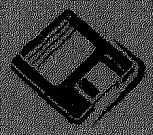
J Mol Med (1997) 75: 467–469

Bioinformatics: Bits and Bytes



by Peer Bork

The breast cancer gene product TSG101: a regulator of ubiquitination?

Chris P. Ponting · Yu-Dong Cai · Peer Bork

Sequence analysis is a powerful tool to obtain structural and functional information about genes and their products. Here we show that TSG101, a gene subjected to somatic mutations in breast cancer, contains an amino terminal domain that is a homologue of ubiquitin conjugating enzymes (UBCs) and not, as previously proposed, DNA-binding domains. As the UBC active site residue is replaced in the TSG101 sequence in a similar manner to several other members of the UBC family, we propose a role for TSG101 in regulating the ubiquitination of short-lived gene products.

C.P. Ponting (✉)
University of Oxford, Old Observatory,
Fibrinolysis Research Unit,
South Parks Road,
Oxford OX1 3RH, UK

Y.-D. Cai · P. Bork
EMBL,
Meyerhofstrasse 1,
D-69012 Heidelberg, Germany,
and
Max-Delbrück-Center for
Molecular Medicine,
Berlin-Buch, Germany

Please send articles to:
Peer Bork
Max-Delbrück-Center
for Molecular Medicine (MDC)
Robert-Rössle-Strasse 10
D-13122 Berlin, Germany
and:
EMBL
Meyerhofstrasse 1
D-69117 Heidelberg, Germany
E-mail: bork@embl-heidelberg.de
http://www.embl-heidelberg.de/~bork/

Recently the tumour susceptibility gene TSG101 has been found to be mutated at high frequency in human breast cancer [1]. The mouse [2] and human TSG101 gene products have been interpreted as containing both a coiled coil that binds stathmin and a coincident leucine zipper that binds DNA, thereby implying both cytoplasmic and nuclear roles for TSG101 [1]. All identified TSG101 deletion mutants result in truncations of this coiled coil region [1]. Li and Cohen [2] suggest the presence of both a Gal4-like binuclear zinc cluster domain and a bacterial type helix-turn-helix motif within the N-terminal region.

However, our analysis indicates that TSG101 contains neither the Gal4-like zinc cluster domain, which to date has found only in fungi, nor the helix-turn-helix motif. Rather, we find the TSG101 N-terminal region to be similar in sequence to the catalytic (UBC) domain of ubiquitin-conjugating enzymes E2 and UBC-related DNA-binding proteins. We propose that TSG101 and UBC molecules are homologues (genes that possess a common evolutionary ancestor) although they are unlikely to possess the same functions (see below). A close homologue (probable orthologue) of human TSG101 occurs in yeasts. These observations provide insight into the modular architecture of TSG101 and suggest a nuclear role for TSG101, potentially as a regulator of ubiquitination. We have become aware that similar conclusions have been reached independently by others [3].

A database search with the sequence preceding the proline-rich region of human TSG101 (residues 1–128) using Tblastn [4] revealed significant similarity (P=4.0×10^{-5}) to a region of *Saccharomyces cerevisiae* chromosome III (gene *Ycl*8c). TSG101 shows significant similarity in the upstream region of the originally annotated Ycl8c gene (see SwissProt YCA8_YEAST), and indeed a novel gene prediction algorithm, GIN, that makes use of a wide variety of sequence information (Y.-D.C. and P.B, unpublished) reveals a larger gene with two introns that includes the homologous upstream region. This alternative Ycl8c is similar to TSG101 over its entire sequence (Blastp P=5.1×10^{-19}), with the exception of the central proline-rich and coiled coil regions which appear to be lacking in Ycl8c. Further confirmation of the Ycl8c sequence was later provided following comparison with a recently deposited *Saccharomyces pastorianus* orthologue (LgYCLUN5c; T. Andersen and T. Nilsson-Tillgren, unpublished, EMBL accession code Z86109).

Since each of the ten next-highest scoring sequences in a Blastp search [4] encompassed the catalytic domain (UBC) of ubiquitin-conjugating enzymes (E2) we were prompted to perform a more detailed analysis of the entire family of UBC domains using profile and motif searches [5]. Successive iterations of database searches using SWise [6] established eight sequences, including TSG101, as newly identified UBC homologues. For example, using a "positive" profile derived from an alignment of 13 yeast UBC domains, a first iteration yielded scores of between 6947 and 8121 for the eight putative homologues, whereas previously known UBCs scored between 7252 and 9591, and the first probable false positive scored 6989.

