activated kinase resulting in an open inactive form [14-16]. Although the crystal structure of FERM domain of merlin is available, the complete structure has not been modeled yet. The non-availability of actin binding site in the carboxy terminal is evident from the available sequence of merlin. We modeled the complete structure of merlin and found a novel central alpha helical domain with a helix- coil-helix. We also found that the indirect actin binding site in the FERM domain of merlin is not available for interaction with other proteins in a closed conformation.

2. Materials and methods

2.1 Similarity search

A homology search for merlin was carried out with BLAST [17] and PSI-BLAST [18] against NCBI nonredundant pdb for similar crystallographic structures. Default parameters were used for BLAST and PSI-BLAST except that, for the latter, the threshold for including homologous protein in the profile was set to 0.05 and the number of iterations was set to 10. The choice of templates was restricted to X-ray crystallographic structures of Moesin from *Spodoptera frugiperda* for which PDB coordinates; 2I1K was available at Brookhaven Protein Data Bank (PDB; www.rcsb.org/pdb/). A domain search was performed against PFAM database [19] with default parameters. The amino acid sequences of ezrin, radixin, moesin and merlin were obtained from Swiss-Prot/TrEMBL database of the ExPASy Molecular biology server (www.expasy.org/sprot/) with the accession numbers P15311, P35241 and P35240, respectively. A multiple sequence alignment of COOH terminal domain of ezrin, radixin, moesin and merlin was done using CLUSTALW (http://align.genome.jp) with matrix BLOSUM62.

2.2 Homology modelling and optimization

The program MODELLER [20] is used to generate initial homology models in the first step of our approach. MODELLER generates protein 3D structures by satisfying spatial restraints

Copyright © 2010, Bioinfo Publications, International Journal of Systems Biology, ISSN: 0975–2900, Volume 2, Issue 1, 2010

imposed by the sequence alignment with the template structure and applying the terms of the CHARMM-22 force-field. 3D protein model is obtained by optimising the molecular probability density function while simultaneously minimising input restraint violations. Preliminary tests showed that a number between 10 and 100 models provides a satisfactory sampling. To optimise the local interactions, all models obtained are subjected to a crude simulated annealing refinement protocol available in MODELLER. Ten models were produced by MODELLER using the standard model- building routine. A single model was automatically selected based on the average between the minimal energy as calculated by MODELLER and minimal steric violations.

2.3 Secondary structural analysis of binding residues of HRS, beta II spectrin and merlin

Secondary structural analysis of the binding residues of HRS, beta II spectrin and merlin proteins was done using the server APSSP2 [21], to identify the types of interaction of both HRS and beta II spectrin with merlin.

Molecular Visualization was done using PyMol (DeLano Scientific Palo Alto, CA, USA. http://www.pymol.org).

3. Results and Discussion

3.1 Merlin lacks the COOH terminal actin binding motif

Multiple sequence alignment of ezrin, moesin and radixin proteins indicates that the C- terminal parts of the three human proteins have a high pattern of sequence conservation that includes a conserved actin binding motif “KYKXL” (Figure 1a). This motif is highly conserved in myosin heavy chains, brush border myosin I (a microvillus protein), CapZbeta subunit and the ezrin protein family [22]. The COOH terminal (25 amino acids) in ezrin, moesin and radixin proteins is the most conserved (Figure 1a). Homology is near the region of the conserved tyrosine (marked Y in Figure 1a) that may be indicative of a common actin binding site. However, Figure 1b shows that the COOH- terminal sequence of merlin neither shows any remarkable homology to the actin binding site of ezrin, moesin and radixin proteins nor conserves the tyrosine. Further, the lack of conservation of actin binding motif in the C-terminal region of merlin in *Fugu rubripes, Danio rerio, C. briggsae*, and *Brugia malayi* may imply that this region does not share the same function [23]. The lack of conserved “KYKXL” motif supports the idea that actin binding motif is absent in the C-terminal region of human merlin.

