

A Minisatellite Sequence within the Propeptide Region of the Vacuolar Carboxypeptidase Y Gene of *Schizosaccharomyces pombe*

SUSHAM S. INGAVALE, RUPINDER KAUR, PARUL AGGARWAL, AND ANAND K. BACHHAWAT*

Institute of Microbial Technology, Chandigarh—160 036, India

Received 11 March 1998/Accepted 7 May 1998

We describe the presence of a minisatellite sequence that displays length polymorphisms in the fission yeast *Schizosaccharomyces pombe*. The minisatellite sequence was found to reside within the propeptide region of the vacuolar carboxypeptidase Y gene. The minisatellite sequence, which was found only at a single locus, was mitotically stable and displayed length polymorphisms between the two varieties of *S. pombe* (*S. pombe* var. *pombe* and *S. pombe* var. *malidevorans*). The minisatellite sequence, however, appeared to be species specific and was absent in other members of the *Schizosaccharomyces* genus. This report constitutes the first experimental demonstration of the presence of such sequences in yeasts.

Minisatellite sequences are found widely dispersed in the genomes of a variety of higher eukaryotic organisms. These minisatellite DNA sequences, also referred to as variable-number tandem repeats (VNTR), are comprised of multiple copies of repeats whose base units are in tens of nucleotides (9 to 60 bp) and are found within the genome in variable numbers. It is this variability which in fact forms the basis of DNA fingerprinting and other genomic analyses (5, 12). In yeasts, however, although length polymorphisms due to microsatellite DNA (repeat units of 1 to 6 bp) have been described (8, 16), length polymorphisms due to naturally occurring nonteleomeric minisatellite DNA have not been observed. In this report, we experimentally demonstrate the existence of such sequences in yeasts.

The vacuolar carboxypeptidase Y (CpY) gene (*cpy1⁺*) of *Schizosaccharomyces pombe* has been cloned independently by several groups. Tabuchi et al. reported a protein of 950 amino acids (accession no. D86560) (19). In contrast to this report however, a protein of 1,002 amino acids was observed from the sequence obtained at the Sanger Centre, Cambridge, United Kingdom, as part of the *S. pombe* genome sequencing project (accession no. D97209), as well as from our own laboratory, where we have cloned the gene by complementation of a *cpy*-deficient mutant (1).

The CpY protein of *S. pombe* is characterized by an unusually long propeptide (554 amino acids in contrast to 91 amino acids for that of *Saccharomyces cerevisiae* [20] and 106 amino acids for that of *Candida albicans* [11]) that contains two distinct repeat regions (Fig. 1). The first repeat unit is 13 amino acids long and is repeated 11 times (7 times in the sequence reported by Tabuchi et al. [19]), and the second repeat unit is 9 amino acids long and is repeated 8 times (Tabuchi et al. [19] have described this as seven repeats, opting not to refer to the eighth unit as one of the repeats; there are, however, no discrepancies in the sequences of this region). The repeat regions fail to display significant homology to anything in the GenBank and EMBL databases by comparison using the BLAST program (1a). When we examined the nucleotide sequences of the

repeat regions we observed only a few differences, even at the nucleotide level, between the repeat units (Fig. 2). This suggested to us that we might be dealing with a minisatellite sequence. We therefore set about to determine whether these two repeat regions indeed corresponded to minisatellite DNA that could display variable numbers of tandem repeats and to examine whether this region was mitotically stable.

Minisatellite analysis was carried out by PCR on different *S. pombe* strains using primers flanking the repeat regions (Fig. 2). We initially used *S. pombe* strains obtained from different laboratories, all of which have their ancestral origin in the isolate of *S. pombe* that was first investigated by Leupold (9).

When PCR was carried out with primers P1 and P2, which flank repeat sequence I (RS I), we were able to detect a band whose size corresponded to the expected 11 repeats. The same size band was observed in all the *S. pombe* strains that we tested (Fig. 3a, lanes 1 and 4 to 8) and was also identical in size to that seen when the pCpY1 plasmid was used as a template (Fig. 3a, lane 11).

