Volume 260, number 1, 67–69 FEB 07938 January 1990

RNA-binding protein-related sequence in a malaria antigen, clustered-asparagine-rich protein

Kei-ichi Kuma, Naoyuki Iwabe, Yu-ichi Kawanishi and Takashi Miyata

Department of Biology, Faculty of Science, Kyushu University, Fukuoka 812, Japan

Received 23 October 1989

Members of the RNA-binding protein superfamily contain RNA binding domains of about 90 amino acids with a highly conserved motif 'GFGF'. Using the conserved motif with some variations G-(F/Y)-(G/A)-(F/Y)-(V/I)-X-(F/Y) as a probe, we screened protein sequences carrying identical amino acids in an NBRF-protein database. It has been found that the C-terminal portion of clustered asparagine-rich protein (CARP), a malaria antigen from *Plasmodium falciparum*, shows an unexpected sequence similarity with the RNA-binding protein superfamily for the C-terminal half of the RNA-binding domain. Dot matrix comparisons and alignment of these sequences as well as a statistical test have revealed highly significant sequence similarities. From these analyses, we conclude that the malaria antigen CARP belongs to a large family of the RNA-binding proteins. An evolutionary implication of the sequence similarity was also discussed.

RNP motif; Malaria antigen; Homology; Evolution

1. INTRODUCTION

Several protein components of the ribonucleoprotein (RNP) particle, including heterogeneous nuclear RNPs (hnRNPs), small nuclear RNPs (snRNPs), poly(A) binding protein and nucleolin, comprise a large family of RNA-binding proteins that share RNA-binding domain of about 90 amino acids [1]. The family members contain a highly conserved stretch of amino acids, 'GFGF', the so-called RNP consensus motif, which is a hallmark characteristic of the RNA-binding domain [1,2]. Recently, members of the superfamily are increasingly found beyond the vertebrate RNPs in *Drosophila pen* [3], *elav* [4], *Sex lethal* (Sxl) [5], and *transformer-2* (*tra-2*) [6] gene products as well as in abscisic acid-induced glycine-rich protein from maize [7,8]. We report here tht a *Plasmodium falciparum* antigen CARP (clustered-asparagine-rich protein) [9] shows a significant similarity in sequence with the RNA-binding domains. Thus the CARP is a new member of the RNA-binding protein superfamily.

2. MATERIALS AND METHODS

The strategy for detecting sequences carrying a stretch of amino acids identical to that of a probe in a protein database is essentially the same as that described previously [10]. We used a stretch of amino acids G-(F/Y)-(G/A)-(F/Y)-(V/I)-X-(F/Y), known as the RNP consensus motif [1], including some variations, as a probe, where (F/Y), for example, means F or Y at the amino acid position and X represents any one of amino acids. Dot matrix comparison and sequence alignment between the CARP and the members of the RNA-binding protein family were carried out by the method described previously [11,12]. On the basis of the alignment, the probability Pr that the observed sequence similarity is realized by chance was estimated by the method described previously [11].

Correspondence address: T. Miyata, Department of Biology, Faculty of Science, Kyushu University, Fukuoka 812, Japan

3. RESULTS AND DISCUSSION

Using the highly conserved amino acids G-(F/Y)-(G/A)-(F/Y)-(V/I)-X-(F/Y) as a probe,we screened protein sequences in the NBRF protein database (release 20.0) as well as recently published sequences complied in our database. Twenty sequences containing a stretch of amino acids identical with the probe were detected. All these sequences are known members of the RNA-binding protein superfamily, except for the malaria antigen CARP containing clusters of asparagine residues from *Plasmodium falciparum* [9]. The CARP sequence was compared with sequences of all members of the RNA-binding protein superfamily by a dot matrix method [11,12]. Fig.1 shows the dot matrix comparison between the CARP and the poly(A)-binding protein (poly(A)-BP) [13]. This figure shows four diagonal lines representing strong sequence smilarities (indicated by arrows) for regions corresponding to the C-terminal half of the RNA-binding domain, where their sequences are strongly conserved among the members of the RNA-binding protein superfamily and the 'GFGF' motif exists [1]. Dot matrix comparisons with other members also show similar results.

Alignment of the CARP sequence with those of known members of this family was shown in fig.2. Per-

