Vol. 163, No. 1, 1989 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
August 30, 1989 Pages 72-78

THE NUCLEOTIDE SEQUENCE OF A HUMAN cDNA ENCODING THE HIGHLY CONSERVED NUCLEOLAR PHOSPHOPROTEIN B23

Xiaozhou Li, L. Jane McNeilage* and Senga Whittingham

The Burnet Clinical Research Unit, The Walter and Eliza Hall Institute of Medical Research and The Royal Melbourne Hospital, Post Office, Royal Melbourne Hospital, Victoria 3050, Australia

Received July 13, 1989

SUMMARY: A cDNA clone containing the complete coding sequence for the human nucleolar phosphoprotein B23 was isolated from a Burkitt's lymphoma cDNA library by immunoscreening with human autoantibodies. The B23 clone contained a 1.3 kb cDNA insert encoding a polypeptide of 294 amino acids with a predicted molecular mass of 32,539 daltons. The deduced B23 amino acid sequence contained 2 acidic domains rich in aspartic and glutamic acid, a feature shared by a number of nuclear and nucleolar proteins. The human B23 amino acid sequence showed 98% homology with rat B23 and 68% homology with the Xenopus laevis nucleolar phosphoprotein, NO38 showing that the primary structure of B23 is highly conserved among these species.

B23, like nucleolin (or C23), is a major nucleolar phosphoprotein expressed in higher amounts in tumor and growing cells than in resting cells (1). Its function is unknown although several lines of evidence suggest that it is involved in ribosome assembly and/or rRNA processing (1-3). Treatment of cells with actinomycin D and other inhibitors of rRNA synthesis (2-3) or the growth of cells in serum-free medium (2) results in the translocation of B23 from the nucleolus to the nucleoplasm. Subsequent re-entry into the nucleolus is dependent upon rRNA

* To whom correspondence should be addressed.

0006-291X/89 $1.50

synthesis but not upon protein synthesis (2-3), indicating that normal rRNA processing is essential for the localization of B23 to the nucleolus. Recently, B23 has been shown to be identical with the nuclear matrix protein, numatrin (4), the synthesis of which is induced in the early stages of cell mitogenesis (4-5).

This article reports the isolation of a 1.3kb cDNA encoding human B23 using a serum containing a high titer antinucleolar autoantibody. Analysis of the deduced amino acid sequence of the B23 cDNA identified several features that have been conserved between human (6,7), rat (8) and X.laevis (9) and are shared by nucleolin (10) and nucleoplasmin (11).

MATERIALS AND METHODS

Cloning. Serum from a patient (Ba) with high titer (1:10,000) antinucleolar autoantibodies reactive by immunoblotting with a 37kD HeLa cell nucleolar polypeptide was used, at a dilution of 1:1,000, to screen (12) a human Burkitt's lymphoma cDNA library (13) in λgt11. A single positive clone containing a 1.3kb cDNA insert was isolated.

DNA sequencing. DNA sequencing of both strands was performed by the dideoxy-chain termination method (14) using ^{35}Sα-dATP (specific activity >1000 mCi/mmol, Amersham, UK) and M13 templates (15) containing the 1.3 kb cDNA insert. Synthetic oligonucleotides were used to prime reactions. The sequence was compiled and analyzed using modified computer programmes of Staden (16).

RESULTS AND DISCUSSION

The cDNA comprises 1320 nucleotides and contains one long open reading frame beginning immediately after the EcoR1 linker at the 5′ end of the insert in frame with β-galactosidase and extending to the termination codon (TAA) at positions 970-972 (Figure 1). There is a putative polyadenylation signal, AATAAA (17), at positions 1265-1270, 15 nucleotides in front of a stretch of 28 adenosine residues presumed to be the poly A tail.

