

This work is dedicated to Professor Jacek Augustyniak

Cloning and characterization of *Arabidopsis thaliana* AtNAP57 – a homologue of yeast pseudouridine synthase Cbf5p[✉]

Jarosław Maceluch, Maciej Kmiecik, Zofia Szweykowska-Kulińska
and Artur Jarmołowski[✉]

*Adam Mickiewicz University, Faculty of Biology, Department of Gene Expression,
Poznań, Poland*

Received: 22 January, 2001; accepted: 15 July, 2001

Key words: Cbf5p, NAP57, dyskerin, H/ACA snoRNAs, pseudouridine synthase, Ψ

Rat Nap57 and its yeast homologue Cbf5p are pseudouridine synthases involved in rRNA biogenesis, localized in the nucleolus. These proteins, together with H/ACA class of snoRNAs compose snoRNP particles, in which snoRNA guides the synthase to direct site-specific pseudouridylation of rRNA. In this paper we present an *Arabidopsis thaliana* protein that is highly homologous to Cbf5p (72% identity and 85% homology) and NAP57 (67% identity and 81% homology). Moreover, the plant protein has conserved structural motifs that are characteristic features of pseudouridine synthases of the TruB class. We have named the cloned and characterized protein AtNAP57 (*Arabidopsis thaliana* homologue of NAP57). AtNAP57 is a 565 amino-acid protein and its calculated molecular mass is 63 kDa. The protein is encoded by a single copy gene located on chromosome 3 of the *A. thaliana* genome. Interestingly, the AtNAP57 gene does not contain any introns. Mutations in the human DKC1 gene encoding dyskerin (human homologue of yeast Cbf5p and rat NAP57) cause *dyskeratosis congenita* a rare inherited bone marrow failure syndrome characterized by abnormal skin pigmentation, nail dystrophy and mucosal leukoplakia.

[✉]This work was supported by the State Committee for Scientific Research (KBN, Poland) grants: 6P04A03511 and 6P04A00312, and by the A. Mickiewicz University Faculty of Biology grant PBWB 203/2000 to Jarosław Maceluch.

[✉]Adam Mickiewicz University, Faculty of Biology, Department of Gene Expression, Międzychodzka 5, 60-371 Poznań, Poland; phone: (48 61) 829 2730; fax: (48 61) 829 2733; e-mail: artjarmo@amu.edu.pl

Abbreviations: EST, expressed sequence tag; hTR, human telomerase RNA; NLS BP, bipartite nuclear localization signal; ORF, open reading frame; Ψ , pseudouridine; RACE, rapid amplification of cDNA ends.

In eukaryotes, 18S, 5, 8S and 28S rRNA molecules are transcribed as a single large precursor, which undergoes a complicated maturation process [1]. This process occurs in a highly specialized part of the nucleus known as the nucleolus [2]. Maturation of the precursor molecule consists of many endo- and exonucleolytic cleavages leading to the production of mature rRNA species and also of modifications of standard nucleosides at specific positions. One of the most frequent post-transcriptional modification of rRNA is pseudouridylation.

Pseudouridine (Ψ , 5-ribosyluracil) was the first modified nucleoside to be discovered, and this is the most abundant modification in all RNA species [3]. Many Ψ residues were found not only in rRNA, but in tRNA, snRNA and snoRNA molecules as well [4–6]. Despite this frequent localization of Ψ in various classes of RNA its role in RNA structure and function has not yet been finally elucidated [3]. The conversion of uridine to pseudouridine involves the breakage of the N1 glycosidic bond and the rotation of the base by 180° around the N3–C6 axis, followed by reformation of a covalent bond at position C5 [7]. This reaction is catalyzed by pseudouridine synthases. Many representants of this large and ancient enzyme family have been identified in both prokaryotes and yeast, with TruA, TruB, RluA and RsuA representing four distinct subfamilies of Ψ -synthases [8, 9].

