Available online at www.sciencedirect.com
SCIENCE DIRECT
Journal of Molecular Graphics and Modelling 24 (2006) 271–277
Journal of Molecular Graphics and Modelling
www.elsevier.com/locate/JMGM

Homology modeling of the DNA-binding domain of human Smad5: A molecular model for inhibitor design

Ramkumar Hariharan, M. Radhakrishna Pillai *

Department of Molecular Medicine, Rajiv Gandhi Centre for Biotechnology and Regional Cancer Centre, Thiruvananthapuram 695014, Kerala, India

Received 11 June 2005; received in revised form 26 September 2005; accepted 26 September 2005
Available online 21 October 2005

Abstract

Members of the Smad protein family function as signal transducers of the transforming growth factor (TGF-beta) superfamily proteins. The human Smad5 protein, a signal transducer downstream of TGF-beta/BMP receptors, is composed of N-terminal DNA binding domain (MH1) and C-terminal protein–protein interaction domain (MH2) connected together by a linker motif. We used homology-modeling techniques to generate a reliable molecular model of the Smad5 MH1 domain based on the crystal structure of Smad3 MH1 domain. Our study presents the structural features of a BMP-regulated, R-Smad subfamily member (consisting of Smad1, Smad5 and Smad8) for the first time. This model provides a structural basis for explaining both functional similarities and differences between Smad3 and Smad5. Also, the structural model of this molecular target would be useful for structure-based inhibitor design because of its high accuracy. The results of our study provide important insights into understanding the structure–function relationship of the members of the Smad protein family and can serve to guide future genetic and biochemical experiments in this area.

Keywords: Homology modeling; Smads; Transforming growth factor-beta; Phylogenetic analysis; Inhibitor design

1. Introduction

The Smads constitute a family of intracellular proteins that function as signal transducers for the transforming growth factor-beta (TGF-beta) superfamily ligands. Activation of Smads leads to their translocation from the cytoplasm to the nucleus, where they regulate transcription of target genes in association with other transcription factors [1–3].

The eight human Smads, named numerically from Smad1 through Smad8, can be grouped into three subfamilies: the receptor-regulated Smads (R-Smads), the common Smads (Co-Smads), and the inhibitory Smads (I-Smads) (Table 1). Most Smad proteins are composed of an N-terminal Mad homology 1 (MH1) domain responsible for DNA binding and a carboxy-terminal Mad homology 2 (MH2) domain that functions in protein–protein interaction. These two conserved domains are connected together by a non-conserved proline-rich linker region essential for ubiquitinized degradation [4].

As shown in Table 1, Smad5 belongs to the receptor-activated Smads, which associate directly with TGF-beta receptor complex. Previous studies have shown that Smad5, together with Smad1 and Smad8, participates in signaling downstream of bone morphogenetic protein (BMP) receptors [5]. Numerous studies have also shown that Smad5 plays key signaling roles in hematopoiesis and osteogenesis [6].

In the present study, we used homology modeling to build the three-dimensional structures for the DNA-binding domain of human Smad5 using the experimentally determined crystal structure of Smad3 MH1 as modeling templates. Our reasons for undertaking this study were two-fold.

Firstly, none of the BMP-regulated R-Smads have been characterized structurally in any detail. Based on biochemical studies to date, Smad5 appears to be the most important member of this subfamily, at least as far as cancer is concerned [7]. We wanted to understand the structural basis for the functional similarities and differences observed between Smad3 and Smad5 proteins. Such a structural analysis would help to explain the high-level of specific DNA-binding ability of Smad5, unique among the BMP-regulated Smads, observed in previous biochemical studies. A biochemical study had

* Corresponding author. Tel.: +91 471 2347975; fax: +91 471 2349303.
E-mail address: mrpillai@vsnl.com (M.R. Pillai).

1093-3263/$ – see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2005.09.009

Table 1
The human Smad family

Sl. no.	Subfamily	Members	Gene name
1	TGF-beta/activin-regulated-R-Smads	Smad2	*MADH2*
		Smad3	*MADH3*
2	BMP-regulated R-Smads	Smad1	*MADH1*
		Smad5	*MADH5*
		Smad8	*MADH9*
3	Common Smad	Smad4	*MADH4*
4	Inhibitory Smads	Smad6	*MADH6*
		Smad7	*MADH7*

The table shows the classification of all eight known human Smad family proteins into four subfamilies and their gene names. Note that the gene name for Smad8 is *MADH9* and not *MADH8*.

already characterized the DNA-binding property of Smad5. The study had concluded that Smad5, but not Smad1 and Smad8, was able to bind DNA specifically at the Smad binding element (SBE) at a level similar to the binding by Smad3 and Smad4 [8]. Even though the high sequence similarity (∼72%) between the DNA binding domains of Smad5 and the structurally well-characterized Smad3 does not suggest novel features in Smad5 MH1 domain architecture, we wanted to reveal the extent of structural conservation. On the other hand, given the high degree of sequence conservation between Smad3 and Smad5 MH1 domains, the Smad5 MH1 model should provide the structural basis for their functional differences.

