

Molecular cloning of PARP (proline/arginine-rich protein) from human cartilage and subsequent demonstration that PARP is a fragment of the NH₂-terminal domain of the collagen $\alpha 2(\text{XI})$ chain

Natalia I. Zhidkova, Randolph G. Brewton and Richard Mayne

Department of Cell Biology, University of Alabama at Birmingham, Birmingham, AL 35294, USA

Received 6 May 1993

We report the molecular cloning of a proline/arginine-rich protein (called PARP) from human cartilage using the polymerase chain reaction (PCR) and degenerate oligonucleotides based on the previously published amino acid sequence of bovine PARP [1]. Subsequently, a reverse transcription-polymerase chain reaction (RT-PCR) was performed with poly(A)-rich RNA from human cartilage using a sense oligonucleotide derived from PARP and an anti-sense oligonucleotide derived from the known sequence of the human collagen $\alpha 2(\text{XI})$ chain [2]. Nucleotide sequencing of the PCR product demonstrated that PARP is a fragment of the NH₂-terminal non-collagenous (NC3) domain of the collagen $\alpha 2(\text{XI})$ chain.

Cartilage; Human collagen type XI; Polymerase chain reaction; Nucleotide sequence

1. INTRODUCTION

In 1990, a proline/arginine-rich protein (called PARP) was isolated from bovine cartilage and the complete amino acid sequence determined [1]. PARP was found to be related to the NH₂-terminal domain of several collagen types, being most closely related (49% identity) to the NH₂-terminal domain of the collagen $\alpha 1(\text{XI})$ chain [1,3,4]. We now report the molecular cloning of human PARP using RT-PCR performed with degenerate oligonucleotide primers designed from the published amino acid sequence of bovine PARP. A subsequent RT-PCR reaction was performed with a sense primer derived from human PARP and an anti-sense primer derived from the published nucleotide sequence of the human $\alpha 2(\text{XI})$ collagen chain [2]. Cloning and nucleotide sequencing of the 1.86 kbp PCR product demonstrated that PARP is a fragment of the amino-terminal non-collagenous (NC3) domain of the collagen $\alpha 2(\text{XI})$ chain.

Correspondence address: R. Mayne, Department of Cell Biology, University of Alabama at Birmingham, Box 302-VH, Birmingham, AL 35294, USA. Fax: (1) (205) 934-7029.

Abbreviations: PCR, polymerase chain reaction; RT-PCR, reverse transcription-polymerase chain reaction; PARP, proline/arginine-rich protein.

The sequence described in this paper has been submitted to the GenBank/EMBL Data Library under the accession number L 18987.

2. MATERIALS AND METHODS

2.1. Isolation of RNA and first strand cDNA synthesis

Chondrocytes were isolated from sternal cartilage obtained from an eight-year-old white female following corrective surgery for pectus excavatum. Minced cartilage was incubated in digestion medium (MEM, 5% L-glutamine, 2% antibiotic/antimycotic containing 127 U/ml collagenase (Worthington Biochemical Corp., CLS1) overnight at 37°C with shaking. Chondrocytes were freed from the matrix by repeated pipetting, filtered through lens paper, and centrifuged to a pellet. The cells were resuspended in MEM supplemental with 10% fetal bovine serum, washed three times with MEM and the pellet subsequently solubilized in 4.0 M guanidine isothiocyanate, 0.1 M Tris-HCl, pH 7.4, 0.1 M 2-mercaptoethanol and 0.5% N-lauryl-sarcosine. Total RNA was obtained by the single step acid-guanidine method [5] and poly(A)-rich RNA selected by oligo(dT)-cellulose chromatography [6]. First strand cDNA for PCR analysis was prepared as described previously [7].

2.2. PCR conditions and cloning of PCR products

Conditions for all PCR reactions were: denaturation, 92°C, 1 min; annealing, 58°C, 1 min; polymerization, 72°C, 3 min; for 30 cycles as described [7]. After agarose gel electrophoresis, individual bands were recovered using GeneClean (BIO 101, LaJolla, CA) as described [7].

To clone PARP, RT-PCR was performed using degenerate 20-mer oligonucleotides designed from the bovine amino acid sequence. Sense and antisense primers were: 5'-GGIACIGCICCC(ACGT)GT(ACGT)-GA(CT)GT-3' and 5'-C(GT)CCAICCCICCC(CT)TC(AG)CA(CT)TC-3' (Primers 1 and 2; for the location of these primers see Fig. 1) in which redundancy was reduced by incorporation of inosine at several sites [8].