3.2 Complete model of human merlin predicts a novel central alpha helical domain with a helix-coil-helix

The amino acid sequence alignment of moesin and merlin is shown in Figure 2. The percentage identity and similarity (identical plus conservative substitutions) for the sequence alignment were 48 and 64%, respectively. Analysis of the predicted structure of merlin reveals three domains: a FERM domain (residues 1-312), an alpha helical region (residues 313-506), and a unique C-terminal domain (residues 506-595). FERM domain of merlin reveals three well-defined sub domains, A, B and C [4,24]. The close packing suggests a coordinated structure and these domains presumably do not function independently. Figure 3 represents the three dimensional structure of the human merlin model that contains all the three predicted domains. The model represents the closed conformation of the human merlin protein since the C-terminal is folded to interact with the sub domains B and C of the FERM domain. Ramachandran plot of the model was evaluated using PROCHECK [25], (Figure 4). SCHIP-1 (Schwannomin interacting protein 1), NHERF (Na+-H+ exchanger regulatory factor), CD44, syntenin, paxillin and HRS, known to interact with FERM domain and C-terminal domain, [26] may not be able to bind with merlin in closed conformation. The residues in FERM domain masked within 4A˚ from the C-terminal are shown in Table 1 and it is interesting to note that no interaction is seen with sub domain A of FERM domain. Therefore our human merlin model predicts that in its closed conformation, C-terminal may interact with sub domains B and C of FERM domain (Figure 5) that retained the same secondary structural features of the original FERM domain modeled earlier [27]. Four distinct alpha helices are found: one interacting with B sub domain region; two with C sub domain; one in the interface of sub domains B and C. The presence of these alpha helices is indicative of the fact that these regions may be involved in the coiled coil type interaction with major interacting proteins like HRS and beta II spectrin. Residue masking was found to be more pronounced in the C sub domain of the FERM domain (Table 1). Our results are in agreement with the predictions made with the crystal structure of a dormant moesin FERM/tail complex revealing that its FERM domain has three compact lobes with the C-terminal segment bound as an extended peptide masking a large surface of the FERM domain [28].

3.3 Secondary structure analysis reveals common binding sites for HRS and beta II spectrin at the C-terminus of merlin

We have done secondary structural analysis [21] of the residues involved in the interaction between HRS, beta II spectrin and merlin C-terminal residue. The results show that the residues of the interacting regions of two proteins (merlin and HRS or merlin and spectrin) adopt a helix-coil (coiled-coil) structure (additional files 1, 2 and 3). It has been shown that merlin binds with beta II spectrin or fodrin through its C terminal region and the interacting region of fodrin includes residues 982-1188, which encompass spectrin repeat 8 [29]. HRS, required

for tumor suppression mediated by merlin, contains 777 amino acids with a “FYVE” domain, a VHS zinc finger domain, a coiled coil domain and two proline-rich regions [4, 30]. Sun et al. [4] generated a series of HRS truncation mutants and defined that the residues 470-497 (which form the predicted coiled-coil domain) are required for merlin binding . The HRS binding domain of merlin was mapped to residues 453-557 [31]. Some of the residues of merlin C-terminus may be common for both HRS and beta II spectrin for interacting with merlin. It has been shown that merlin with an L46R, L360P, L535P, or Q538P missense mutation demonstrated reduced affinity for HRS binding and L360P, K413E, and L535P missense mutations demonstrated reduced affinity for beta II spectrin binding [29, 26]. Density gradient centrifugation analysis showed that the less buoyant merlin has direct or indirect association with the F-actin cytoskeleton. However, in the open conformation merlin dissociates from one or more of its binding partners [32]. It has been shown that during transition to a growth suppressive state, merlin dissociates from actin based structures [14]. Merlin mediated growth suppression requires HRS [4]. Since both HRS and beta II spectrin are thought to share a common binding site in the C-terminal region of merlin protein, the loss of interaction of actin could be due to the dissociation of beta II spectrin from merlin C-terminus. It has been proposed that, merlin can interact directly with actin through its N terminal residues and indirectly through association with beta II spectrin. As the residues involved in the proposed direct interaction are highly conserved among all merlin and ezrin, moesin and radixin proteins, these residues are considered as common protein binding motif rather than actin binding motifs. Therefore, indirect binding through beta II spectrin may be considered as a possible way of merlin- actin binding and may be possible only in the open conformation of merlin. The presence of an overlapping binding site for both HRS and beta II spectrin in the C-terminus of merlin could be indicative of the dual functional nature of merlin. Our model predicts that, in its closed conformation merlin prevents other proteins to bind to the FERM domain through masking of its B and some part of C sub domains. However, experimental studies in this direction are necessary to validate the predictions made in this study