Secondly, recent biochemical studies have demonstrated that Smad5 is overexpressed and not deleted in colorectal cancer and in hepatoma cell lines [7,9]. This conflicts with earlier findings, which had assigned Smad5 a 'candidate tumor suppressor' status [10]. Hence, we felt that a reliable model of Smad5 MH1 domain would guide future biochemical and genetic efforts in its evaluation as a potential therapeutic target. Furthermore, an accurate molecular model of Smad5 MH1 domain could aid in the structure-based inhibitor design for antagonists against the DNA-binding function of human Smad5, an emerging molecular target.

2. Methodology

The entire computational analysis was done on a Pentium 4, 2.8 GHz processor running on Windows XP professional.

2.1. Amino acid sequence alignment

The MH1 domain sequences of all known human Smads were retrieved from SWISS-PROT. A multiple sequence alignment for these sequences was then generated using CLUSTAL W (default parameters) [11]. CHROMA was used to annotate and format the multiple sequence alignment [12].

2.2. Phylogenetic analysis

An unrooted phylogenetic tree for the human Smad family based on MH1 sequences was constructed using algorithms contained within the PHYLIP Phylogeny Inference Package, version 3.5c [13]. Prior to this, the gap rich regions of the MH1

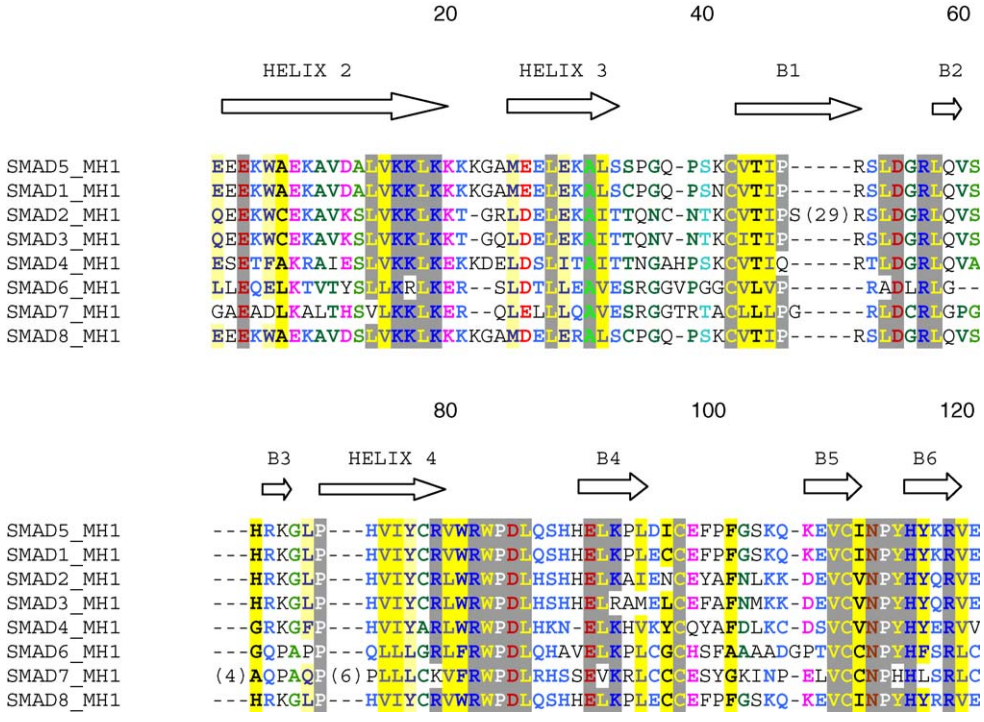

Fig. 1. Multiple sequence alignment of Mad homology 1 domains of all known human Smads. Highly conserved columns are shown with background color shading. The positions of the secondary structural elements for Smad5 are shown as arrows above the sequence alignment. B1–B6 represents the six short beta strands identified in the Smad5 MH1 domain model. Note that Smad2 and Smad7 domains contain long insertions that disrupt the B1 and helix 4 elements. CHROMA was used to format the multiple sequence alignment.

multiple sequence alignment was removed using GENEDOC, a free standing sequence alignment editor and formatter [14]. PROTDIST was used on these sequences to calculate a distance matrix according to the Dayhoff PAM probability model [15]. The calculated distances represent the expected fraction of amino acid substitutions between each sequence pair. This distance matrix was then used to estimate the phylogenies using the Neighborhood Joining (NJ) method [16]. Bootstrapping was carried out using SEQBOOT (1000 replicates for the PAM substitution model). CONSENSE was then used to generate the consensus tree by the majority-rule method. The final unrooted tree diagram was generated using PHYLODENDRON [17].