To obtain a PCR product bridging PARP and the $\alpha 2(\text{XI})$ chain the sense primer from PARP was 5'-CTCCTGACTCTCTACAGTGC-3' (Primer 3; Fig. 1) and the antisense primer was 5'-CTGACGTCCAG-GATAGCCAG-3' (Primer 4; positions 953–972 of the published nucleotide sequence [2]).

Products of PCR were cloned into the pCRII vector using the TA Cloning kit as described by the manufacturer (Invitrogen, San Diego, CA).

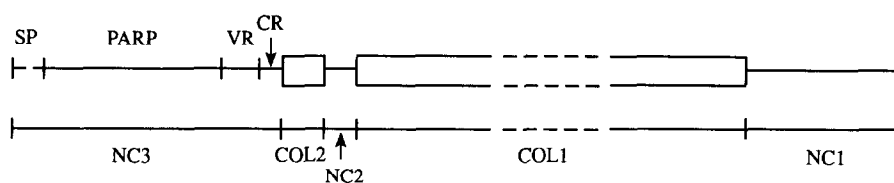
1	GGGACGGCGCCGGTGTGATGTG	GGG TAC CGA GTG GCA CGA CCT GCC CAG CTC AGT GCA CCC ACT CGC CAG CTT TTC CCA GGA	81
	1 →	A Y R V A R P A Q L S A P T R Q L F P G	
82	GGA TTT CCA AAA GAT TTT CCT CTG CTG ACT GTT GTC CGC ACC CGC CCT GGT CTC CGA GCT CCC CTC CTG ACT CTC TAC AGT	162	
	S	A	
163	GTC CAG GGT GTC CGA CAG CTG GGC CTG GAG CTG GGC CGA CCT GTC CGC TTC CTG TAT GAA GAC CAG ACA GGG CGG CCT CAA	243	
	A Q G V R Q L G L E L G R P V R F L Y E D Q T G R P Q		
244	CCT CCC TCT CAG CCA GTC TTC CGA GGC CTC AGC CTA GCA GAT GGC AAG TGG CAC CGT GTG GCT GTG GCT GTG AAG GGC CAG	324	
	P P S Q P V F R G L S L A D G K W H R V A V A V K G Q		
325	TCT GTC ACC CTC ATT GTT GAC TGC AAG AAG CGA GTC ACC CGG CCT CTC CCC CGA AGT GCT CGT CCA GTA TTG GAC ACC CAT	405	
	S V T L I V D C K K R V T R P L P R S A R P V L D T H		
406	GGA GTG ATC ATC TTT GGT GGC CGT ATT CTG GAT GAA GAA GTC TTT GAG GGT GAT GTC CAG GAG CTG GGC ATT GTC CCA GGG	486	
	G V I I F G A R I L D E E V F E G D V Q E L A I V P G		
487	GTC CAG GCA GCC TAT GAA TCA TGT GAA CAG AAG GAG CTG GAA TGC GAG GGG GGC CAG AGG GAA AGA CCC CAA AAC CAA CAG	567	
	V Q A A Y E S C E Q K E L E C E G G Q R E R P Q N Q Q		
568	CCT CAC AGA GCC CAG AGA TCT CCA CAG CAG CAA CCA TCA AGA CTT CAC AGG CCA CAA AAT CAG GAA CCC CAG AGC CAG GAC	648	
	P H R A Q * R S P Q Q Q P S R L H R P Q N Q E P Q S Q D		
649	CCC ACC CCA GGT GAA GAG GAA GAA ATC CTG GAG TCG AGC CTC TTG CCA CCC CTT GAG GAG GCT GCC CAT GGA CCC CGA GGG	729	
	P T P G E E E E I L E S S L L P P L E E A A H G P R G		
730	CTG AAG GGA GAG AAA GGA GAG CCT GCA GTG TTG GAA CCT GGT ATG CTC GTG GAG GGG CCC CCT GGC CCA GAA GGG CCT GCG	810	
	L K G E K G E P A V L E P G M L V E G P P G P E G P A		
811	GGA TTG ATT GGT CCC CCT GGC ATC CAG GGG AAC CCA GGC CCA GTT GGA GAC CCT GGA GAG AGG GGC CCC CCT GGC CGA GCA	891	
	G L I G P P G I Q G N P G P V G D P G E R G P P G R A		
892	GGG CTC CCT GGA TCA GAT GGG GCT OCT GGT OCT OCT GGC ACA TCT CTC ATG CTC CCA TTC CGG TTT GGC AGT GGT GGG GGT	972	