4. Summary and conclusion

We present here the model of closed (active) conformation of human merlin protein. It has been proposed that, beta II spectrin which requires merlin in its open conformation to interact with actin cannot bind, since C-terminal region is found to be masked by its interaction with FERM domain. Our model predicts that merlin, in its closed conformation, prevents other proteins to bind to the FERM domain through its B and some part of C sub domains.

References

[1] Ramesh V. (2004) Nature Reviews Neuroscience, 5, 462-470.
[2] Rouleau G.A., Merel P., Lutchman M., Sanson M., Zucman J., Marineau C., Hoang-Xuan K., Demczuk S., Desmaze C. and Plougastel B. (1993) Nature, 363, 515-521.
[3] Trofatter J.A., MacCollin M.M., Rutter J.L., Murrell J.R., Duyao M.P., Parry D.M., Eldridge R., Kley N., Menon A.G. and Pulaski K. (1993) Cell, 75, 826.
[4] Sun C.X., Haipek C., Scoles D.R., Pulst S.M., Giovannini M., Komada M. and Gutmann D.H. (2002) Human Molecular Genetics, 11, 3167-3178.
[5] Tsukita S., Yonemura S. and Tsukita S. (1997) Trends in Biochemical Sciences, 22, 53-58.
[6] Nguyen R., Reczek D. and Bretscher A. (2001) Journal of Biological Chemistry, 276, 7621-7629.
[7] Arpin M., Algrain M. and Louvard D. (1994) Current Opinion in Cell Biology, 6, 136-141.
[8] Vaheri A., Carpen O., Heiska L., Helander T.S., Jaaskelainen J., Majander-Nordenswan P., Sainio M., Timonen T. and Turunen O. (1997) Current Opinion in Cell Biology, 9, 659-666.
[9] Bretscher A., Chambers D., Nguyen R. and Reczek D. (2000) Annual Review of Cell and Developmental Biology, 16, 113-143.
[10] McClatchey A.I., Saotome I., Mercer K., Crowley D., Gusella J.F., Bronson R.T. and Jacks T. (1998) Genes & Development, 12, 1121-1133.
[11] LaJeunesse D.R., McCartney B.M. and Fehon R.G. (1998) Journal of Cell Biology, 141, 1589-1599.
[12] Morrison H., Sherman L.S., Legg J., Banine F., Isacke C., Haipek C.A., Gutmann D.H., Ponta H. and Herrlich P. (2001) Genes & Development, 15, 968-980.
[13] Sherman L., Xu H.M., Geist R.T., Saporito-Irwin S., Howells N., Ponta H., Herrlich P. and Gutmann D.H. (1997) Oncogene, 15, 2505-2509.
[14] Shaw R.J., Paez J.G., Curto M., Yaktine A., Pruitt W.M., Saotome I., O'Bryan J.P., Gupta V., Ratner N., Der C.J., Jacks T. and McClatchey A.I. (2001) Developmental Cell, 1, 63-72.
[15] Kissil J.L., Johnson K.C., Eckman M.S. and Jacks T. (2002) Journal of Biological Chemistry, 277, 10394-10399.
[16] Rong R., Surace E.I., Haipek C.A., Gutmann D.H. and Ye K. (2004) Oncogene, 23, 8447-8454.
[17] Altschul S.F., Gish W., Miller W., Myers E.W. and Lipman D.J. (1990) Journal of

 Copyright © 2010, Bioinfo Publications, International Journal of Systems Biology, ISSN: 0975–2900, Volume 2, Issue 1, 2010