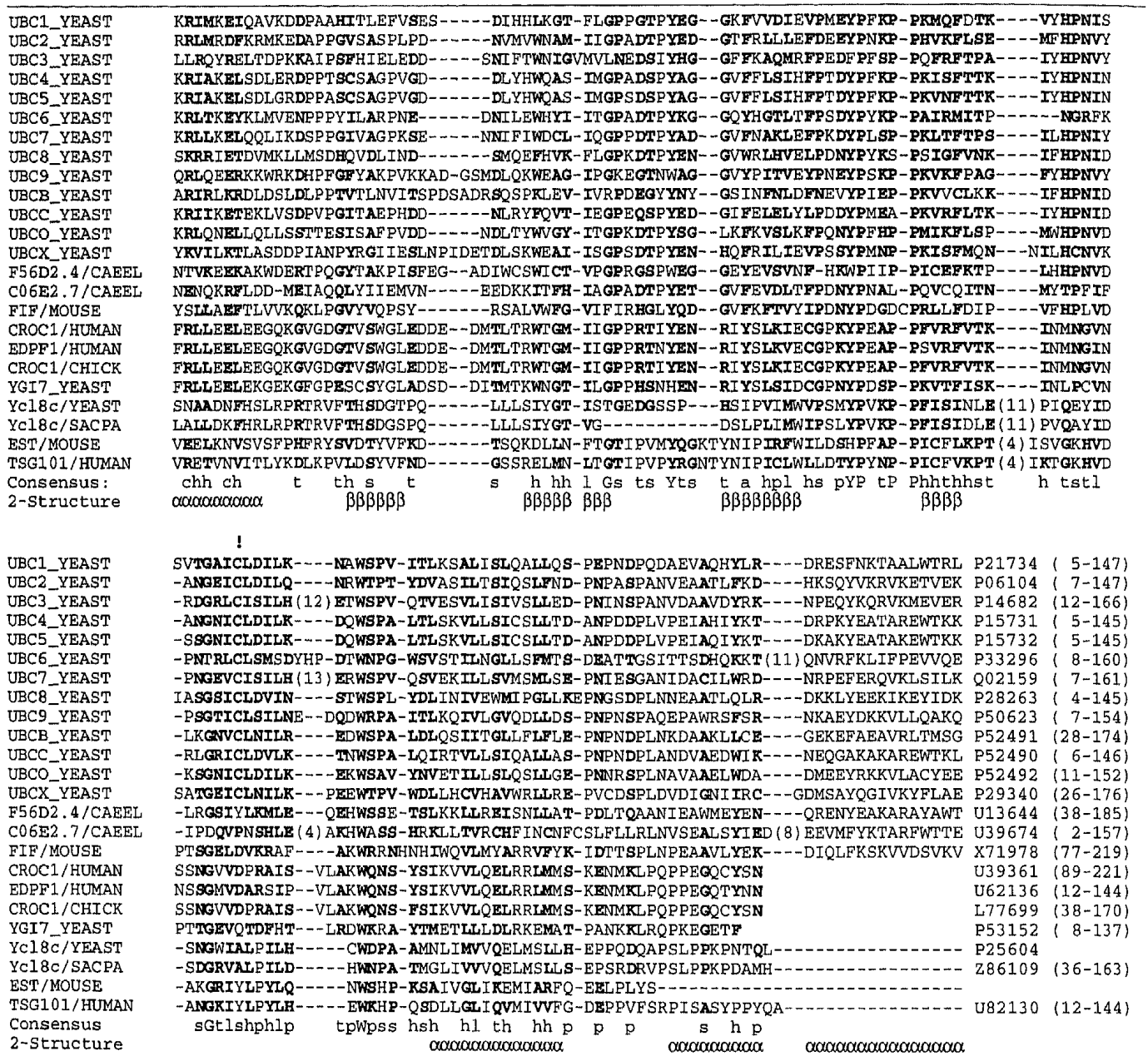


Fig. 1 Alignment of TSG101, Ycl8c and other UBC homologues that lack the active site cysteine, with the 13 yeast E2-type UBC domains contained in the completely sequenced yeast genome; the 13 represent all sub-branches of a phylogenetic tree constructed from the set of UBC sequences that are currently detectable in databases (data not shown). *Exclamation* mark (!), active site cysteine of E2-type UBCs; *numbers in parentheses*, residues omitted from the alignment; SwissProt or EMBL accession codes and residue limits are shown following the alignment; *dots*, insertions/deletions; *hyphens*, incomplete sequences or deleted secondary structures; the known secondary structure of yeast UBC4 [16] is shown beneath the alignment. TSG101, Ycl8c, and CROC-1 appear to have dispensed with the C-terminal helix of UBC domains; however comparisons with the known structures of UBCs [11, 16] indicate that its deletion would be unlikely to disrupt the UBC fold. A paralogue of mouse TSG101 was found using Tblastn [4] to be partially encoded by an expressed sequence tag (EST; accession code AA199318)

These eight candidate UBC homologues were: TSG101; CROC-1, a DNA-binding protein that mediates transcriptional activation of the human c-fos promotor from the direct repeat enhancer (S.L. Li, EMBL accession codes U39361 and U39360); a putative human enterocyte differentiation promoting factor (J. Faria, GE Wild, EMBL code U62136); a murine mutant fused toes protein (FIF; F. van der Hoeven et al., unpublished, EMBL accession code X71978), which is a morphogenic protein; yeast Ycl8c; an additional yeast protein (YGI7_YEAST); and two hypothetical proteins from *Caenorhabditis* elegans, F56d2.4 and C06e2.7.