2.3. Homology modeling

The entire MH1 domain sequence of Smad5 was retrieved from the NCBI protein sequence databank since the corresponding entry in SWISS-PROT lacks a part of the N-terminal region. Suitable template for modeling the Smad5 MH1 domain was found via a simple BLASTp search of PDB [18]. The crystal structure of Smad3 DNA binding domain (PDB id: 1ozj chain A), sharing 73% sequence identity with the target sequence was used as template for building a homology model of the Smad5 MH1 domain. Although target–template similarity was high, to avoid problems arising out of bad target–template alignment, the Smad5 MH1 domain sequence was manually threaded into the template structure using DeepView (Swiss-Pdb viewer) version 3.7 [19]. The resulting project file was sent to the SWISS-MODEL automated homology model-building server for model calculation. Stereochemical quality check was performed on the returned model using the recently described VADAR server [20]. A few problematic side chain conformations were identified and rectified. The resulting structure was energy minimized using DeepView and the process was repeated.

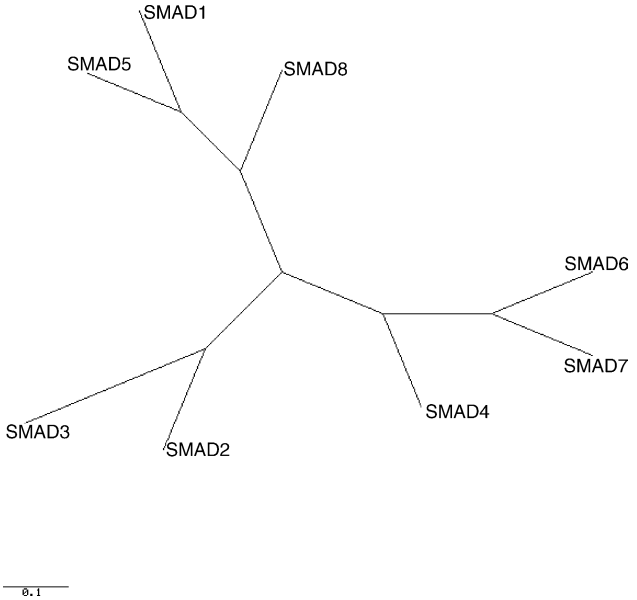

Fig. 2. Phylogenetic relationships between the members of the human Smad family. PHYLIP was used to generate the unrooted tree. PHYLODENDRON, a freestanding program was used to create the tree diagram.

2.4. Structure representation

Structure representations of the generated molecular models were done using PyMOL, a free-standing molecular graphics package.