Molecular Biology, 215, 403-410.
[18] Altschul S.F., Madden T.L., Schaffer A.A., Zhang J., Zhang Z., Miller W. and Lipman D.J. (1997) Nucleic Acids Research, 25, 3389-3402.
[19] Bateman A., Birney E., Cerruti L., Durbin R., Etwiller L., Eddy S.R., Griffiths-Jones S., Howe K.L., Marshall M. and Sonnhammer E.L. (2002) Nucleic Acids Research, 30, 276-280.
[20] Sal A. & Blundell T.L. (1993) Journal of Molecular Biology, 234, 779-815.
[21] Raghava (2002) CASP5, A-132.
[22] Turunen O., Wahlstrom T. and Vaheri A. (1994) Journal of Cell Biology, 126, 1445-1453.
[23] Golovnina K., Blinov A., Akhmametyeva E.M., Omelyanchuk L.V. and Chang L.S. (2005) BMC Evolutionary Biology, 5, 69.
[24] Baines A.J. (2006) BMC Genomics, 7, 85.
[25] Laskowski R.A., Rullmannn J.A., MacArthur M.W., Kaptein R. and Thornton J.M. (1996) Journal of Biomolecular NMR, 8, 477-486.
[26] Scoles D.R., Huynh D.P., Chen M.S., Burke S.P., Gutmann D.H. and Pulst S.M. (2000) Human Molecular Genetics, 9, 1567-1574.
[27] Shimizu T., Seto A., Maita N., Hamada K., Tsukita S., Tsukita S. and Hakoshima T. (2002) Journal of Biological Chemistry, 277, 10332-10336.
[28] Pearson M.A., Reczek D., Bretscher A. and Karplus P.A. (2000) Cell, 101, 259-270.
[29] Neill G.W. and Crompton M.R. (2001) Biochemical Journal, 358, 727-735.
[30] Komada M. and Kitamura N. (1995) Molecular and Cell Biology, 15, 6213-6221.
[31] Gutmann D.H., Haipek C.A., Burke S.P., Sun C.X., Scoles D.R. and Pulst S.M. (2001) Human Molecular Genetics, 10, 825-834.
[32] Stickney J.T., Bacon W.C., Rojas M., Ratner N. and Ip W. (2004) Cancer Research, 64, 2717-2724.