CROC-1 homologues, including FIF and yeast YGI7, were shown to be significantly similar to known UBC homologues, using Blastp ($P<3\times10^{-5}$). Importantly, human and mouse TSG101, and *S. cerevisiae* and *S. pastorianus* Ycl8c scored higher than any other sequences when databases were scanned using either positive or negative profiles generated from an alignment of known UBC homologues, including FIF, CROC-1 and yeast YGI7. The motif searching algorithm MoST [7] identified YGI7_YEAST and CROC-1 as UBC homologues with $E<6.0\times10^{-4}$ using a block encompassing strands β3–β4 of known UBC homologues (I=80%).

E2 enzymes catalyse conjugation of ubiquitin to protein substrates during selective degradation of short-lived regulatory proteins (reviewed in [8, 9]). Proteins with a wide variety

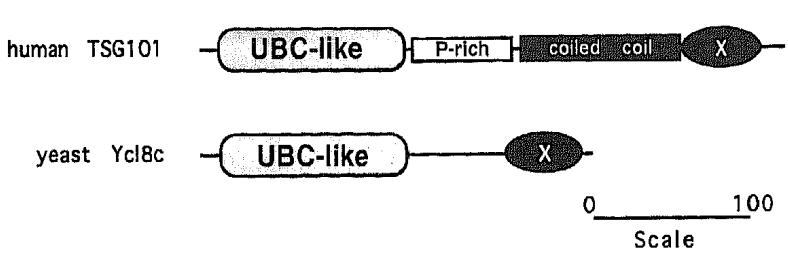


Fig. 2 Schematic representation of the domain organisation of human TSG101 and *S. cerevisiae* Ycl8c (domains approximately to scale). Each contains the N-terminal UBC homologous domain and a C-terminal region of unknown function. The *S. pastorianus* Ycl8c orthologue appears not to possess the C-terminal domain. TSG101 contains proline- (*P*-) rich and coiled coil regions

of cellular functions are subject to ubiquitin-mediated proteolysis, and include tumour suppressors such as p53, cell cycle regulators, and the transcription factors, c-jun and c-fos (reviewed in [8, 10]). The diversity of UBC isoforms (15 UBC homologues in yeast and over 20 in Arabidopsis thaliana [11]) indicates that E2 proteins have a range of substrate specificities. Indeed, E2s are known to be involved in DNA repair [12] and in the regulation of G^1 to S phase transition in the cell cycle [13]. However, studies on the in vivo degradation of the yeast Matα2 repressor demonstrate an overlap in E2 substrate specificities mediated in part by homotypic heterodimerisation of E2s [14].