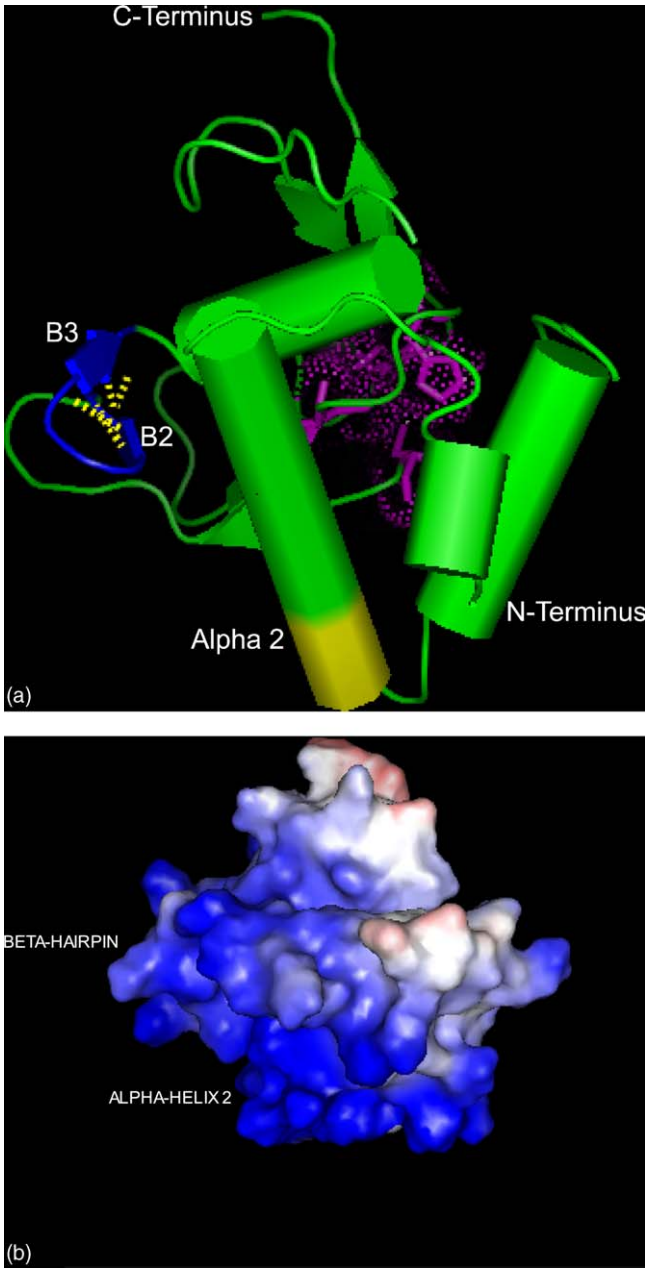

Fig. 3. Molecular models of the Smad5 DNA-binding domain. (a) Green colored solid cylinders represent the alpha helices. Beta strands are shown as flat arrows in the same color. The nuclear localization signal (yellow) maps to the N-terminal alpha 2 helix. The beta-hairpin believed to be responsible for specific DNA binding, is shown in blue, whereas, the hydrogen bonds stabilizing the motif is shown as yellow dashes. The sidechains of the four key residues that can potentially coordinate a zinc atom are shown in magenta, with a dotted surface. (b) The vacuum electrostatics generated for the model shows that the two key regions of the molecule responsible for DNA-binding are basic (blue) with other surface regions of the molecule being predominantly uncharged (white). PyMOL was used to create structural representations.

SWISS-PROT accession numbers for human Smad entries are given below Smad1 (Q15797), Smad2 (Q15796), Smad3 (P84022), Smad4 (Q13485), Smad5 (Q99717), Smad6 (O43541), Smad7 (O15105), Smad8 (O15198). (In SWISS-PROT, human Smad8 is still entered under its previous name i.e. Smad9.)

3. Results

3.1. Sequence alignment and phylogenetic analysis

All available human Smad MH1 domain sequences were compiled from SWISS-PROT for multiple sequence alignment. The sequence alignment itself is shown in Fig. 1. The results show that evolutionary conservation of MH1 domain sequences is high within the human Smad family, especially within the secondary structural elements. Also, with the exception of Smad2 and Smad7, insertions and deletions are infrequent in this domain.

The unrooted phylogenetic tree in Fig. 2 shows the relationship of Smad5 with other human Smads. Not surprisingly, Smad5 is more closely related to Smad1 and Smad8, the BMP-regulated Smads than to the other members. However, since the MH1 domains of Smad1 and Smad8 have not been experimentally determined, Smad3 MH1 domain structure was used as the modeling template.

3.2. Homology modeling

The predicted structure of the DNA binding domain of Smad5 is shown in Fig. 3a and b. The overall structure of the Smad5 MH1 domain is similar to that of Smad3. This Smad5 DNA binding domain adopts a compact globular fold with four alpha-helices, six short beta-strands and five loops. The most striking feature of this model is the beta-hairpin, formed by strands B2 and B3. This motif is striking in the sense that it was identified as a novel DNA-binding motif when it was first described in the crystal structure of Smad3 MH1 domain. The beta-hairpin, protruding outward from the globular MH1 core, is held rigidly in place by intrachain hydrogen bonds. The highly conserved nuclear localization signal (KKLKK), known to be present in the domain, maps to the N-terminal alpha helix 2 in this structure. The four key residues, known to be involved in coordinating a zinc atom in the Smad3 MH1 domain, also map to corresponding regions in this model.

The positions in Smad5 MH1 domain that differ from equivalent positions in the template were mapped on to the model. The results are shown in Fig. 4. Most of the non-conserved residues map to the surface loops of the domain, whereas synonymous positions map onto secondary structural elements.