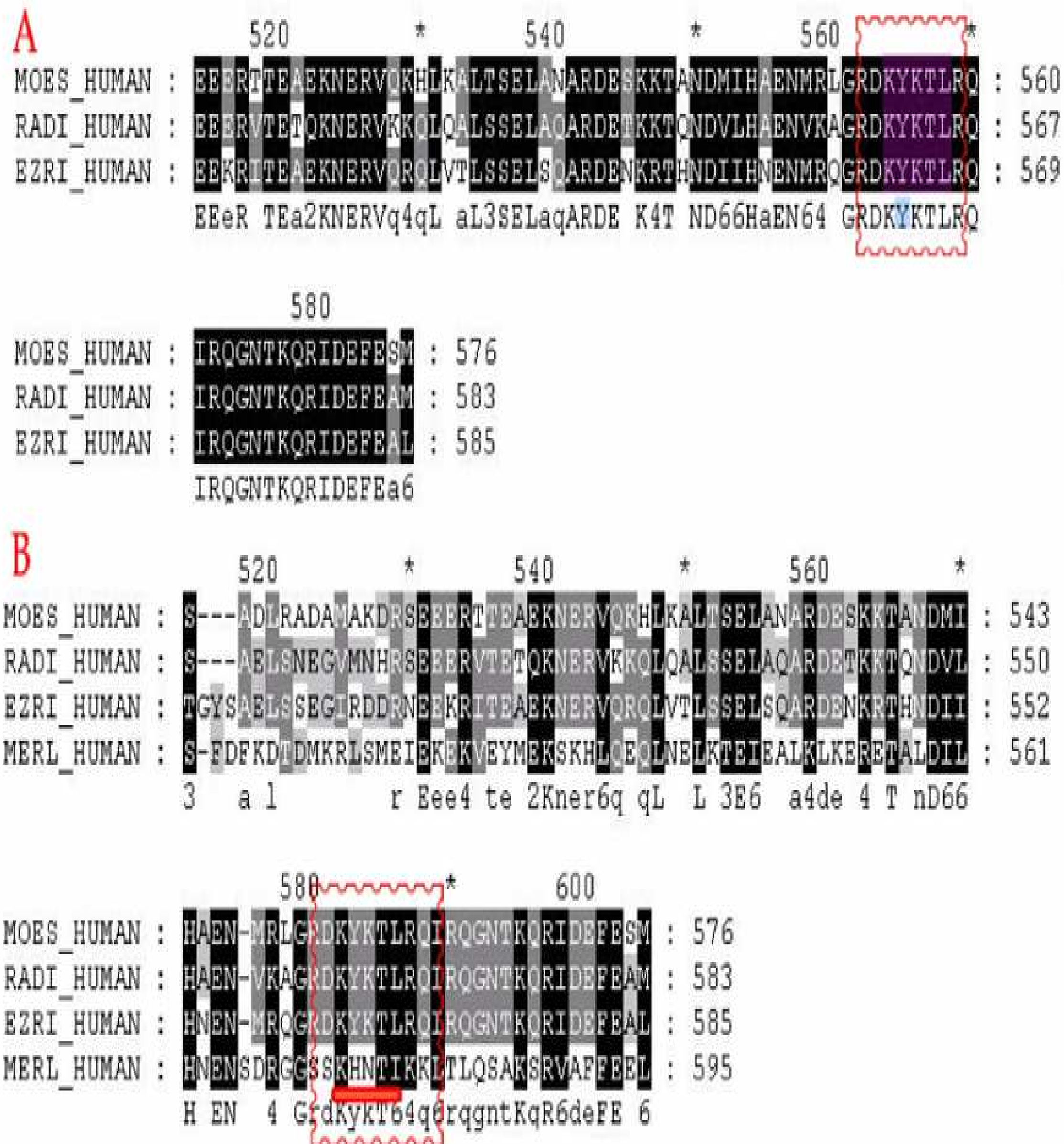

Fig. 1- Sequence alignment of COOH terminal actin binding domain in ezrin, moesin and radixin. (a) Sequence homology is centered on tyrosine marked in blue box. The “KYKTL” block represents the common actin binding site of ezrin, moesin and radixin proteins. (b) The alignment indicates the absence of “KYKTL” motif in merlin protein. The region is underlined in red in the figure. It is relevant to note that the conserved tyrosine in the motif is replaced by histidine in merlin.