	G L P G S D G A P G P P G T S L M L P F R P G S G G G		
973	GAC AAG GGC CCT GTG GTG GGC GCC CAG GAG GCT CAG GCC CAG GCG ATC CTG CAG CAG GCG AGG CTG GCG CTC CGT GGA CCC	1053	
	D K G P V V A A Q E A Q A Q A I L Q Q A R L A L R G P		
1054	CCT GGC CCC ATG GGA TAC ACA GGG CGC CCT GGA CCC TTG GGC CAA CCT GGG AGC CCT GGC CTG AAA GGA GAG TCT GGA GAC	1134	
	P G P M G Y T G R P G P L G Q P G S P G L K G E S G D		
1135	TTA GGA CCT CAG GGC CCC AGA GGA CCT CAG GGC CTC ACA GGC TCC CTG GGC AAG GCT GGG CGA AGG GGC CGG GCA GGT CCT	1215	
	L G P Q G P R G P Q G L T G S L G K A G R R G R A G P		
1216	GAT GGA GCC CGA GGG ACC CTG GGA GAT CCT GGA GTG AAG GGT GAC CGA GGT TTT GAT GGA CTC CCA GGG CTC CCT GGA GAG	1296	
	D G A R G T L G D P G V K G D R G F D G L P G L P G E		
1297	AAG GGC CAT AGG GGT GAT ACT GGT CCC AGG GGC CTT CCT GGT CCC CCT GGT GAG GAT GGA GAG AGG GGA GAT GAC GGG GAG	1377	
	K G H R G D T G P R G L P G P P G E D G E R G D D G E		
1378	ATT GGG CCT CGA GGG CTG CCT GGA GAG TCG GGA CCT CGA GGT CTC CTT GGC CCC AAA GGC CCA CCT GGT ATT CCT GGA CCC	1458	
	I G P R G L P G E S G P R G L L G P K G P P G I P G P		
1459	CCT GGC GTC CGA GGC ATG GAT GGT CCC CAG GGC CCC AAA GGG AGC TTG GGA CCC CAG GGA GAG CCA GGA CCT CCT GGA CAA	1539	
	P G V R G M D G P Q G P K G S L G P Q G E P G P P G Q		
1540	CAG GGC ACC CCT GGG ACC CAG GGT CTT CCC GGG CCC CAG GGC GCC ATC GGC CCT CAT GGA GAG AAG GGT CCT CAA GGG AAG	1620	
	Q G T P G T Q G L P G P Q G A I G P H G E K G P Q G K		
1621	CCA GGG CTC CCC GGC ATG CCT GGC TCA GAC GGA CCC CGG GGT CAC CCA GGG AAG GAA GGT CCC CCT GGA ACC AAA GGG AAA	1701	
	P G L P G M P G S D G P P G H P G K E G P P G T K G K		
1702	CCA GGT CCC TCT GGA CCT CAG GGA CCT CTA GGA TAC CCA GGA CCT CAA GGG GTC AAG GGT GTG AAC GGA ATT CGG GGT CTG	1782	
	P G P S G P Q G P L G Y P G P Q G V K G V N G I R G L		
1783	AAG GGT CAT AAG GGT GAG AAG GGT GAG GAT GGC TTT CCT GGG TTC AAA GGT GAC ATA GGC GTG AAA GGT GAC AGG GGC GAA	1863	
	K G H K G E K G E D G F P G F K G D I G V K G D R G E		
1864	GTT GGA GTC CCT GGT TCC AGG GGA GAG GAT GGT CCT GAG GGG CCA AAG GGA CGC ACT GGA CCG ACT GGA GAC CCT GGG CCC	1944	
	V G V P G S R G E D G P E G P K G R T G P T G D P G P		
1945	CCA GGG CTC ATG GGC GAG AAG GGC AAG CTG GGT GTT CCT GGT CTG CTT GGC TAT CCT GGA GGT CAG 2010		
	P G L M G E K G K L G V P G L P G Y P G R Q		
	← 4		

NC3

COL2

NC2

COL1



Total amino acids/domain (from references 2, 9-12).