The accuracy of the model was estimated as sequence divergence from the template as well as a stereochemical

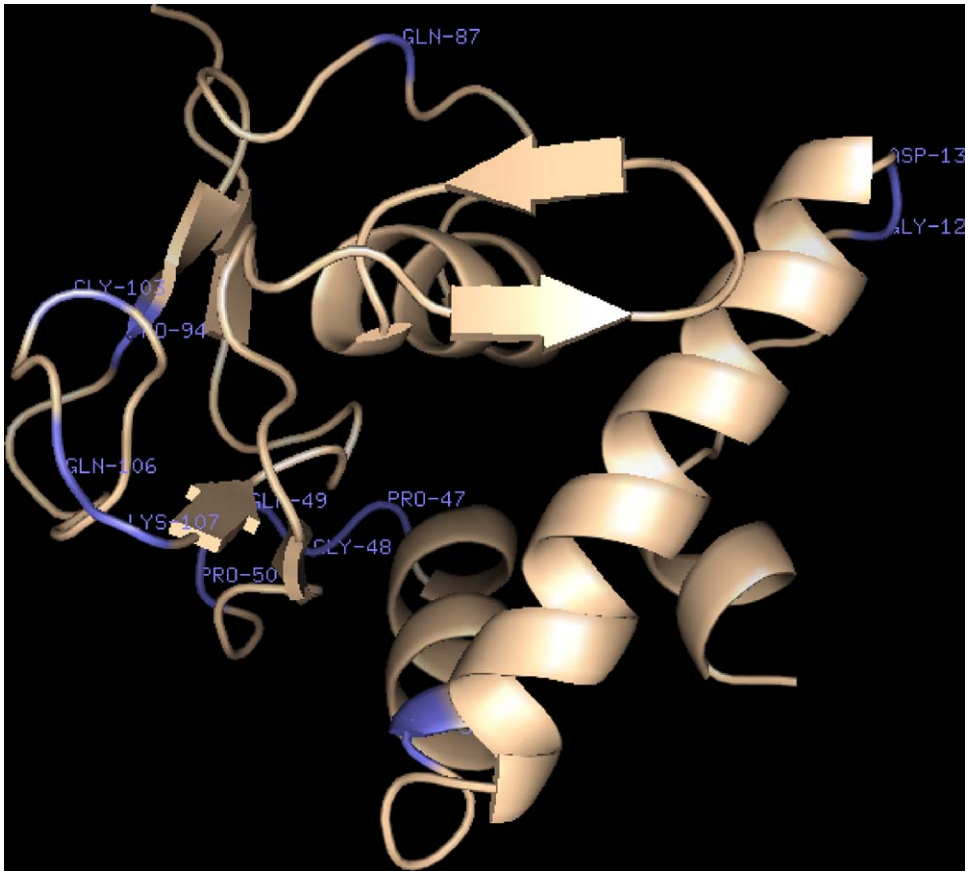

Fig. 4. Ribbon model of Smad5 MH1 domain showing structurally non-conserved residues. The model is shown in wheat color, whereas, the variant non-synonymous residues (differing from equivalent positions in Smad3 MH1 domain) are shown with labels in blue.

```
Smad3MH1  7      FTPPIVKRLL GWKKGEQNGQ EEKWCEKAVK SLVKKLKKTG -QLDELEKAI
Smad5MH1  1         PAVKRLL GWK---QGDE EEKWAEKAVD ALVKKLKKKK GAMEELEKAL
                    * ***** ***   *  . **** ****  .*******      ..*****.

Smad3MH1             hhhhhh             hhhhhhhhhh hhhhhhh       hhhhhhhh
Smad5MH1             hhhhh              hhhhhhhhhh hhhhhhhh      hhhhhhhh


Smad3MH1  56     TTQNVNTKCI TIPRSLDGRL QVSHRKGLPH VIYCRLWRWP DLHSHHELRA
Smad5MH1  45     SSPGQPSKCV TIPRSLDGRL QVSHRKGLPH VIYCRVWRWP DLQSHHELKP
                 ..    .**. ********** ********** *****.**** **.*****..

Smad3MH1         hh       s ssss     s ss  sss hh hhhhhh           ssss
Smad5MH1         hh       s ssss     s ss  sss hh hhhhhh           ssss


Smad3MH1  106    MELCEFAFNM KKDEVCVNPY HYQRVET
Smad5MH1  95     LDICEFPFGS KQKEVCINPY HYKRVES-
                 ...***.*   *  ***.*** ** ***.

Smad3MH1         s            sssss    sssss
Smad5MH1         s            sssss    sssss  -
```