 Copyright © 2010, Bioinfo Publications, International Journal of Systems Biology, ISSN: 0975–2900, Volume 2, Issue 1, 2010

```
2I1K_A          -----------------MPKSMNVRVTTMDAELEFAIQQTTTGKQLFDQVVKTIGLREVW
MERL_HUMAN      MAGAIASRMSFSSLKRKQPKTFTVRIVTMDAEMEFNCEMKWKGKDLFDLVCRTLGLRETW
                                  **::.**:.*****:**  : . .**:*** * :*:****.*

2I1K_A          FFGLQYTDSKGDLTWIKLYKKVMQQDVKKENPLQFKFRAKFYPEDVADELIQEITLKLFY
MERL_HUMAN      FFGLQYT-IKDTVAWLKMDKKVLDHDVSKEEPVTFHFLAKFYPENAEEELVQEITQHLFF
                *******  *. ::*:*: ***:::**.**:*: *:* ******:. :**:**** :**:

2I1K_A          LQVKNAILSDEIYCPPETSVLLASYAVQARHGDHNPAVHGPGFLANDRLLPQRVTDQHKM
MERL_HUMAN      LQVKKQILDEKIYCPPEASVLLASYAVQAKYGDYDPSVHKRGFLAQEELLPKRVINLYQM
                ****: **.::******:***********::**::*:**  ****::.***:** : ::*

2I1K_A          SREEWEQSITNWWQEHRGMLREDAMMEYLKIAQDLEMYGVNYFEIRNKKNTELWLGVDAL
MERL_HUMAN      TPEMWEERITAWYAEHRGRARDEAEMEYLKIAQDLEMYGVNYFAIRNKKGTELLLGVDAL
                : * **: ** *: ****  *::* ****************** *****.*** ******

2I1K_A          GLNIYEKDDKLTPKIGFPWSEIRNISFNDRKFIIKPIDKKAPDFVFFAPRVRVNKRILAL
MERL_HUMAN      GLHIYDPENRLTPKISFPWNEIRNISYSDKEFTIKPLDKKIDVFKFNSSKLRVNKLILQL
                **:**: :::*****.***.******:.*::* ***:***   * * :.::**** ** *

2I1K_A          CMGNHELYMRRRKPDTIDVQQMKAQAREEKLAKQAQREKLQLEIAARERAEKKQQEYQDR
MERL_HUMAN      CIGNHDLFMRRRKADSLEVQQMKAQAREEKARKQMERQRLAREKQMREEAERTRDELERR
                *:***:*:*****.*:::************  ** :*::*  *   **.**:.::* : *

2I1K_A          LRQMQEEMERSQANLLEAQDMILRLEEQLRQLQAAKEELEQRQNELQAMMQRLEETKNME
MERL_HUMAN      LLQMKEEATMANEALMRSEETADLLAEKAQITEEEAKLLAQKAAEAEQEMQRIKATAIRT
                * **:**   ::  *:.:::     * *: :  :   : * *:  * :  ***:: *

2I1K_A          AAERQKLEDEIRAKQEEVSRIQQEVELKDSETRRLQEEVEDARRKQDEAAAALLAATTPQ
MERL_HUMAN      EEEKRLMEQKVLEAEVLALKMAEESERRAKEADQLKQDLQEAREAERRAKQKLLEIATKP
                  *:: :*:::   :  . :: :* * : .*: :*::::::**. : .*   **  :*

2I1K_A          HHHVAERADTDPDHDNASD---AGSESGGGDLARGPDDLVDPVADRRTLAERNERLHNQL
MERL_HUMAN      TYPPMNPIPAPLPPDIPSFNLIGDSLSFDFKDTDMKRLSMEIEKEKVEYMEKSKHLQEQL
                 :    :    :     * .*    ..* * . . :       ::    ::    *:.::*::**

2I1K_A          KALKQDLARSCDETKETAMDKIHREN-VRQGRDKYKTLREIRKGNTKRRVDQFENM
MERL_HUMAN      NELKTEIEALKLKERETALDILHNENSDRGGSSKHNTIKKLTLQSAKSRVAFFEEL
                : ** ::      : :***:* :*.**  * * .*::*::::   .:* **  **::
```

Fig. 2- Sequence alignment of merlin and moesin template (2I1K). Alignment shows 48% of sequence identity between the sequences

