UDC 593.4:575.852:577.217
ISSN 1330-9862

scientific note

(FTB-1276)

HMGB2 Protein from the Marine Sponge *Suberites domuncula***

Helena Ćetković and Lada Lukić-Bilela*

Department of Molecular Biology, Ruđer Bošković Institute, Bijenička 54, HR-10000 Zagreb, Croatia

Received: October 16, 2003
Accepted: November 10, 2003

Summary

Investigation of the phylogenetically conserved genes/proteins from marine sponges (Porifera), the most primitive metazoan phylum, can be applied for the reconstruction of ancient structures of genome/proteom complexity in the ancestral organism common to all multicellular animals. The complete nucleotide sequence of *Suberites domuncula* (Demospongiae) cDNA coding for 183-amino-acid protein (21.3 kDa), which displays high overall similarity in primary structure and organization of domains with HMGB canonical proteins, belonging to High Mobility Group (HMG) family of nuclear proteins, is reported here. The major role of these non-specific DNA binding proteins is to facilitate the formation of complex nucleoprotein assemblies and nucleosome remodelling. The encoded protein, named HMGB2SD, contains two typical, highly conserved DNA-binding domains: box A (69 aa) and box B (71 aa). Short C-terminal »tail« is only 13 aa long. HMG2SD displays the highest overall similarity (58 %) with HMGB2 proteins from mammals (pig, human, rat, mouse), higher than with *Drosophila melanogaster* (53 %) or *Caenorhabditis elegans* (31 %) homologues. This is in accordance with previous results, which showed the best homology between sponge and mammalian homologues/ortologues. Our results further confirm that sponges are an excellent model for molecular evolutionary studies.

Key words: Porifera, *Suberites domuncula*, High Mobility Group proteins, HMGB2, molecular evolution

Introduction

The High Mobility Group (HMG) chromosomal proteins are among the most abundant and ubiquitous non-histone proteins found in nuclei of higher eukaryotes. They were originally isolated from the mammalian cells and named according to their electrophoretic mobility in polyacrilamide gels. HMGB 1 and 2 proteins, one of the three classes of HMG proteins, bend DNA and bind preferentially to distorted DNA structures. They appear to act primarily as architectural facilitators in the assembly of nucleoprotein complexes in which the bound DNA is often tightly bent (*1*). HMGB proteins have been reported to play vital roles in transcription, replication, recombination and cellular differentiation (*2*). They show a preferential interaction with supercoiled plasmid DNA (*3*) and non-B-type DNA such as four-way junction (*4*), cruciform DNA (*5*), cisplatin modified DNA (*6*) and B-Z junction (*7*). Recent genetic and biochemical evidence suggests that these proteins can facilitate nucleosome remodeling (*8*).

These proteins have highly conserved primary sequences and tertiary structures (*9,10*); they consist of two similar, but distinct, tandem DNA-binding domains

* Corresponding author; Phone: ++385 1 4571 239; Fax: ++385 1 4561 177; E-mail: cetkovic@rudjer.irb.hr
** This paper was presented at the 8th Croatian Biological Congress in Zagreb, 27 Sept.–2. Oct. 2003

called HMG-boxes A and B with non-identical amino acid sequences and a long acidic C-terminal »tail« consisting of ~30 (HMG1) or ~20 (HMG2) acidic (Asp and Glu) residues (*11*,*12*). HMGB2 protein binds with DNA duplex in a sequence-nonspecific manner by two HMG boxes, then bends and unwinds the DNA. Both HMG boxes are required for full activity of the DNA structural distortion, but C terminal tail is necessary for inducing the peculiar distorted structures of higher affinity to HMGB2 (*2*). The HMG box family appeared more than 1000 million years ago (*13*), indicating that many proteins important for the function of complex organisms had already been created in the common ancestor of sponges and other multicellular animals, which strongly speaks in favour of the monophyletic origin of all metazoan phyla, including Porifera (*14*,*15*).

Sponges (Porifera) are an excellent model organism for molecular evolutionary studies; they date back to at least 580 million years ago (*16*). As they represent the lowest metazoan phylum that existed prior to the Cambrian explosion (*17*), they can be considered as living fossils (*18*). Systematic analyses of genes, cDNAs and their deduced protein sequences from sponges allow an experimental approach to reflect upon the molecular events that occurred during the transition from Protozoa to Metazoa. Sponges have long been considered as rich sources for novel bioactive compounds rendering them useful for applications in medicine. The new evolutionary insights, in addition to the technique of sponge cell culture, especially the primmorph system, have increased the attractiveness of Porifera also for biotechnology (*19*,*20*).