(a)

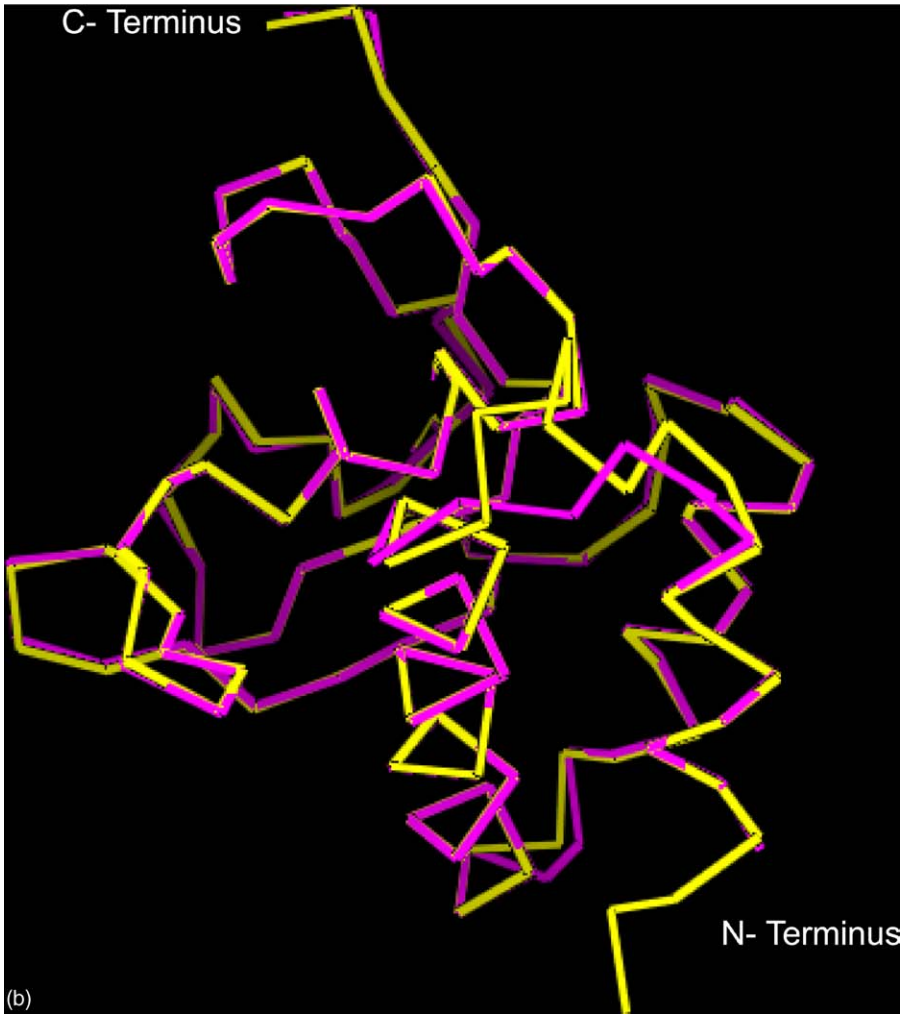

Fig. 5. (a) Shows the target–template sequence alignment used to generate the molecular model of Smad5 DNA-binding domain. Conserved positions are denoted by '*' below them. The secondary structure assignments of various regions are shown in the next line with 's' representing strands and 'h' for helices. (b) Shows the results of superimposition of the Smad3 MH1 onto the predicted structure of the Smad5 DNA-binding domain.

quality check of the model. The target–template alignment used to build the model is shown in Fig. 5a. For model validation as well as to quantify the overall structural similarity between the MH1 domains of Smad3 (the template) and Smad5 (target), the molecular model of Smad5 MH1 domain was superimposed on the crystal structure of Smad3 DNA binding domain. The root mean square deviation (RMSD) of the model with the experimentally determined structure of Smad3 (all backbone atoms), calculated using DeepView, is 0.46 Å. Fig. 5b shows the results of the superimposition.

4. Discussion

The molecular model of the DNA binding domain of human Smad5, generated by homology modeling, presents the structural features of a BMP-regulated, R-Smad subfamily member (consisting of Smad1, Smad5 and Smad8) for the first time. Of the eight known human Smads, the MH1 domain of only Smad3 has been characterized structurally [21].

One of the objectives of this study was, given the high sequence identity (∼73%) between Smad3 and Smad4 MH1 domains, to what extent was this similarity present at the structural level? A comparison between the corresponding domain structures shows that all the distinct structural elements shown to play key roles in the DNA-binding property of Smad3, are conserved almost entirely in the Smad5 MH1 domain structure. These structural elements include the protruding 11-residue beta-hairpin, the basic alpha-helix 2 at the N-terminus and the zinc-coordination site formed by the side chains of four conserved residues.