Our studies lead us to conclude that human and mouse TSG101 are homologues of the catalytic UBC domain of E2 enzymes. Furthermore, close homologues of TSG101 occur in two yeasts, *S. cerevisiae* and *S. pastorianus*. However, surprisingly, each of these four molecules is unlikely to conjugate ubiquitin since a cysteine residue that in E2 enzymes mediates reversible attachment of ubiquitin is substituted (Figs. 1, 2). Although not suggested previously, it would appear that inactive UBC homologues are not uncommon. In our sequence database searches we have identified seven additional UBC homologues that contain similar active site substitutions (Fig. 1). Of these, several are closely related to CROC-1, a DNA-binding protein reported to mediate transcriptional activation of the human c-fos promotor from the direct repeat enhancer (S.L. Li, EMBL accession codes U39361 and U39360).

By analogy to CROC-1 and considering the known propensity of UBC homologues to form heterotypic dimers [14], TSG101 may possess UBC- and DNA-binding functions in a similar manner to Rad18, a DNA-binding protein that also binds the UBC homologue, Rad6 [15]. Other potential binding partners are ubiquitin, ubiquitin homologues, ubiguitin-activating enzymes, and ubiquitin-ligases. Alternatively the cellular transformation that results from TSG101 knockout [2] may suggest a dominant role in regulating the degradation of tumour suppressors such as p53 or transcriptional regulators such as c-jun and c-fos as multiple binding sites may have been retained despite the loss of catalytic UBC activity.

References

1. Li L, Francke U, Cohen SN (1997) The TSG101 tumor susceptibility gene is located in chromosome 11 band p15 and is mutated in human breast cancer. Cell 88:143–154
2. Li L, Cohen SN (1996) tsg101: a novel tumor susceptibility gene isolated by controlled homozygous functional knockout of allelic loci in mammalian cells. Cell 85:319–329
3. Koonin EV, Abagyan RA (1997) TSG101, a gene mutated in human breast cancer, may be the prototype of a class of dominant negative regulators of the ubiquitin pathway. Nat Genet (in press)
4. Altschul SF, Boguski MS, Gish W, Wootton JC (1994) Issues in searching molecular sequence databases. Nat Genet 6:119–129
5. Bork P, Gibson TJ (1996) Applying motif and profile searches. Meth Enzymol 266:162–184
6. Birney E, Thompson JD, Gibson TJ (1996) PairWise and SearchWise: finding the optimal alignment in a simultaneous comparison of a protein profile against all DNA translation frames. Nucleic Acids Res 24:2730–2739
7. Tatusov RL, Altschul SF, Koonin EV (1994) Detection of conserved segments in proteins: iterative scanning of sequence databases with alignment blocks. Proc Natl Acad Sci USA 91:12091–12095
8. Hochstrasser M (1995) Ubiquitin, proteasomes, and the regulation of intracellular protein degradation. Curr Opin Cell Biol 7:215–223
9. Jentsch S (1992) The ubiquitin conjugation system. Annu Rev Genet 26:177–205
10. Rolfe M, Chiu MI, Pagano M (1997) The ubiquitin-mediated proteolytic pathway as a therapeutic area. J Mol Med 75:5–17
11. Cook WJ, Jeffrey LC, Sullivan ML, Vierstra RD (1992) Three-dimensional structure of a ubiquitin-conjugating enzyme (E2). J Biol Chem 267:15116- 15121
12. Jentsch S, McGrath JP, Varshavsky A (1987) The yeast DNA repair gene Rad6 encodes a ubiquitin-conjugating enzyme. Nature 329:131–134
13. Goebl MG, Yochem J, Jentsch S, McGrath JP, Varshavasky A, Byers B (1988) The yeast cell cycle gene Cdc34 encodes a ubiquitin-conjugating enzyme. Science 241:1331–1335
14. Chen P, Johnson P, Sommer T, Jentsch S, Hochstrasser M (1993) Multiple ubiquitin-conjugating enzymes participate in the in vivo degradation of the yeast MATα2 repressor. Cell 74:357–369
15. Bailly V, Lamb J, Sung P, Prakash S, Prakash L (1994) Specific complex formation between yeast RAD6 and Rad18 proteins: a potential mechanism for targeting RAD6 ubiquitin-conjugating activity to DNA damage sites. Genes Dev 8:811–820
16. Cook WJ, Jeffrey LC, Xu Y, Chau V (1993) Tertiary structures of class I ubiquitin-conjugating enzymes are highly conserved: crystal structure of yeast Ubc4. Biochemistry 32:13809–13817

Note added in proof A sequence identical to the mouse EST sequence shown in Fig.1 has recently been submitted to databases (Burbelo PD, EMBL accession code U97571).