The PCR results with RS II and primers P3 and P4 revealed a band that was detected in all the *S. pombe* strains that we tested and corresponded in size to the expected eight repeat units (Fig. 3b, lanes 1 and 4 to 8) and was also identical in size to the bands seen with the plasmid control (Fig. 3a, lane 11).

These results indicated that the two repeat regions were mitotically stable and could possibly be used as genetic identification markers. We therefore decided to examine other isolates of *S. pombe*. The *S. pombe* species has been subdivided into two varieties: *S. pombe* var. *pombe* (which is normally referred to as *S. pombe*) and *S. pombe* var. *malidevorans* (which was previously referred to as *S. malidevorans* but has subsequently been found to be a variety of *S. pombe*) (18). We examined two such isolates of *S. pombe* by the PCR analysis described above. In the case of RS I, an identical band that corresponded to 11 repeats was observed in both strains (Fig. 3a, lanes 9 and 10). However, in the case of RS II, bands migrating faster than would bands of the expected size were observed (Fig. 3b, lanes 9 and 10), indicating a loss of some repeat units. This was also confirmed by Southern blotting (data not shown). To determine exactly how many and which of the repeat units were lacking, we cloned the smaller RS II PCR product from both the above-mentioned *S. pombe* var. *malidevorans* isolates and sequenced this region. A loss in one

* Corresponding author. Mailing address: Institute of Microbial Technology, Sector 39-A, Chandigarh—160 036, India. Phone: 0172-690004. Fax: 0172-690585 or 0172-690632. E-mail: abachhawatt@usa.net.

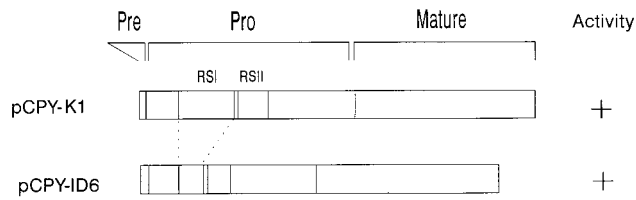


FIG. 1. Schematic representation of the *S. pombe* CpY protein indicating the different regions. pCPY-K1 is a *Kpn*I subclone that contains the *cpy1*⁺ open reading frame, and pCPY-ID6 is a derivative of pCPY-K1 that contains an in-frame deletion of six repeat units in RS I. The diagram has been drawn to scale (1.0 cm = 195 amino acids). Pre, prepeptide region; Pro, propeptide region.

of the internal repeat units, cII (Fig. 2), was observed. Preferential variation in the number of internal repeat units as opposed to variation in the number of the flanking repeats is a feature quite typical of VNTR loci (4).

Although Tabuchi et al. have reported a functional protein that differed in that it contained only seven repeats in RS I (19), we observed no length polymorphism in this region among the strains that we tested. However, despite their mitotic stability VNTR loci do undergo changes in their lengths at low frequencies, and to investigate this point further, we examined whether deletion of a few of the repeat units within this region would still lead to a functional protein. We therefore constructed pCPY-ID6, an in-frame CpY deletion mutant lacking six of the repeat units of RS I (Fig. 1). This was constructed by partial digestion with *Bst*XI followed by religation. The deletion was confirmed by sequencing. This construct was transformed into the *cpy1-1* mutant (10), and the transformants were found to exhibit CpY activity similar to the parent plasmid pCPY-K1, as seen by a CpY colorimetric plate assay (6) (data not shown). This indicated that the CpY enzyme can accommodate length polymorphism in RS I. It is