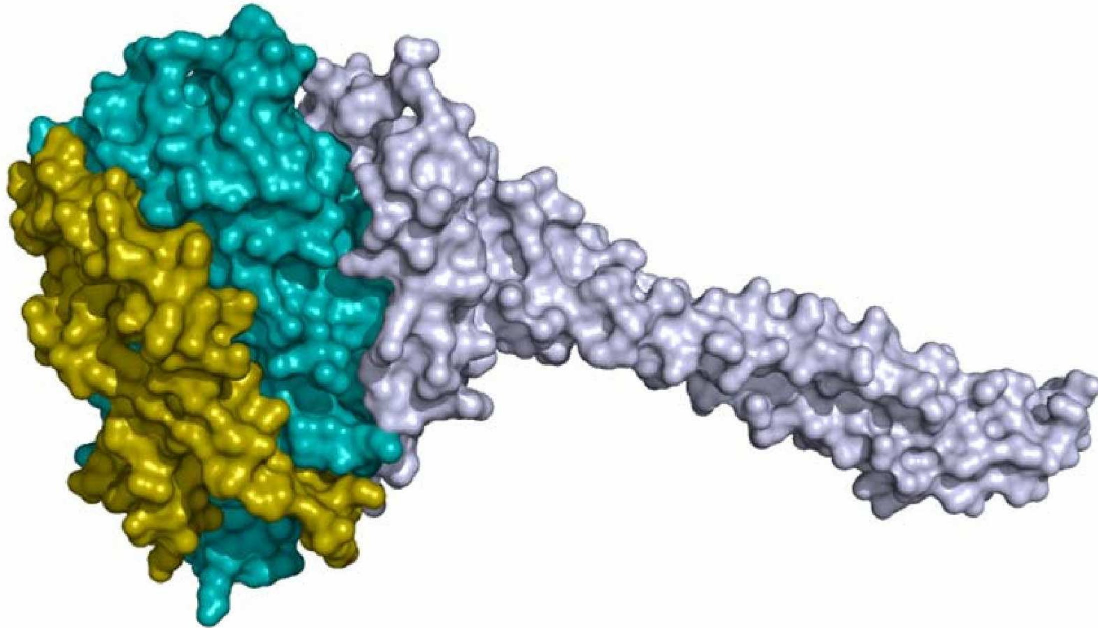

Fig. 3-The complete model of closed conformation of merlin. The cyan colored region shows the FERM domain, the light blue colored region represents the central helical domain and the yellow colored region represents the COOH terminal domain overlaying FERM domain.

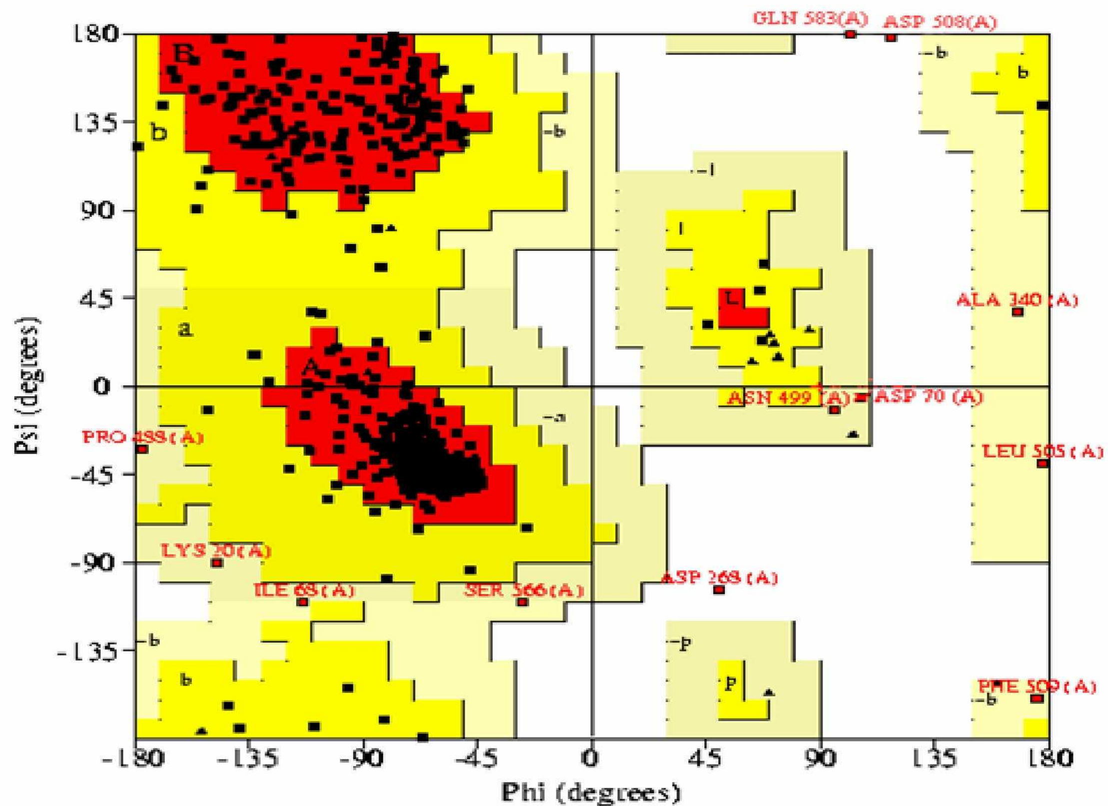

Fig. 4- Ramachandran plot of closed conformation of merlin generated by Procheck. The 578 residues of merlin were evaluated for its stereo quality. 94.8% of all residues were in favored regions, 3.2% of all residues were in allowed regions and 1.5% in generously allowed region.