The first potential initiation codon (ATG) was identified in frame with β-galactosidase at positions 88-90. Three

```
GAATTCCTCCTGCGCGGTTGTTCTCTGGAGCAGCGTTCTTTTATCTCCGTCCGCCTTCTCTCCTACCTAA
                                                                      70
                    1
                    M  E  D  S  M  D  M  D  M  S  P  L  R  P  Q  N  Y  L
GTGCGTGCCGCCACCCGATGGAAGATTCGATGGACATGGACATGAGCCCCCTGAGGCCCCAGAACTATCT
                                                                     140
     20                                                            40
  F  G  C  E  L  K  A  D  K  D  Y  H  F  K  V  D  N  D  E  N  E  H  Q
TTTCGGTTGTGAACTAAAGGCCGACAAAGATTATCACTTTAAGGTGGATAATGATGAAAATGAGCACCAG
                                                                     210
                                                          60
  L  S  L  R  T  V  S  L  G  A  G  A  K  D  E  L  H  I  V  E  A  E  A
TTATCTTTAAGAACGGTCAGTTTAGGGGCTGGTGCAAAGGATGAGTTGCACATTGTTGAAGCAGAGGCAA
                                                                     280
                                                    80
M  N  Y  E  G  S  P  I  K  V  T  L  A  T  L  K  M  S  V  Q  P  T  V  S
TGAATTACGAAGGCAGTCCAATTAAAGTAACACTGGCAACTTTGAAAATGTCTGTACAGCCAACGGTTTC
                                                                     350
                                          100
  L  G  G  F  E  I  T  P  P  V  V  L  R  L  K  C  G  S  G  P  V  H  I
CCTTGGGGGCTTTGAAATAACACCACCAGTGGTCTTAAGGTTGAAGTGTGGTTCAGGGCCAGTGCATATT
                                                                     420
                                120
  S  G  Q  H  L  V  A  V  E  E  D  A  E  S  E  D  E  E  E  E  D  V  K
AGTGGACAGCACTTAGTAGCTGTGGAGGAAGATGCAGAGTCAGAAGATGAAGAGGAGGAGGATGTGAAAC
                                                                     490
                    140
L  L  S  I  S  G  K  R  S  A  P  G  G  G  S  K  V  P  Q  K  K  V  K  L
TCTTAAGTATATCTGGAAAGCGGTCTGCCCCTGGAGGTGGTAGCAAGGTTCCACAGAAAAAAGTAAAACT
                                                                     560
     160                                                           180
  A  A  D  E  D  D  D  D  D  D  D  E  E  D  D  D  E  D  D  D  D  D  F
TGCTGCTGATGAAGATGATGACGATGATGATGAAGAGGATGATGATGAAGATGATGATGATGATTTT
                                                                     630
                                                    200
  D  D  E  E  A  E  E  K  A  P  V  K  K  S  I  R  D  T  P  A  K  N  A
GATGATGAGGAAGCTGAAGAAAAAGCGCCAGTGAAGAAATCTATACGAGATACTCCAGCCAAAAATGCAC
                                                                     700
                                          220
Q  K  S  N  Q  N  G  K  D  S  K  P  S  S  T  P  R  S  K  G  Q  E  S  F
AAAAGTCAAATCAGAATGGAAAAGACTCAAAACCATCATCAACACCAAGATCAAAAGGACAAGAATCCTT
                                                                     770
                                240
  K  K  Q  E  K  T  P  K  T  P  K  G  P  S  S  V  E  D  I  K  A  K  M
CAAGAAACAGGAAAAAACTCCTAAAACACCAAAAGGACCTAGTTCTGTAGAAGACATTAAAGCAAAAATG
                                                                     840
                          260
  Q  A  S  I  E  K  G  G  S  L  P  K  V  E  A  K  F  I  N  Y  V  K  N
CAAGCAAGTATAGAAAAAGGTGGTTCTCTTCCCAAAGTGGAAGCCAAATTCATCAATTATGTGAAGAATT
                                                                     910
                    280
C  F  R  M  T  D  Q  E  A  I  Q  D  L  W  Q  W  R  K  S  L  *
GCTTCCGGATGACTGACCAAGAGGCTATTCAAGATCTCTGGCAGTGGAGGAAGTCTCTTTAAGAAAATAG
                                                                     980

TTTAAACAATTTGTTAAAAAATTTTCCGTCTTATTTCATTTCTGTAACAGTTGATATCTGGCTGTCCTTT
                                                                    1050

TTATAATGCAGAGTGAGAACTTTCCCTACCGTGTTTGATAAATGTTGTCCAGGTTCTATTGCCAAGAATG
                                                                    1120

TGTTGTCCAAAATGCCGTTTAAGTTTTTAAAGATGGAACTCCACCCTTTGCTTGGTTTTAAGTATGTATG
                                                                    1190

GAATGTTATGATAGGACATAGTAGTAGCGGTGGTCAGACATGGAAATGGTGGGGAGACAAAAATATACAT
                                                                    1260

GTGAAATAAAACTCAGTATTTTAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGGAATTC
                                                                    1320
```