We report here the complete nucleotide sequence and phylogenetic analysis of a cDNA from the demospongian sponge *Suberites domuncula*, which encodes a HMGB2 protein belonging to the High Mobility Group family of nuclear proteins.

Materials and Methods

Isolation and characterization of HMGSD cDNA

Live specimens of sponge *Suberites domuncula* (Porifera, Demospongiae, Tetractinomorpha, Hadromerida, Suberitidae) were collected from the Northern Adriatic Sea near Rovinj, Croatia. The sponge material was immediately frozen in liquid nitrogen until usage. Total RNA was extracted from sponge tissue and polyadenylated mRNA was isolated from total RNA as described previously (*21*). The preparation of *S. domuncula* cDNA library in λZAP Express™ vector (Stratagene, La Jolla, USA) was already described (*22*). During the screening of the sponge library for recombinant phages encoding evolutionary conserved genes, a λ phage carrying a cDNA for HMGSD was also identified. Phagemid pBK-CMV with HMGSD cDNA insert was excised *in vivo* from lambda vector using 704 helper virus and *E. coli* strain XLORL cells using Rapid Excision Kit (Stratagene).

The nucleotide sequence of the cDNA insert was determined on the automated DNA sequencer ALF Express (Pharmacia, Uppsala, Sweden) using Thermo-Sequenase Cy5 Terminator Kit (Amersham-Pharmacia, Uppsala, Sweden). The universal primers for sequencing from the insert ends, as well as one specifically designed, were used to obtain the complete primary structure of HMGSD cDNA. The sequence of the HMGSD cDNA insert was obtained at least twice.

Sequence analysis

Nucleotide and protein sequences were stored and analyzed using PC/GENE 14.0 programs from IntelliGenetics (Mountain View, CA, USA). Homology searches and sequence retrieval were done *via* the Internet server at the National Center for Biotechnology Information, National Institute of Health, Bethesda, MD, USA (http://www.ncbi.nlm.nih.gov). Multiple sequence alignment (MSA) was performed with CLUSTAL X program (*23*) and its graphic presentation by the program GeneDoc (*24*).

Result and Discussion

The complete sequence has been deposited to EMBL databank under the accession number AY436705.

Homology searches, performed *via* e-mail server BLASTX, identified the protein encoded by this cDNA from *S. domuncula* as HMGB protein, belonging to High Mobility Group (HMG) of nuclear proteins. The protein was therefore named HMGB2SD. The cDNA is 1093 nucleotides (nt) long, excluding the polyA tail and contains a 5'non-coding sequence of 243 nt, followed by the open reading frame (nt 244-794). 3'non-coding sequence is 298 nt long and extremely rich in AT residues, which has already been noticed (*12*). The open reading frame codes for HMGB2SD protein of 183 aa, with a calculated M_r of 21377.88 Da and a pI of 6.75. Non-translated regions at both ends of the cDNA are very long, which was often found in sponge transcripts (*25*). Careful inspection of the untranslated sequences did not show the presence of significant secondary structures or additional ORFs. Like in the other cDNAs from marine sponges, the typical signal for polyadenilation, AATAAA (*26*), is not present at the 3'-end and it is unclear which sequence corresponds to this signal in sponges.