Copyright © 2010, Bioinfo Publications, International Journal of Systems Biology, ISSN: 0975–2900, Volume 2, Issue 1, 2010

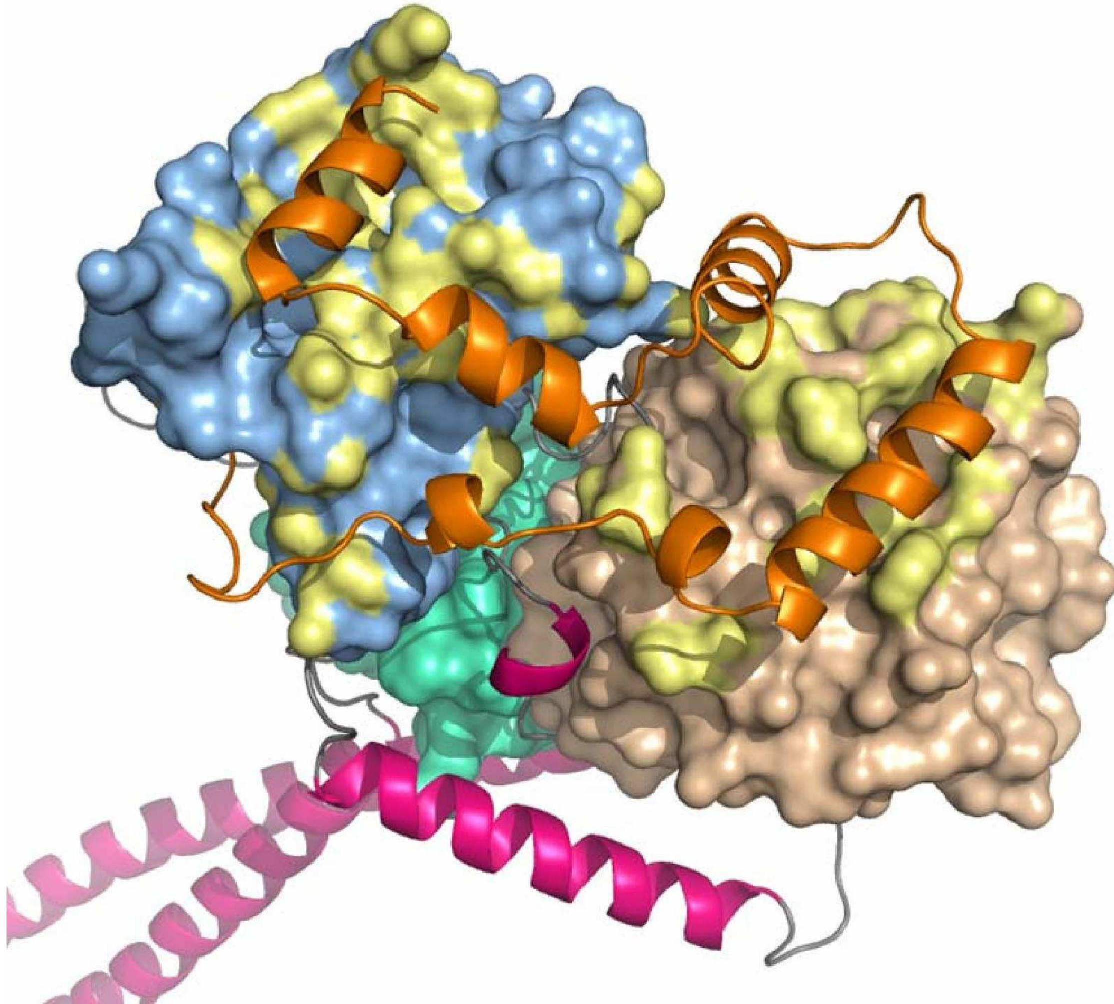

Fig. 5- The COOH terminal domain masking the subdomains B & C of FERM domain. COOH terminal domain (orange) masks the subdomain C (pale blue) at major portions, while subdomain B (wheat) is masked partially. Subdomain A (green) is free from masking. The yellow color patched on the subdomains (B&C) indicates the residues masked by COOH terminal domain.