Figure 1. Nucleotide and predicted amino acid sequence of the λgt11 cDNA insert. Two negatively charged clusters of repeated glutamine and aspartic acid residues are shown in the boxed regions. The polyadenylation signal (AATAAA) is underlined. The numbers on the right refer to the nucleotide sequence and the superscripted numbers refer to the amino acid sequence.

additional in frame ATG codons are present at positions 100-102, 106-108 and 112-114 respectively. Comparison of the predicted human B23 amino acid sequence with sequences derived from highly homologous cDNA clones encoding rat B23 (8) and the X.laevis nucleolar phosphoprotein N038 (9) suggests that the first ATG is the correct initiation site (8). Furthermore, in vitro transcription of the human cDNA generated a 1.3 kb transcript which yielded, upon in vitro translation, a ^{35}S-methionine-labeled translation product that co-migrates with the 37kDa immunoreactive HeLa cell polypeptide, identified by the Ba serum (Li and McNeilage, unpublished observations). Thus the predicted polypeptide comprises 294 amino acids including the amino terminal methionine and has a calculated molecular weight of 32.5kDa. This value is smaller than 37kDa, observed on SDS-PAGE for B23. This anomalous migration may be due to the presence of 2 regions of negatively charged amino acids.

This distinctive feature comprises repeated glutamic and aspartic acid residues located at positions 120-132 and 161-187 (Figure 1). The former contains 11 Glu/Asp residues out of 13, the latter 26 out of 28. Extended Asp/Glu clusters are a feature of a number of other nuclear and nucleolar proteins (reviewed in 18) although the function of these negatively charged motifs, termed A^- or acidic regions (18) is not known.

Comparison of the predicted amino acid sequence of our cDNA clone with other mammalian nucleolar proteins identified 3 highly homologous sequences. The C-terminal 81 amino acids of our cDNA are identical with the predicted 82 C-terminal amino acids of partial cDNAs encoding rat and human B23 (7). This identity extends beyond the coding region for a further 114 nucleotides (7). The 20 amino acids from position 117 to position 136, which precede and include tl first negatively charged domain, show 95% identity with the amino acid sequence

determined for a major phosphopeptide of rat B23 (6). A second potential phosphorylation site with the sequence DT(P)PAK positions 198-202 (Figure 1) has been identified in B23 (19). This site is not phosphorylated in vitro (19) and has the predicted secondary structure found at or near phosphorylation sites regulated by cyclic AMP-dependent kinase (20), an extra-nucleolar protein kinase. Phosphorylation of this site may serve a regulatory role. Evidence that this may be regulated by an extranucleolar protein kinase is supported by the recent finding that the nuclear matrix protein, numatrin, which has a role in the transduction of receptor-mediated mitogenesis (4,5) is identical to B23 (4). Our sequence also showed 97.6% identity with the sequence recently derived for rat B23 (8) (Figure 2)

```
H B23 MEDSMDMD-MSPLRPQNYLFGCELKADK-DYHFKVDNDENEHQLSLRTVS
R B23 ********-********************-**********************
NO 38 ********NIA******F**********K**S****D*************
                                                        100
H B23 LGAGAKDELHIVEAEAMNYEGSPIKVTLATLKMSVQPTVSLGGFEITPPV
R B23 **************************************************
NO 38 ***S******v****GI****KT**iA**S**P****I*************
                                                        150
H B23 VLRLKCGSGPVHISGQHLVAVEEDAESEDEEEEDVKLLSISGKRSAPGGG
P B23                  ***************d*****
R B23 ********************************d*******GM**********
NO 38 i****S******Y********L-**L**S*d*d*eHEPSPKNA**I**DSA
                                                        200
H B23 SKVPQKKVKLAADEDDDDDDEEDDDEDDDDDDFDDEEAEEKAPVKKSIRD
R B23 N**********DE*d*e**e*d*-***********e**T***V*****v**
NO 38 ****R**TR*---*eeee*Sd****d*e****E**d*E**ET****T-DS
                                                        250
H B23 TPAKNAQKSNQNGKDSKPSSTPR----SKGQESFKKQEKTPKTPKGPSSV
H HB2                 P*********----**********************
R B23 ***************L***-***----**********************
NO 38 *KS*A***L*H***A*AL*T*QKTPKTPEQKGKQDTKPQ******T*L*S
                                                        300
H B23 EDIKAKMQASIEKGGSLPKVEAKFINYVKNCFRMTDQEAIQDLWQWRKSL
H HB2 **************************************************
R B23 **************************************************
NO 38 *e******TYL***NV******V**A*********TEN*KV*E***K**Q**

H B23 ----
R B23 ----
NO 38 KDGK*
```