Selection of uridine residues for pseudouridylation by prokaryotic Ψ -synthases occurs with a high degree of site-specificity, both for tRNA and rRNA substrates. In these cases, each synthase is generally responsible for pseudouridylation of one or more [10] U positions in a particular RNA species. However, dual-specificity synthases catalyzing the formation of Ψ residues in both tRNAs and rRNAs [11], tRNAs and snRNAs [12], and in both cytoplasmic and mitochondrial tRNAs [13] have been described. In eukaryotic RNAs (especially rRNA), due to the numerous localization of Ψ [4, 6], site-selection for pseudo-

uridylation has to proceed in another way. It seems that this process involves a specific class of small nucleolar RNAs (snoRNAs) known as box H/ACA snoRNAs, that act as guides to direct site-specific pseudouridylation of rRNA [14–16]. These snoRNAs possess a hairpin-hinge-hairpin-tail structure, with the hinge region containing the conserved sequence block ANANNA (H box) and the ACA motif found in the tail structure, three nucleotides away from the 3'-end [17]. Target rRNAs have short signal sequences located around U to be converted to Ψ . Site selection for Ψ synthesis occurs by base-pair interactions between antisense elements in the H/ACA snoRNAs and sequences in the rRNA on both sides of the target uridine. This creates a pocket structure in which the target U is unpaired and accessible to the Ψ -synthase. The snoRNAs interact with many proteins creating functional ribonucleoprotein particles (snoRNPs). Each box H/ACA snoRNA associates with at least four specific proteins, one of which is pseudouridine synthase, known as Cbf5p in yeast [18] or NAP57 [19] in mammals.

Cbf5p was originally isolated as a low-affinity *in vitro* centromeric DNA binding protein [20]. CBF5 is an essential gene encoding a 55 kDa highly charged protein with a domain containing ten tandem KKE/D repeats near the C terminus. Cbf5p is a highly conserved protein with homologues found in many organisms (including rat – NAP57 and human – dyskerin). Cbf5p and its orthologues have high sequence homology to *Escherichia coli* TruB pseudouridine synthase which catalyzes the conversion of uridine to Ψ at position 55 in tRNA [21]. The region of TruB believed to be the conserved U-binding domain, is the area showing the greatest homology to Cbf5p. Furthermore, Cbf5p shares the KP and XLD motifs found in three out of the four distinct families of known and putative Ψ synthases [22]. Cbf5p coimmunoprecipitates with all members of the H/ACA class snoRNAs [18]. This suggests that Cbf5p directly inter-

acts with the conserved H/ACA box to provide metabolic stability to the snoRNA and participate in assembly of the snoRNP molecule. It has been shown that genetic depletion of Cbf5p inhibits Ψ formation in rRNA, resulting in accumulation of the unmodified rRNA molecules, and in delay or arrest of the yeast cell cycle [22]. Cbf5p shows high homology (71% identity, 85% similarity) to the rat nucleolar protein NAP57, which was identified as a protein associated with nucleolar shuttling protein Nopp140 [19]. Nucleolar localization and homology with known Ψ synthases suggests that Cbf5p and NAP57 might be the rRNA Ψ synthases in yeast and mammals, respectively, guided to their target sites by box H/ACA snoRNAs [23]. In these snoRNP particles Cbf5p/NAP57 are the active Ψ synthesizing components. Interesting new features of the human Cbf5p homologue, dyskerin, have been recently revealed (see Results and Discussion).

In this work we present the identification of *A. thaliana* Cbf5p/NAP57 homologue. AtNAP57, as we named the protein, shares a high degree of sequence identity with mammalian (rat NAP57 and human dyskerin) and yeast (Cbf5p) proteins. It is the first plant homologue that belongs to the family of Ψ synthases.

MATERIALS AND METHODS

GenBank accession number. The nucleotide and protein sequence of AtNAP57 have been submitted to the GenBank database under accession No. AF234984.

Plant material and growth conditions. Seeds of the plant *A. thaliana* (ecotype Columbia) were supplied by Lehle Seeds (Round Rock, U.S.A.). They were grown in a greenhouse (22°C with 16 h light photoperiod) on soil irrigated with mineral nutrients as suggested by the seed producer.