REPEAT SEQUENCE I

```

P1 → ...
600 CAT CAC GAG CCT GGT GAG CAC ATG CCT CCT CCT CTG 638 aI
639 CAC CAC AAG CCC GGT GAG CAC ATG CCT CCT CCT CTG 677 bI
678 CAC CAC GAA CCT GGA GAG CAC ATG CCT CCT CCT CTG 716 cI
717 CAC CAC GAA CCT GGA GAG CAC ATG CCT CCT CCT CTG 755 dI
756 CAC CAC GAA CCT GGA GAG CAC ATG CCT CCT CCT CTG 794 eI
795 CAC CAC GAA CCT GGA GAG CAC ATG CCT CCT CCT CTG 833 fI
834 CAT CAC GAG CCT GGA GAG CAC ATG CCT CCT CCT CTG 872 aII
873 CAC CAC GAA CCT GGA GAG CAC ATG CCT CCT CCT CTG 911 cII
912 CAC CAC GAA CCT GGA GAG CAC ATG CCT CCT CCT CTG 950 eII
951 CAT CAC GAG CCT GGA GAG CAC ATG CCT CCT CCT CTG 989 aI
990 CAT CAC GAG CCT GGA GAG CAC ATG CCT CCT CCT CTG 1028 dI
... ← P2

```

REPEAT SEQUENCE II

```

P3 → ...
1065 GAG CAT CAT CCT GGA CCT GAG GAC AAG 1091 aII
1092 GAA CAC CAT AAG GGA CCC AAG GAT AAG 1118 bII
1119 GAG CAC CAT AAG GGA CCT AAG GAT AAG 1145 cII
1146 GAG CAC CAT AAG GGA CCT AAG GAT AAG 1172 dII
1173 GAG CAC CAT AAG GGA CCT AAG GAT AAG 1199 eII
1200 GAG CAC CAC AAG GGA CCT AAG GAC AAG 1226 fII
1227 GAG CAC CAC AAG GGA CCT AAG GAC AAG 1253 gII
1254 GAG CAC CAT CAA GGA CCT AAG GAG AAG 1280
... ← P4

```

FIG. 2. Alignment of nucleotide sequences of the tandem repeat units of RS I and RS II that lie within the CpY propeptide region of *S. pombe* ABP13 *h*⁻ *ade6-210 ura4-D18* (obtained from Fission Yeast Course, Cold Spring Harbor, N.Y.). Primers P1 (5'-TTGATGACGACGCTCCCAAGC-3'), P2 (5'-CCTTCA TGCTCTCAAGCTCATGG-3'), P3 (5'-CCATGAGCTTGAGGAGCATGA AGG-3'), and P4 (5'-TTGCATGTTTTGCTCAGGAGCTTC-3') used for PCR analysis of the repeat regions are also indicated schematically. The code represents the code for each of the repeat units and indicates a specific sequence. The *Bst*XI sites at which in the in-frame deletion pCPY-ID6 was constructed are underlined.