	NC3				COL2	NC2	COL1	NC1
	SP	PARP	VR	CR				
$\alpha 2(\text{XI})$	-	218	45	22	51	37	1014	190
$\alpha 1(\text{XI})$	36	225	159	22	51	36	1014	264
$\alpha 1(\text{V})$	36	225	185	22	51	39	1014	266

SP = signal peptide; PARP = proline/arginine rich peptide; VR = variable region; CR = constant region

Fig. 2. A model for the organization of the domains and sub-domains of the $\alpha 2(\text{XI})$, $\alpha 1(\text{XI})$ and $\alpha 1(\text{V})$ chains.

2.3. DNA sequencing

Dideoxynucleotide sequencing was performed on either single or double stranded DNA obtained after cloning PCR products in the pCRII vector. For PARP clones, Sequenase 2.0 (US Biochemicals) was used and for PARP- $\alpha 2(\text{XI})$ bridging clones Taq polymerase was used (TAQuence 2.0, US Biochemicals). Nucleotide sequences of PARP were obtained from a single clone (NZ1), and for PARP- $\alpha 2(\text{XI})$ bridging clones two independent clones (called NZ2 and NZ3) were sequenced.

3. RESULTS

In order to clone PARP, RT-PCR was initially performed with degenerate primers designed from the protein sequence of bovine PARP. The location of these primers (called Primer 1 and 2) and the nucleotide sequence for a clone of the PCR product called NZ1 is shown in Fig. 1. Primer 1 was a degenerate primer designed from the amino terminal sequence of bovine PARP (GTAPVDV [1]). However, subsequent nucleotide sequencing of NZ1 demonstrated that priming had occurred 24 amino acids further downstream at the sequence GICPADV.

To obtain a clone bridging between PARP and the $\alpha 2(\text{XI})$ chain, RT-PCR was performed using sense and anti-sense primers located at sites 3 and 4 in Fig. 1. The nucleotide sequence of this clone (called NZ2) is also shown in Fig. 1, together with a comparison of the derived protein sequence of human PARP with bovine

PARP. Close agreement was observed with 17 largely conservative differences, demonstrating the successful cloning of human PARP. In addition, complete agreement was found for the nucleotide sequence at the 3' end of NZ2 with the published nucleotide sequence of human $\alpha 2(\text{XI})$ in the same region [2]. This result therefore establishes that PARP is derived from the amino-terminal non-collagenous (NC3) domain of the $\alpha 2(\text{XI})$ chain.

With the sequence of the amino terminus of the $\alpha 2(\text{XI})$ chain (Fig. 1) and the complete sequences of the $\alpha 1(\text{XI})$ and $\alpha 1(\text{V})$ chains now available [9,10-12], it is possible to recognize several different domains and a model labeling these structures is shown in Fig. 2. In the same manner as several other collagens, the domains are numbered from the 3' end so that the major collagenous domain is designated COL1. A second shorter collagenous domain is labeled COL2 and the NC3 domain is further subdivided into four subdomains. In several previous studies homologies between the sequence of PARP and the $\alpha 1(\text{XI})$, $\alpha 1(\text{V})$, $\alpha 1(\text{IX})$, $\alpha 1(\text{XII})$, and $\alpha 1(\text{XIV})$ chains were noted suggesting a common ancestral origin for this structure [1,3,4].

The NC3-PARP subdomain is followed by the NC3-variable region (NC3-VR) which is much shorter for $\alpha 2(\text{XI})$ than for $\alpha 1(\text{XI})$ and $\alpha 1(\text{V})$ (Fig. 2). Previous comparison of this region for the $\alpha 1(\text{XI})$ and $\alpha 1(\text{V})$ chains showed very little homology [12], and this region