Cloning of AtNAP57 cDNA. To clone cDNA of the *A. thaliana* gene encoding a

homologue of yeast protein Cbf5p we performed 5' and 3' RACE (rapid amplification of cDNA ends) experiments [24, 25]. To obtain gene specific primers for RACE we searched the GenBank (www.ncbi.nlm.nih.gov) database with the protein sequence of rat NAP57 [26]. We got one *A. thaliana* EST partial cDNA sequence (GenBank F20038). On the basis of this sequence, we designed three 5' RACE and three 3' RACE primers. To characterize the 3' end of AtNAP57 transcript, total RNA was isolated from whole two weeks old *A. thaliana* seedlings using the RNeasy Plant Mini Kit (Qiagen). The RNA was then reverse-transcribed with SuperScript II RNase H⁻ RT from Gibco BRL using the oligo-dT primer Qt. Specific cDNAs were then amplified by two rounds of PCR with other EST gene specific nested primers (NAP1, NAP2, NAP3) and the adapter primers Qo and Qi, using Taq Polymerase (Qiagen). To amplify the 5' region of the transcript we used the antisense 5' gene specific primer NAP4 to generate cDNA in the reverse transcription reaction. An A-tailing step (the terminal transferase was from Roche Molecular Biochemicals) was carried out to attach an adapter sequence to the 5' end of cDNA. Specific cDNAs were then amplified by two rounds of PCR with gene specific primers (NAP5, NAP6) and Qt, Qo and Qi adapter primers. The PCR conditions for both 5' and 3' RACE were as follows: the 1st round: 95°C for 15 min, 50°C for 2 min, 72°C for 40 min, 30 cycles of 94°C/30 s, 55°C/30 s, 72°C/1 min, and final extension 72°C for 10 min. The 2nd round: 95°C for 3 min, 30 cycles of 94°C/30 s, 55°C/30 s, 72°C/1 min, and final extension 72°C for 10 min. All amplified fragments were purified from agarose gels using the QIAquick Gel Extraction Kit (Qiagen), cloned into pGEM T-Easy vector (Promega) and manually sequenced (the fmol DNA Sequencing System was supplied by Promega and [α^{35} S]dATP radioisotope was from ICN Biomedicals) using both vector and gene specific primers (listed below).

DNA and protein sequence analysis. The obtained DNA sequences were assembled and analyzed with the Lasergene computer program (DNA STAR Inc., U.S.A.). Protein alignments were performed with the ClustalW [27] and Boxshade [www.ch.embnet.org] programs. Computer-assisted searches for nucleotide and amino-acid sequences were carried out using the BLAST tools [26]. Conserved domain searches were performed with the help of Reverse Position Specific BLAST (GenBank Conserved Domain Database) [28] and ProfileScan (Prosite Database) [www.isrec.isb-sib.ch] programs.

Primers used in reactions. Below, the list of primers used to amplify and sequence AtNAP57 cDNA is shown. Primers NAP1-NAP6 and Qt, Qo, Qi were used in RACE reactions, primers NAP7-NAP14 and Universal

Qt -	CCAGTGAGCAGAGTGACGAGGAC- TCGAGCTCAAGCTTTTTTTTTTTT- TTTTT
Qo -	CCAGTGAGCAGAGTGAC
Qi -	GAGGACTCGAGCTCAAGC
NAP1 -	GCGGAGGTCGACATCTCAC
NAP2 -	GTCATCAATCTCGATAAACCC
NAP3 -	AGTGGTACTCTTGACCCAAAAG
NAP4 -	CCCGCGAACCCCTTCTTCTCACGA- GGTC
NAP5 -	CGACCGCCTCAACGTCCGTACC
NAP6 -	GCCACAGAGTTTCACTCCAGCCA- TC
Universal -	GTAAAACGACGGCCAGT
Reverse -	CAGGAAACAGCTATGAC
NAP7 -	GACTACATGATCAAGCCACAGAG
NAP8 -	CCTCATTGTCTGTATTGACCG
NAP9 -	CCTCCTTTGATCTCTGCTG
NAP10 -	GGGTTTCTTGTGAGGCGGG
NAP11 -	GACAACTCCAGAGATGAG
NAP12 -	GGGTGAGGCGATTGCCG
NAP13 -	GGATAAGCATGGGAAGCCTAATG
NAP14 -	GCCCTGCTCCTGTAACAACC
NAP15 -	GGAGAAAGAAGAGGAAGCC
NAP16 -	CACCAAAGTCTGAGAAGAAGA

and Reverse primers (specific for pGEM T-Easy vector) were used in sequencing reactions. All primers are listed in 5' to 3' direction; they were supplied by Sigma-ARK Scientific.

RESULTS AND DISCUSSION

To clone cDNA encoding the *A. thaliana* homologue of Cbf5p/NAP57, we applied a RACE strategy. Gene specific primers were designed on the basis of *Arabidopsis* EST sequence with high homology to yeast Cbf5p and rat NAP57 (F20038). These primers and *Arabidopsis* total RNA were used in RACE reactions to amplify the 5' and 3' parts of cDNA of the putative *Arabidopsis* Cbf5p/NAP57 homologue. The 5' RACE reaction yielded a product about 200 bp in size (Fig. 1, lane 2) and the 3' reaction gave an about 1300 bp product (Fig. 1, lane 4). Both RACE products were directly cloned into pGEM T-Easy vector and sequenced. A short overlapping sequence in these products allowed us to compose a virtual, full length cDNA sequence. However, this sequence did not seem to be complete (it did not contain the STOP codon), so we decided to perform another 3' RACE reaction with two new primers (NAP15, NAP16) to elongate the 3' part of the sequence. This reaction yielded an about 300 bp product (not shown) that was also cloned and sequenced. Based on the obtained sequence data we designed two additional edge primers to directly amplify the full length cDNA molecule by PCR. The product was also cloned and sequenced to confirm that all earlier analyzed parts of cDNA were derived from the same mRNA. This gave us the complete cDNA sequence of the *A. thaliana* protein that is homologous to yeast Cbf5p and rat NAP57 (Fig. 2). We named the gene and its product AtNAP57 (*Arabidopsis thaliana* homologue of NAP57).