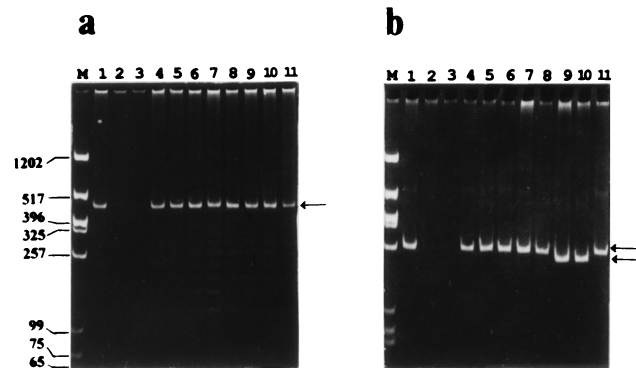


FIG. 3. PCR analysis of RS I (a) and RS II (b) regions. PCRs were carried out (95°C for 1 min, 60°C for 1 min, and 72°C for 1 min for 30 cycles), and the PCR products were run on 5% acrylamide gels. P1 and P2 were used to amplify RS I; P3 and P4 were used to amplify RS II. Genomic DNA isolated from the different yeast strains by the glass bead method (7) was used as the template for the PCRs. For marker DNA (lanes M) we used a *Hin*II/*Xba*I digest of pBSK(-). Lanes: 1, *S. pombe* ABP13 *h*⁻ *ade6-210 ura4-D18* (obtained from Fission Yeast Course); 2, *S. japonicus* var. *versatilis* (obtained from Fission Yeast Course); 3, *S. octosporus* (obtained from Fission Yeast Course); 4, *S. pombe* KN1 *h*⁻ *ade6-210 ura4-D18 leu1-32 pmd1::ura4*⁺ (13); 5, *S. pombe* ABP 320 *h*⁺ *arg3-D4 ura4-D18* (21); 6, *S. pombe* GP 25 *h*⁻ *his5-303 lys1-131* (3); 7, *S. pombe* *h*⁹⁰ *swi4-1* NCYC 1992; 8, *S. pombe* NCYC 1990; 9, *S. pombe* var. *malidevorans* NCYC 2387; 10, *S. malidevorans* NCYC 683; and 11, plasmid pCPY1 containing the original *cpy1*⁺ clone that was isolated in this laboratory (1). Arrows indicate the repeat regions. Molecular sizes of marker DNA are given (in base pairs) to the left of panel a.

possible, therefore, that RS I represents a second minisatellite sequence within this region even though we were not able to observe any polymorphism among the strains that we tested.

We extended the minisatellite analysis to other members of the *Schizosaccharomyces* genus, namely, *S. japonicus* and *S. octosporus*, as we felt it might reflect on the origin of these repeats. However, we failed to detect any bands corresponding to these regions upon PCR (Fig. 3, lanes 2 and 3). This was also confirmed by Southern blotting, by which we failed to observe any bands when RS I and RS II were used as probes (data not shown). This indicated that the repeats were species specific and did not occur in other members of the *Schizosaccharomyces* genus.

These studies provide a demonstration of the existence of a minisatellite DNA sequence that exhibits the property of possessing a variable number of tandem repeats in yeasts. The limited amount of intergenic and noncoding regions in yeasts may be responsible for the relative lack of these sequences in these unicellular eukaryotes. In retrospect, therefore, it is perhaps not surprising that a minisatellite DNA sequence should be detected in the propeptide region of a precursor protease such as CpY. Although these proregions play an essential role in the maturation of the proteases functioning as intramolecular chaperones, there is a very low level of conservation among the proregions of CpYp. Furthermore, structure-function analysis of the proregion indicates a great deal of structural flexibility (14, 15). It is unclear at this stage whether the minisatellite sequence of *S. pombe* also has some functional significance, since Shinde and coworkers recently reported that an alteration in the proregion can alter the folding and kinetic properties of the subtilisin enzyme (17).

While the results that we have reported relate to the fission yeast, it is likely that such sequences exist in other yeasts as well. A recent report described the presence of a putative minisatellite sequence in *Saccharomyces carlsbergensis* based on sequence comparisons with other minisatellite regions (2). Our results, while demonstrating that such sequences do exist

in yeasts, also show that they display the property of possessing a variable number of tandem repeats that perhaps are variety specific. These sequences could well provide a useful system for examining how such polymorphisms occur and originate, in addition to possibly adding another dimension to DNA fingerprinting and taxonomic analyses in yeasts and other unicellular eukaryotes.

S.S.I. and R.K. contributed equally to this work.

We thank A. Carr for the genomic libraries and M. Yoshida, S. Wadell, and G. Cottarel for the strains. We thank Raj Kumar for technical assistance and K. Ganesan for helpful discussions.

This work was supported in part by a grant-in-aid (project BT/R&D/15/40/93) from the Department of Biotechnology, Government of India. S.S.I., R.K., and P.A. are Senior Research Fellows of the Council of Scientific & Industrial Research, Government of India.