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/90/$3.50

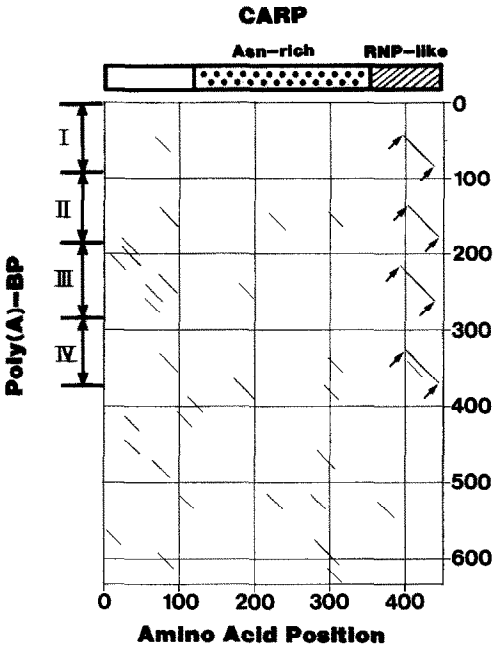

Fig.1. Dot matrix comparison between the poly(A)-binding protein (poly(A)-BP) and the malaria antigen CARP. Four diagonal lines showing significant sequence similarities between the two sequences are indicated by arrows. The four RNA-binding domains (I–IV) of the poly(A)-BP and the Asn-rich and RNP-related domains of the CARP are also shown.

cent amino acid identity $I(\%)$ of the CARP with each member of this family and the probability Pr that such sequence similarity is realized by chance are also shown in fig.2. Evidently the sequence similarities between the CARP and the members of the RNA-binding protein superfamily are highly significant, except for the second repeat unit of nucleolin and the second repeat unit of hnRNP A1 protein (fig.2). A weak sequence similarity is also found in the less conserved N-terminal half of the RNA-binding domain. In the CARP the highly conserved 'GFGF' motif is located at the precise position to those in the RNA-binding domain. Furthermore a unique structural feature of asparagine-rich domain in CARP also resembles to the simple structure of clustered amino acids often found at the N-terminal or C-terminal region of the family members (see references cited in fig.2). From these results, we conclude that the CARP is a new member of the RNA-binding protein superfamily and possibly has been derived from a member of this superfamily by gene duplication.

Strikingly, the malaria antigen CARP bears a structural resemblance to protein components of RNP. These two protein species differ from each other in both function and localization within the cell. Unlike many *P. falciparum* antigens known to date, the CARP lacks tandem repeats of antigen sequences. Instead, it contains a region rich in asparagine residues that are distributed in clusters rather than in repeats. Combining the fact that similar simple structures exist in some members of the RNA-binding family, it may be possible to speculate that a certain gene for RNA-binding protein with a simple structure of clustered amino acids has been used as an antigen during the evolution of an ancestral malaria species. There is another example of such unexpected sequence similarity: A malaria antigen