HMGB2SD is composed of two typical tandem HMG-box domains (A and B) and a long acidic C-terminal »tail«, which consists of 13 acidic (Asp and Glu) residues. The HMG-box A is 69 aa long and HMG-box B is 71 aa long. The nucleotide and amino acid sequence of HMGB2SD are shown in Fig. 1. Two HMG boxes (A and B) and C-terminal »tail« are indicated. HMGB2SD was compared with all HMGB proteins found in protein databases, as well as in »non-redundant« database using BLASTP at NCBI. The highest homology was found with HMG2 from mammals (human, pig, rat), slightly lower with zebra fish (*Danio rerio*) and even lower with insects (*Drosophila melanogaster* and *Anopheles gambiae*) (Table 1). HMGB2 protein from human and other higher vertebrates shows the highest similarity to HMGB2SD protein; also, it was found that sponge proteins are often similar to those of higher vertebrates. In contrast, the most diverged homologue of HMGs is found in *C. elegans*, which is distinctive because of the lack of one of the two DNA-binding boxes. Based on previous data, which indicates that evolutionary changes in sponges

```
GGCACGAGGGCCAGTACAGAGATCTAGGTACTATTTGGCGGGATATAGGCTTAGCACTATTGTTTGCCATAGACA   75
TACATACAGTGCTAATACAGGTGTATATCCTCCGACTGTATACATAGTCAATAGCACATATCCAACAAGTTGCAA  150
GTCTTTTTGGATCTCTTAGCTGAATATTAGCCCCTTACTGTTTGAAAAGGAGGTGAGAATCGCTGAAAGCCGACT  225
ATCTACTGCAACATCACCATGCCACCAAAGAAGGACCCAAACAAGCCCAAAGGTAGAACCAGCGCTTATGCGTTC  300
                     M  P  P  K  K  D  P  N  K  P  K  G  R  T  S  A  Y  A  F   19
TACGTGCAGGAAAGAAGAGACATCTACAGGAAAAATGGCGATACAGTCGTGTTTGCGCCTTTCTCTCAAGAGTGT  375
 Y  V  Q  E  R  R  D  I  Y  R  K  N  G  D  T  V  V  F  A  P  F  S  Q  E  C   44
GCTGAGTTGTGGAAGAATGTGAAAGACAAGAAGGCTTATCAAGATATGGCCGCTGTCGACAAAGAAAGATACGAC  450
 A  E  L  W  K  N  V  K  D  K  K  A  Y  Q  D  M  A  A  V  D  K  E  R  Y  D   69
AGAGAAATGGCCGAGTATGTTCCACCAGATGATGGTGAGAAATCCAAGAAGAAAAAGCAGAAAGACAAAACGAAG  525
 R  E  M  A  E  Y  V  P  P  D  D  G  E  K  S  K  K  K  K  Q  K  D  K  T  K   94
CCTAAAAGATCTTTGACAGCTTTCCTGTTCTTCTGTTCAGAAGAAAGACCCAAAATGAAAGAGAAAAATCCTGGA  600
 P  K  R  S  L  T  A  F  L  F  F  C  S  E  E  R  P  K  M  K  E  K  N  P  G   119
TCCTCTGTCGGTGATCTGGCTAAATTGCTCGGGGCTAAGTGGAAGGGAATGTCTGAAGATGATAAACAACCATTC  675
 S  S  V  G  D  L  A  K  L  L  G  A  K  W  K  G  M  S  E  D  D  K  Q  P  F   144
AGTGACATGGCGCAAGATGATAAAGATAGATATAACGACGAAATGGCATTGTGGAAAAAGGGTCAATTCAACAGA  750
 S  D  M  A  Q  D  D  K  D  R  Y  N  D  E  M  A  L  W  K  K  G  Q  F  N  R   169
CAAGAGGAGGAGGAAGGAGCAGATTATAGTGAAGAGGAAGATTAGTTTATTAAACAATGGTACTAGAGCACTGTT  825
 Q  E  E  E  E  G  A  D  Y  S  E  E  E  D  -  183
ATGTATTTACTAGCTATCTTAATGTGTTCTTTGGGCCAATCTTAAGTGCATATGTGCATAACTATATACGTTCCT  900
TACACGCCTTTCCTTGCTAAGTTTAATCTTTTGTTTCTGTTGTAAAGCTCTGTGTTGCCCTCCTTATTTACTTGT  975
CACCATTATTTCTTTTACCACAATAGGTGCATTCACACATCAGTACGTTATTATTAGTTAGACTTGCATTTATTT 1050
TTACTCACCTCTATTACATGTTTAAAAGCTGTTATTACACACT  1093
```

Fig. 1. The nucleotide sequence of the cDNA and the predicted amino acid sequence of *Suberites domuncula*. HMG box A is underlined, HMG box B is double underlined and C-terminal »tail« is underlined with wavy line

Table 1. The percentage of identity and overall similarity (in parenthesis) between four metazoan HMGB2 proteins. The accession numbers of HMGB2s are as in Fig. 2

	Suberites domuncula	human/mammalian	*Danio rerio*	*Drosophila melanogaster*
Suberites domuncula	100 %	42 % (58 %)	41 % (59 %)	36 % (53 %)
human/mammalian	/	100 %	73 % (87 %)	44 % (62 %)
Danio rerio	/	/	100 %	47 % (64 %)
Drosophila melanogaster	/	/	/	100 %