A biochemical study had already characterized the DNA-binding property of Smad5 [8]. The study had concluded that Smad5, but not Smad1 and Smad8, was able to bind DNA specifically at the Smad binding element (SBE) at a level similar to the binding by Smad3 and Smad4. This unique DNA-binding property of Smad5 can at least be partly explained by the presence of the key DNA-binding and zinc-coordinating elements in the correct structural context. Structural characterization of Smad1 and Smad8 MH1 domains at least using bioinformatics techniques could reveal the basis for their inability to bind DNA in a sequence specific manner. However, it is more likely that protein–protein interactions involving the MH2 domain and the linker motif (which has been shown to be important in transcriptional activation by Smads), in the case of Smad1 and Smad8, could prevent the nuclear export of these proteins from the cytoplasm thereby indirectly interfering with these proteins' DNA-binding property [22].

Most of the amino-acid residue differences between Smad5 and Smad3 map onto the surface loops of the MH1 domain model. Since, surface loops play an important role in mediating protein–protein interactions; this might be the basis of functional differences between the two proteins. However, since the DNA-binding elements in both proteins are almost identical in structure and position, most functional differences between Smad3 and Smad5 can be attributed to differences in the linker motif and MH2 domain (which is the Smad domain mainly involved in protein–protein interactions). The molecular modeling of the linker motif and MH2 domain can be the subject of another future study.

In an attempt to seek out the best available template structure to carry out homology modeling of the human Smad5 MH1 domain and to find evolutionary basis for the functional differences between the different human Smad family members, we carried out a detailed phylogenetic analysis of human Smads based on their MH1 domain sequences. Fig. 2 shows the unrooted phylogenetic tree of human Smads.

The overall topology of the tree is in agreement with the well-known fact that three distinct subfamilies exist within the human Smad family. Further, the analysis clearly shows the evolutionary divergence of the three sub-families away from each other, leading to functional specialization within the human Smad family. The inhibitory I-Smads, Smad6 and Smad7, arise from a common node and are separated from the other members of the family by Smad4, the only common-Smad sub-family member reported in humans.

Although the receptor regulated R-Smads, that transduce signals downstream of BMP-receptors (Smad1, Smad5 and Smad8) do not share a common node, the tree clearly shows that they form a distinct subgroup and are separated evolutionarily from the two activin-regulated R-Smads, viz, Smad2 and Smad3, branching out from the same node. Thus, even though the unrooted phylogenetic tree of the human Smads presented here, does not allow us to deduce their common ancestor, it clearly mirrors and provides evolutionary basis, behind the functional differences among the human Smad family members.

Finally, although Smad5 was initially given a candidate tumor suppressor status, two recent studies have shown that it is over-expressed in colorectal cancer and hepatoma cell lines [7,9]. Moreover, TGF-beta has been shown to have both tumor suppressor and oncogenic roles depending on the stage of the tumor [23]. Thus, Smad family members may turn out to be future therapeutic targets amenable to small-molecule perturbation and modulation. This warrants the elucidation of the role of Smad5 in carcinogenesis.

The molecular model of the human Smad5 DNA-binding domain presented in this study, mainly because of high template–target homology, is of sufficient quality as to be used in structure-based inhibitor studies and can therefore, aid in the design of small molecule modulators of Smad5 function and also serve to guide future biochemical experiments aimed at investigating Smad5 function.

5. Conclusion

In this paper, we have presented a reliable molecular model of the DNA-binding domain of human Smad5, thereby providing a structural basis for the biological functions of this protein. Similarities and differences between the human Smad family members, especially functional differences between the receptor regulated R-Smads, have been explained using inferences drawn from this structure, existing crystal structures, as well as from an evolutionary perspective by employing bioinformatic protocols to perform a detailed phylogenetic

analysis of all the human Smads. However, the most important conclusion of our modeling results is that, due to the high target–template similarity, the homology model presented here is of a sufficient quality to serve as a structural model for future structure-based inhibitor design efforts, focusing on human Smad5 as a molecular target in cancer.

Acknowledgement

Ramkumar Hariharan is supported by a research fellowship from the Council of Scientific and Industrial Research (CSIR), Government of India [F. No. 9/553 (14)/2003].