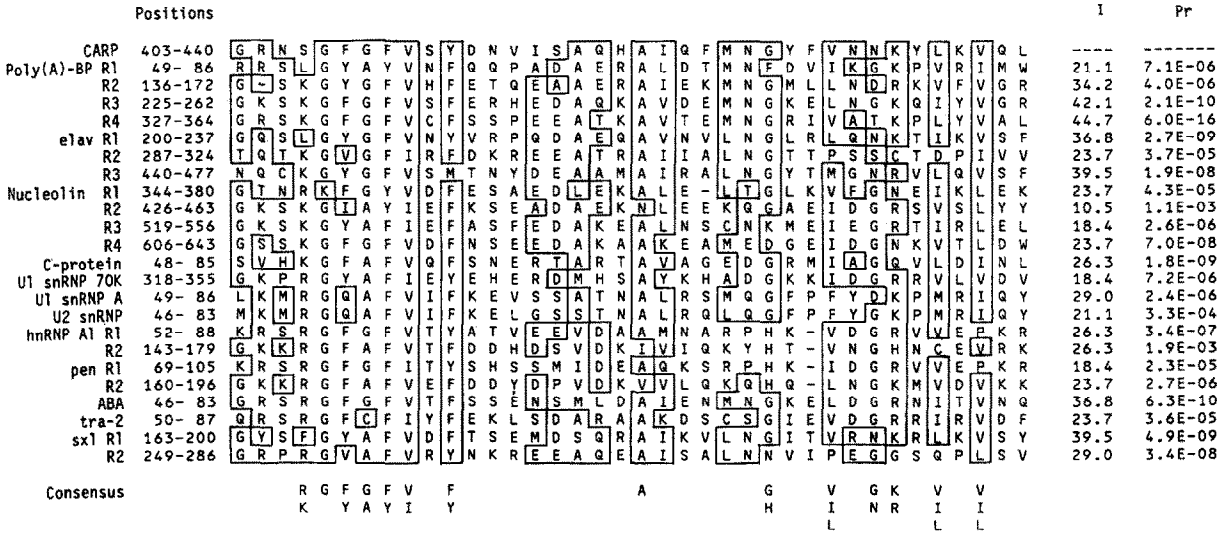

Fig.2. Alignment of the amino acid sequence of the CARP with those of the RNA-binding protein superfamily. Highly conserved amino acids are boxed and also are shown beneath the alignment. The amino acid positions of the aligned region are shown. Gaps (–) were introduced to increase the sequence similarity. I, the percent amino acid identity of CARP with each member of the superfamily. Pr, the probability that such similarity is realized by chance; the Pr value 7.1E–06, for example, means 7.1×10^{-6}. Poly(A)-BP R1, for example, represents the first repeat unit of the poly(A)-BP. Abbreviations of proteins and sequence data sources: CARP [9]; Poly(A)-BP, human poly(A)-binding protein [13]; *elav, Drosophila elav* gene product [4]; Nucleolin,from hamster [14]; C-protein, *Xenopus laevis* hnRNA C-protein [15]; U1 snRNP 70K, from human [16]; U1-associated A protein, from human [17]; U2 snRNP, from human [18]; hnRNP A1, from human [19]; *pen, Drosophila pen* gene product [3]; ABA, maize abscisic acid-induced glycine-rich protein [7]; *tra-2, Drosophila* sex-determining *tra-2* gene product [6]; *sxl, Drosophila sex-lethal* gene product [5].

Ag63 from *P. falciparum* share a close sequence similarity with the heat shock protein (*HSP70*) family [20], an enzyme carrying ATPase activity [21]. It may be interesting to know whether other malaria antigens also share sequence similarities with proteins of such distinct functions as enzymes and nuclear factors. Lens crystallins are unique structural proteins that have evolved by recruiting or duplicating already existing enzymes [22].

Acknowledgements: This work was supported in part by grants from the Ministry of Education, Science and Culture of Japan.

REFERENCES

[1] Dreyfuss, G., Swanson, S.S. and Pinol-Roma, S. (1988) Trends Biochem. Sci. 13, 86–91.
[2] Adam, S.A., Nakagawa, T., Swanson, M.S., Woodruff, T.K. and Dreyfuss, G. (1986) Mol. Cell. Biol. 6, 2932–2943.
[3] Haynes, S.R., Rebbert, M.L., Mozer, B.A., Forquignon, F. and Dawid, I.B. (1987) Proc. Natl. Acad. Sci. USA 84, 1819–1823.
[4] Robinow, S., Camos, A.R., Yao, K.-M. and White, K. (1988) Science 242, 1570–1572.
[5] Bell, L.R., Maine, E.M., Schedl, P. and Cline, T.W. (1988) Cell 55, 1037–1046.
[6] Amrein, H., Gorman, M. and Noethigher, R. (1988) Cell 55, 1025–1035.
[7] Gomez, J., Sanchez-Martines, D., Stiefel, V., J., Puigdomenech, P. and Pages, M. (1988) Nature 334, 262–264.
[8] Mortenson, E. and Dreyfuss, G. (1989) Nature 337, 312.
[9] Wahlgren, M., Aslund, L., Franzen, L., Sundvall, M., Wahlin, B., Berzins, K., McNical, L.A., Bjoerkman, A., Wigzell, H. Perlmann, P. and Pettersson, U. (1986) Proc. Natl. Acad. Sci. USA 83, 2677–2681.
[10] Hayashida, H., Kuma, K. and Miyata, T. (1988) Proc. Jpn. Acad. 64B, 113–118.
[11] Toh, H., Hayashida, H. and Miyata, T. (1983) Nature 305, 827–829.
[12] Miyata, T., Hayashida, H., Kikuno, R., Toh, H. and Kawade, Y. (1985) in: Interferon 6, (Groesser, I. ed.) pp.1–30, Academic Press, London.
[13] Grange, T., Martin de Sa, C., Oddos, J. and Pictet, R. (1987) Nucleic Acids Res. 15, 4771–4787.
[14] Lapeyre, B., Bourbon, H. and Amalric, F. (1987) Proc. Natl. Acad. Sci. USA 84, 1472–1476.
[15] Preugschat, F. and Wold, B. (1988) Proc. Natl. Acad. Sci. USA. 85, 9669–9673.
[16] Theissen, H., Etzerodt, M., Reuter, R., Schneider, C., Lottspeich, F., Argos, P., Luehrmann, R. and Philipson, L. (1986) EMBO, J. 5, 3209–3217.
[17] Sillekens, P.T.G., Habets, W.J., Beijer, R.P. and Venrooij, W.J. (1987) EMBO J. 6, 3841–3848.
[18] Lahiri, D.K. and Thomas, J.O. (1986) Nucleic Acids Res. 14, 4077–4094.
[19] Buvoli, M., Biamonti, G., Tsoulfas, P., Bassi, M.T., Ghetti, A., Riva, S. and Morandi, C. (1988) Nucleic Acids Res. 16, 3751–3770.
[20] Bianco, A.E., Favaloro, J.M., Burkot, T.R., Culvenor, J.G., Crewther, P.E., Brown, G.V., Anders, R.F., Coppel, R.L. and Kemp, D.J. (1986) Proc. Natl. Acad. Sci. USA 83, 8713–8717.
[21] Chappell, T.G. Welch, W.J., Schlossman, D.M., Palter, K.B., Schlesinger, M.J. and Rothman, J.E. (1986) Cell 45, 3–13.
[22] Wistow, G.J. and Piatigorsky, J. (1988) Annu. Rev. Biochem. 57, 479–504.