Figure 2. Alignment of amino acid sequences of the rat B23 clone (R B23;8), X.laevis NO38 clone (NO38;9), a human B23 phosphopeptide (P B23;6) and a partial cDNA clone encoding the carboxyl 82 amino acids of human B23 (H HB2;7) relative to the human B23 sequence (H B23). The sequences, shown in single-letter amino acid code, were aligned according to the modified programs of Staden (16). Spaces (-) have been introduced to give maximum alignment and residues identical with those in the human B23 sequence are represented by asterisks. Conservative amino acid substitutions are represented in lower case letters.

and 68% identity with the sequence for the X.laevis nucleolar phosphoprotein NO38 (9) (Figure 2).

Comparison of our sequence with that for nucleolin (13), showed that both proteins contained extended domains rich in glutamic and aspartic acid residues. However, outside these regions there were significant differences in both structural features and potential modification sites. Nucleolin contains four potential phosphorylation sites (serines) within 3 acidic domains in the N terminal region of the protein (10), whereas only one potential phosphorylation site, serine at position 125, is observed in the corresponding acidic domains of B23. In addition, nucleolin contains 5 potential N-glycosylation sites (10) compared with none in B23.

ACKNOWLEDGMENTS

We thank Dr G. Peterson, Drs I. Hariharan, S. Foote and A. Cowman for advice and materials, Dr T. Kyne for his help with the computer analysis and Dr F. Firkin, St Vincent's Hospital, Melbourne for sending us serum from patient Ba. We also thank Mrs J. Lygnos for her excellent secretarial assistance. S.W. and L.J. McN. are supported by The National Health and Medical Research Council, Australia.

REFERENCES

1. Spector, D.L., Ochs, R.L. and Busch, H. (1984) Chromosoma 90,139-148.
2. Chan, P.K., Aldrich, M. and Busch, H. (1985) Exp. Cell Res. 161,101-110.
3. Yung, B.Y.M., Busch, H. and Chan, P.K. (1985) Biochimica et Biophysica Acta 826,167-173.
4. Feuerstein, N., Chan, P.K. and Mond, J.J. (1988) J. Biol. Chem. 263,10608-10612.
5. Feuerstein, N., Spiegel S. and Mond, J.J. (1988) J. Cell Biol. 107,1629-1642.
6. Chan, P.K., Aldrich, M., Cook, R.G. and Busch, H. (1987) J. Biol. Chem. 261,1862-1872.
7. Chan, P.K., Chan, W.Y., Yung, B.Y.M., Cook, R.G., Aldrich, M.B., Ku, D., Goldknopf, I.L. and Busch, H. (1987) J. Biol. Chem. 261,14335-14341.
8. Chang, J.H., Dumbar, T.S. and Olsen, M.O.J. (1988) J. Biol. Chem. 263,12824-12827.
9. Schmidt-Zachmann, M.S., Hügle-Dörl, B. and Franke, W.W. (1987) EMBO J. 6,1881-1890.
10. Lapeyre, B., Bourbon, H. and Amalric, F. (1987) Proc. Natl. Acad. Sci. USA. 84,1472-1476.

11. Bürglin, I.R., Mattaj, I.W., Newmeyer, D.D., Zeller, R. and De Robertis, E.M. (1987) Genes Dev. 1,97-107.
12. Huynh, T.V., Young, R.A. and Davis, R.W. (1984) pp. 49-78. In: D. Glover (ed.), DNA cloning techniques: A practical approach. IRL. Press, Oxford, Vol I, pp. 49-78.
13. Hariharan, I.K., and Adams, J.M. (1987) EMBO J. 6,115-119.
14. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74,5463-5467.
15. Messing, J. (1983) Methods In Enzymology 101,20-78.
16. Staden, R. (1982) Nucleic Acids Res. 10,4731-4751.
17. Proudfoot, N.J., and Brownlee, G.C. (1976) Nature 263,211-214.
18. Earnshaw, W.C. (1987) J. Cell. Biol. 105,1479-1482.
19. Jones, C.E., Busch, H. and Olsen, M.O.J. (1981) Biochimica et Biophysica Acta. 667,209-212.
20. Carmichael, D.F., Geahlen, R.L., Allen, S.M. and Krebs, E.A. (1982) J. Biol. Chem. 257,10440-10445.