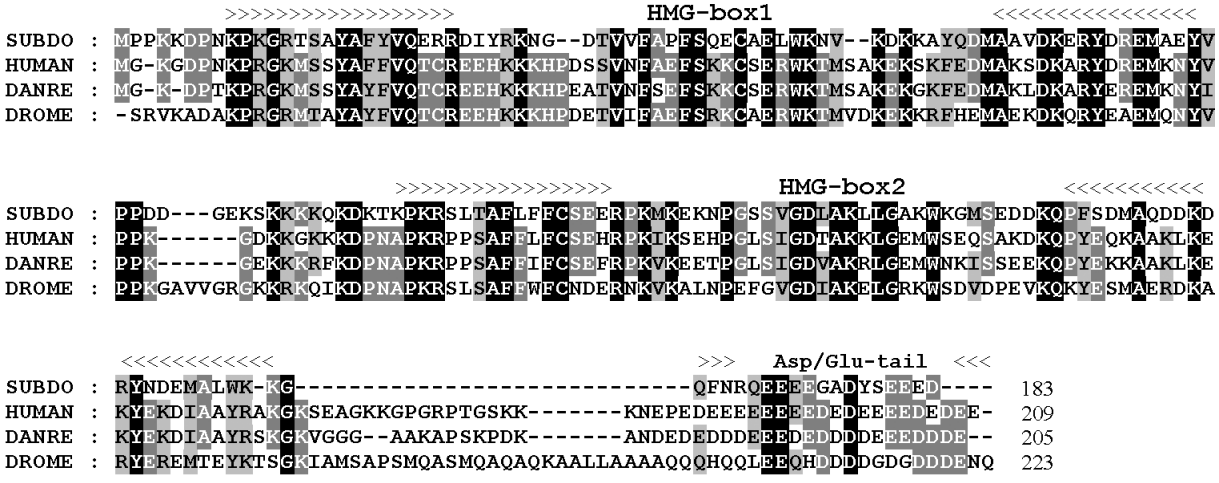

Fig. 2. Multiple alignment of four HMGB2 aa sequences from the multicellular organisms: *Suberites domuncula* (SUBDO, AY436705), *Homo sapiens* (HUMAN, P26583), *Danio rerio* (DANRE, AAH45917), *Drosophila melanogaster* (DROME, Q24537). HMG boxes A and B, as well as C-terminal »tail« are marked with arrows. 100 % conserved aa (identical + similar) are shown in white on black and 75 % conserved aa in white on gray

and mammals were relatively slow compared to the faster processes in insects and Nematodes, our analysis suggests that high similarity of sponge proteins to mammalian HMGs might be a consequence of the same evolutionary mechanisms. Multiple sequence alignment of four metazoan HMGB, produced by CLUSTAL X, is shown in Fig. 2. The degree of divergence between various HMGBs is mainly due to the sequence dissimilarities in the inter-domain regions, as the boxes are highly conserved through all HMG-box proteins, hence verifying the crucial involvement of boxes in the function of this protein family. High homology between sponge and mammalian homologous proteins has already been noted during our previous studies of ancient versions of genes/proteins from sponges. Although it is generally believed that sponges (Porifera) were the first to branch off from the common ancestor of Metazoa, more than 580 million years ago (*16*), many orthologues that are existing in sponge and mammalian genomes are missing in more recently evolved organisms like *C. elegans* (*27, 28*). Therefore, it is plausible to speculate that sponges, as well as humans or mammals, show, on average, slower molecular evolutionary rates in protein coding sequences than *D. melanogaster* and especially *C. elegans* (*28*). As a result, sponge proteins are often more similar to human than to *C. elegans* and *D. melanogaster* homologues/ortologues. This phenomenon is the most apparent in the comparisons of ancient proteins (genes), which probably existed in the ancestral progenitor common to all Metazoa (*27*).

Acknowledgements

This work was done in the Laboratory of Molecular Genetics, Ruđer Bošković Institute, under the supervision of Sc. D. Vera Gamulin and supported by grant from the Croatian Ministry of Science and Technology (P0098072 to V. G.)