REFERENCES

- Aggarwal, P., and A. K. Bachhawat. Unpublished observations.
- Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**:403–410.
- Andersen, T. H., and T. A. Nilsson-Tillgren. 1997. A fungal minisatellite. *Nature* **386**:771.
- Cottarel, G. 1995. The *Saccharomyces cerevisiae* *HIS3* and *LYS2* genes complement the *Schizosaccharomyces pombe* *his5-303* and *lys1-131* mutations respectively: new selectable markers and new multipurpose, multicopy shuttle vectors pSP3 and pSP4. *Curr. Genet.* **28**:380–383.
- Gray, I. C., and A. J. Jeffreys. 1991. Evolutionary transience of hypervariable minisatellite in man and the primates. *Proc. R. Soc. Lond. Ser. B* **243**:241–253.
- Jeffreys, A. J., V. Wilson, and S. L. Thein. 1985. Hypervariable 'minisatellite' regions in human DNA. *Nature* **314**:67–73.
- Jones, E. W. 1991. Tackling the protease problem in yeast. *Methods Enzymol.* **194**:428–453.
- Kaiser, C., S. Michaelis, and A. Mitchell. 1994. *Methods in yeast genetics: a laboratory course manual*, 1994 edition. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- Kashi, Y., D. King, and M. Sella. 1997. Simple sequence repeats as a source of genetic variation. *Trends Genet.* **13**:74–78.
- Leupold, U. 1950. Die Vererbung von Homothallie und Heterothallie bei *Schizosaccharomyces pombe*. *C. R. Trav. Lab. Carlsberg Ser. Physiol.* **24**:381–480.
- Moreno, S., A. Klar, and P. Nurse. 1991. Molecular genetic analysis of fission yeast *Schizosaccharomyces pombe*. *Methods Enzymol.* **194**:795–823.
- Mukhtar, M., D. A. Logan, and N. F. Kaufer. 1992. The carboxypeptidase Y-encoding gene from *Candida albicans* and its transcription during yeast-to-hyphae conversion. *Gene* **121**:173–177.
- Nakamura, Y., et al. 1987. Variable number of tandem repeat (VNTR) markers for human gene mapping. *Science* **235**:1616–1622.
- Nishi, K., M. Yoshida, M. Nishimura, M. Nishikawa, M. Nishiyama, S. Horinouchi, and T. Beppu. 1992. Leptomycin B resistance gene of *Schizosaccharomyces pombe* encodes a protein similar to the mammalian P-glycoprotein. *Mol. Microbiol.* **6**:761–769.
- Ramos, C., J. R. Winther, and M. C. Kielland-Brandt. 1994. Requirement of the pro-peptide for in vivo formation of active carboxypeptidase Y. *J. Biol. Chem.* **269**:7006–7012.
- Ramos, C., and J. R. Winther. 1996. Exchange of regions of the carboxypeptidase Y propeptide. *Eur. J. Biochem.* **242**:29–35.
- Richard, G.-F., and B. Dujon. 1996. Distribution and variability of trinucleotide repeats in the genome of the yeast *Saccharomyces cerevisiae*. *Gene* **174**:165–174.
- Shinde, U., J. J. Liu, and M. Inouye. 1997. Protein memory through altered folding mediated by intramolecular chaperones. *Nature* **386**:771–773.
- Sipiczki, M., J. Kucsera, S. Ulaszewski, and J. Zsolt. 1982. Hybridization studies by crossing and protoplast fusion within the genus *Schizosaccharomyces* Lindner. *J. Gen. Microbiol.* **128**:1989–2000.
- Tabuchi, M., D. Iwaihara, Y. Ohtani, N. Ohuchi, J.-I. Sakurai, T. Morita, S. Iwahara, and K. Takegawa. 1997. Vacuolar protein sorting in fission yeast: cloning, biosynthesis, transport, and processing of carboxypeptidase Y from *Schizosaccharomyces pombe*. *J. Bacteriol.* **179**:4179–4189.
- Valls, L. A., C. P. Hunter, J. H. Rothman, and T. H. Stevens. 1987. Protein sorting in yeast: the localization of the yeast vacuolar carboxypeptidase Y resides in the pro-peptide. *Cell* **48**:887–897.
- Wadell, S., and J. R. Jenkins. 1995. *arg3⁺*, a new selectable marker system for use in *Schizosaccharomyces pombe*: application of *ura4⁺* as a removable selectable marker. *Nucleic Acids Res.* **23**:1836–1837.