←

Fig. 1. The nucleotide sequence and deduced amino sequence of PCR-derived clones NZ1 and NZ2 which bridge from PARP to the COL1 domain of $\alpha 2(\text{XI})$. Sites 1 and 2 are the location of degenerate primers used initially to obtain clone NZ1. Sites 3 and 4 represent the location of primers used to obtain clone NZ2 which bridges from PARP to $\alpha 2(\text{XI})$. The collagenous domains COL1 and COL2 are shown by boxed sequences between which is the NC2 domain. The carboxyl-terminus of PARP is designated by, and differences between the amino acid sequence of bovine and human PARP are shown on the third line for this sub-domain. The sequence from * to ● represents the variable region of the NC3 domain (VR-NC3) and the short sequence after ● to COL2 represents the constant region of NC3 (CR-NC3). The conserved crosslinking site (KGHR) in COL1 is indicated by ▲. The start of the published nucleotide sequence of the human $\alpha 2(\text{XI})$ chain is indicated by †.

for $\alpha 2(XI)$ did not appear closely related to either the $\alpha 1(XI)$ or $\alpha 1(V)$ sequences. The remaining NC3-constant region (NC3-CR), COL2, NC2 and COL1 domains all showed a high degree of homology between the $\alpha 2(XI)$, $\alpha 1(XI)$ and $\alpha 1(V)$ chains.

4. DISCUSSION

The successful cloning of human PARP and its origin from the $\alpha 2(XI)$ chain allow a comparison to be made with other chains of the type V/type XI family. The results show close homologies with the $\alpha 2(XI)$ and $\alpha 1(V)$ chains but marked differences from the $\alpha 2(V)$ chain as noted previously [13]. However, at one subdomain (called NC3-VR) marked differences were observed for the $\alpha 2(XI)$, $\alpha 1(XI)$ and $\alpha 1(V)$ chains both in length and in deduced amino acid sequence. This domain is, however, acidic for all three chains and does not contain cysteine residues. It appears that the NC3 domain may form an extended structure that cannot be visualized after rotary shadowing [14]. With the availability of the amino-terminal sequences of the $\alpha 1(XI)$ and $\alpha 2(XI)$ chains, it will now be possible to analyze the processing events that occur before the final tissue forms of these chains are achieved [14].

Acknowledgements: This study was supported by Research Grants AR 30481 and HD 20691 from the National Institutes of Health. The human tissues utilized in this study were obtained through the UAB Cooperative Human Tissue Network which is funded by the National Cancer Institute.

REFERENCES

- [1] Neame, P.J., Young, C.N. and Treep, J.T. (1990) *J. Biol. Chem.* 265, 20401-20408.
- [2] Kimura, T., Cheah, K.S.E., Chan, S.D.H., Lui, V.C.H., Mattei, M.-G., van der Rest, M., Ono, K., Solomon, E., Ninomiya, Y. and Olsen, B.R. (1989) *J. Biol. Chem.* 264, 13910-13916.
- [3] Wälchli, C., Trueb, J., Kessler, B., Winterhalter, K.H. and Trueb, B. (1993) *Eur. J. Biochem.* 212, 483-490.
- [4] Bork, P. (1992) *FEBS Lett.* 307, 49-54.
- [5] Chomczynski, P., and Sacchi, N. (1987) *Anal. Biochem.* 162, 156-159.
- [6] Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. (1979) *Biochemistry* 18, 5294-5299.
- [7] Brewton, R.G., Ouspenskaia, M.V., van der Rest, M. and Mayne, R. (1992) *Eur. J. Biochem.* 205, 443-449.
- [8] Knuth, K., Roberds, S., Poteet, C. and Tamkun, M. (1988) *Nucleic Acids Res.* 16, 10932.
- [9] Yoshioka, H. and Ramirez, F. (1990) *J. Biol. Chem.* 265, 6423-6426.
- [10] Bernard, M., Yoshioka, H., Rodriguez, E., van der Rest, M., Kimura, T., Ninomiya, Y., Olsen, B.R. and Ramirez, F. (1988) *J. Biol. Chem.* 263, 17159-17166.
- [11] Takahara, K., Sato, Y., Okazawa, K., Okamoto, N., Noda, A., Yaoi, Y. and Kato, I. (1991) *J. Biol. Chem.* 266, 13124-13129.
- [12] Greenspan, D.S., Cheng, W. and Hoffman, G.G. (1991) *J. Biol. Chem.* 266, 24727-24733.
- [13] Petit, B., Freyria, A.-M., van der Rest, M. and Herbage, D. (1992) in: *Biological Regulation of Chondrocytes* (Adolphe, M. Ed.) pp. 33-84, CRC Press, Boca Raton.
- [14] Thom, J.R. and Morris, N.P. (1991) *J. Biol. Chem.* 266, 7262-7269.