References

1. J. O. Thomas, A. A. Travers, *TRENDS Biochem. Sci. 26* (2001) 167–174.
2. Y. Nakamura, M. Shimizu, M. Yoshida, *J. Biochem. (Tokyo) 131* (2002) 153–160.
3. S. Waga, S. Mizuno, M. Yoshida, *J. Biol. Chem. 265* (1990) 19424–19428.
4. J. R. G. Pohler, D. G. Norman, J. Bramham, M. E. Bianchi, D. M. J. Lilley, *EMBO J. 17* (1998) 817–826.
5. M. E. Bianchi, M. Beltrame, G. Paonessa, *Science, 243* (1989) 1056–1059.
6. Q. He, U. M. Ohndorf, S. J. Lippard, *Biochemistry, 39* (2000) 14426–14435.
7. S. Waga, S. Mizuno, M. Yoshida, *Biochem. Biophys. Res. Commun. 153* (1988) 334–339.
8. A. A. Travers, *EMBO Rep. 4* (2003) 131–136.
9. M. Bustin, D. A. Lehn, D. Landsman, *BBA-Gene Struct. Express, 1049* (1990) 231–243.
10. M. Bustin, R. Reeves, *Prog. Nucleic Acid. Res. Mol. Biol. 54* (1996) 35–100.
11. K. Tsuda, M. Kikuchi, K. Mori, S. Waga, M. Yoshida, *Biochemistry USA, 27* (1988) 6159–6163.
12. H. Shirakawa, K. Tsuda, M. Yoshida, *Biochemistry USA, 29* (1990) 4419–4423.
13. V. Laudet, D. Stehelin, H. Clevers, *Nucleic Acids Res. 21* (1993) 2493–2501.
14. W. E. G. Müller, *Naturwissenschaften, 82* (1995) 321–329.
15. W. E. G. Müller, I. M. Müller, B. Rinkevich, V. Gamulin, *Naturwissenschaften, 82* (1995) 36–38.
16. C. W. Li, J. Y. Chen, T. E. Hua, *Science, 279* (1998) 879–882.
17. A. H. Knoll, *Proc. Natl. Acad. Sci. USA, 91* (1994) 6743–6750.
18. W. E. G. Müller, *Naturwissenschaften, 82* (1998) 11–25.
19. G. Le Pennec, S. Perovic, M. S. A. Ammar, V. A. Grebenjuk, R. Steffen, F. Brümmer, W. E. G. Müller, *J. Biotechnol. 100* (2003) 93–108.
20. R. Osinga, *J. Biotechnol. 100* (2003) 91–92.
21. K. Pfeifer, M. Haasemann, V. Gamulin, H. Bretting, F. Fahrenholz, W. E. G. Müller, *Glycobiology, 3* (1993) 179–184.
22. M. Kruse, I. M. Müller, W. E. G. Müller, *Mol. Biol. Evol. 14* (1997) 1326–1334.
23. J. D. Thompson, D. G. Higgins, T. J. Gibson, *Nucleic Acids Res. 22* (1994) 4673–4680.
24. K. B. Nicholas, H. B. Jr. Nicholas, D. W. Deerfield, *Embnew. News, 4* (1997) 14.
25. H. Cetkovic, I. M. Müller, W. E. G. Müller, V. Gamulin, *Gene, 216* (1998) 77–84.
26. D. Zarkower, P. Stephenson, M. Sheets, M. Wickens, *Mol. Cell Biol. 6* (1986) 2317–2323.
27. V. Gamulin, I. M. Müller, W. E. G. Müller, *Biol. J. Linn. Soc. 71* (2000) 821–828.
28. H. Cetkovic, W. E. G. Müller, V. Gamulin, *Genomics* (in press).

HMGB2 protein iz morske spužve *Suberites domuncula*

Sažetak

Istraživanja filogenetski sačuvanih gena/proteina u morskim spužvama, najjednostavnijih Metazoa, mogu se primijeniti za rekonstrukciju strukture proteina i kompleksnosti genoma/proteoma u ancestralnom organizmu koji je bio zajednički predak svih višestaničnih životinja. U ovom je radu analizirana cDNA iz morske spužve *Suberites domuncula* (Demospongia) koja kodira protein dug 183 aminokiseline (ak), molekularne mase 21.3 kDa, što pokazuje najviši stupanj sličnosti u primarnoj strukturi i organizaciji domena sa HMGB proteinima iz obitelji HMG (High Mobility Group) proteina jezgre. Najvažnija je uloga tih nespecifičnih DNA-vežućih proteina posredovanje pri stvaranju nukleoprotein-

skih kompleksa i remodeliranju nukleosoma. Protein iz *S. domuncula*, nazvan HMGB2SD, čine dvije tipične, jako sačuvane DNA-vežuće domene: box A (69 ak) i box B (71 ak), te C terminalni »rep« od 13 aminokiselina. Nedavne analize brojnih cDNA i gena u spužvama, koji kodiraju evolucijski sačuvane proteine, pokazuju da su njihovi proteini sličniji svojim homolozima/ortolozima iz sisavaca. Isto je pokazano i na primjeru HMGB2SD proteina, koji pokazuje viši stupanj sličnosti (58 %) sa HMGB2 proteinima iz sisavaca (svinja, čovjek, štakor, miš), nego sa HMG2 iz kukca *D. melanogaster* (53 %) ili oblića *C. elegans* (31 %). To je prvi protein iz obitelji HMG proteina nađen u spužvama, te povrđuje da su one idealni modelni organizam u molekularno evolucijskim istraživanjima.