References

[1] J. Massague, S.W. Blain, R.S. Lo, TGF beta signaling in growth control, cancer and heritable disorders, Cell 103 (2000) 295–309.
[2] J.L. Wrana, Regulation of Smad activity, Cell 100 (2000) 189–192.
[3] P. Ten Dijke, K. Miyazono, C.H. Heldin, Signaling inputs converge on nuclear effectors in TGF-beta signaling, Trends. Biochem. Sci. 25 (2000) 64–70.
[4] Y. Shi, Structural insights on Smad function in TGF-beta signaling, Bioessays 23 (2001) 223–232.
[5] L. Attisano, J.L. Wrana, Smads as transcriptional co-modulators, Curr. Opin. Cell. Biol. 12 (2000) 235–243.
[6] B. Liu, N. Mao, Smad5: signaling roles in hematopoiesis and osteogenesis, Int. J. Biochem. Cell. Biol. 36 (5) (2004) 766–770.
[7] D.B. Zimonjic, M.E. Durkin, C.L. Keck-Waggoner, S.W. Park, S.S. Thorgeirsson, N.C. Popescu, SMAD5 gene expression, rearrangements, copy number, and amplification at fragile site FRA5C in human hepatocellular carcinoma, Neoplasia 5 (5) (2003) 390–396.
[8] W. Li, F. Chen, R.P. Nagarajan, X. Liu, Y. Chen, Characterization of the DNA-binding property of Smad5, Biochem. Biophys. Res. Commun. 286 (5) (2001) 1163–1169.
[9] J.R. Van Beijnum, P.T. Moerkerk, A.J. Gerbers, A.P. De Bruine, J.W. Arends, H.R. Hoogenboom, S.E. Hufton, Target validation for genomics using peptide-specific phage antibodies: a study of five gene products overexpressed in colorectal cancer, Int. J. Cancer 101 (2) (2002) 118–127.
[10] J. Zavadil, P. Svoboda, H. Liang, L.V. Kottickal, L. Nagarajan, An antisense transcript to SMAD5 expressed in fetal and tumor tissues, Biochem. Biophys. Res. Commun. 255 (3) (1999) 668–672.
[11] D. Higgins, J. Thompson, T. Gibson, J.D. Thompson, D.G. Higgins, T.J. Gibson, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, Nucleic Acids. Res. 22 (1994) 4673–4680.
[12] L. Goodstadt, C.P. Ponting, CHROMA: consensus-based coloring of multiple alignments for publication, Bioinformatics 17 (9) (2001) 845–846.
[13] J. Felsenstein, Inferring phylogenies from protein sequences by parsimony, distance, and likelihood methods, Meth. Enzymol. 266 (1996) 418–427.
[14] K.B. Nicholas, H.B. Nicholas Jr., D.W. Deerfield, GeneDoc: analysis and visualization of genetic variation, Embnew News 4 (1997) 14.
[15] M. Dayhoff, Atlas of Protein Sequence and Structure, National Biomedical Research Foundation, Washington, DC, 1978.
[16] N. Saitou, M. Nei, The neighbor-joining method: a new method for reconstructing phylogenetic trees, Mol. Biol. Evol. 4 (1987) 406–425.
[17] http://iubio.bio.indiana.edu/treeapp/, PHYLODENDRON, Phylogenetic Tree Drawing.
[18] Altschul, F. Stephen, G. Warren, M. Webb, E.W. Myers, D.J. Lipman, Basic local alignment search tool, J. Mol. Biol. 215 (1990) 403–410.
[19] N. Guex, M.C. Peitsch, SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling, Electrophoresis 18 (15) (1997) 2714–2723.
[20] L. Willard, A. Ranjan, H. Zhang, H. Monzavi, R.F. Boyko, B.D. Sykes, D.S. Wishart, VADAR: a web server for quantitative evaluation of protein structure quality, Nucleic Acids. Res. 31 (13) (2003) 3316–3319.
[21] Y. Shi, Y.F. Wang, L. Jayaraman, H. Yang, J. Massague, N.P. Pavletich, Crystal structure of a Smad MH1 domain bound to DNA: insights on DNA binding in TGF-beta signaling, Cell 94 (1998) 585–594.
[22] L. Xu, Y.G. Chen, J. Massague, The nuclear import function of Smad2 is masked by SARA and unmasked by TGF beta-dependent phosphorylation, Nat. Cell. Biol. 2 (2000) 559–562.
[23] E. Piek, A.B. Roberts, Suppressor and oncogenic roles of transforming growth factor-beta and its signaling pathways in tumorigenesis, Adv. Cancer Res. 83 (2001) 1–54.