The cDNA sequence of AtNAP57 consists of 1935 nucleotides (GenBank AF234984). It

contains an 1698 bp ORF, starting at the nucleotide 59, encoding a 565 amino-acids pro-

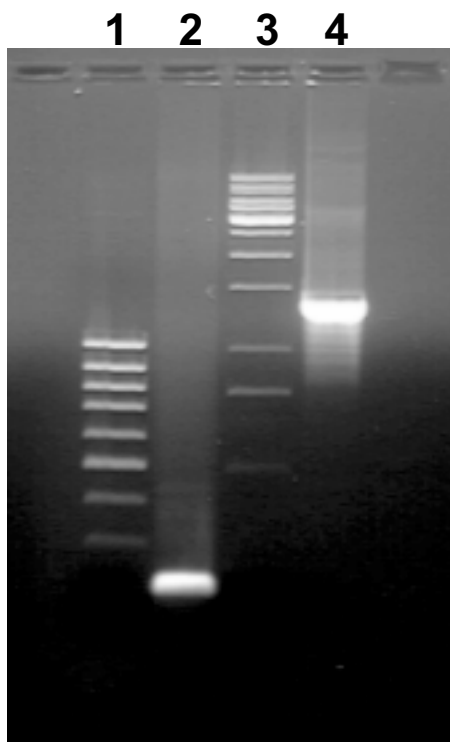


Figure 1. Results of the RACE reactions

Products of the 5' and 3' RACE reactions of AtNAP57 separated on 2% agarose gel and visualized by ethidium bromide staining. Lane 1, GeneRuler™ 100bp DNA Ladder (MBI Fermentas, size of bands (in bp): 1031, 900, 800, 700, 600, 500, 400, 300, 200-non visible, 100-non visible, 80-non visible). Lane 2, 5' RACE reaction product (about 200 bp). Lane 3, GeneRuler™ 1kb DNA Ladder (MBI Fermentas, size of bands: 10000, 8000, 6000, 5000, 4000, 3500, 3000-stronger, 2500, 2000, 1500, 1000, 750, 500, 250-non visible). Lane 4, 3' RACE reaction product (about 1300 bp).

tein (Fig. 2). Predicted size of this protein is 63 kDa. AtNAP57 is a strongly charged protein with many basic amino acids on its C-terminus (isoelectric point at 9.158). Genomic localization and structure of AtNAP57 gene were resolved by computer analysis. It revealed that the AtNAP57 gene is located within the sequence of *A. thaliana* chromosome 3 (AL138655). Interestingly, this AtNAP57 gene does not contain introns (genomic sequence is identical with cDNA). AtNAP57 contains conserved structural domains (Fig. 2) which may point to the protein

function. Moreover, it has the TruB family pseudouridylate synthase (N terminal domain) motif [21] throughout amino-acid residues 101–238 and the PUA motif (pos. 287–362) – a novel RNA binding domain detected in archaeal and eukaryotic pseudouridine synthases (designated PUA after pseudouridine synthase and archaeosine transglycosylase) [29]. A bipartite nuclear localization signal (NLS BP) [30] was also detected at positions 505–522. Other putative NLSs occur on both, the N and C ends of AtNAP57. A comparison of the amino-acid sequence of AtNAP57 with its most known homologues: yeast Cbf5p (L12351), rat NAP57 (Z34922) and human dyskerin (O60832), revealed that AtNAP57 is a highly conserved protein. Its high homology within these proteins is shown in Fig. 3. The arabidopsis protein shares 72% identity and 85% homology with yeast Cbf5p and, respectively, 65%/80% with human dyskerin, and 67%/81% with rat NAP57 protein. The highest homology is observed over the central region of AtNAP57 (pos. 40–420) that contains the TruB and PUA conserved domains. The highly charged, Lys and Glu rich C-terminal domain with the KKE/KKD repeats found in yeast Cbf5p is also present in the arabidopsis protein.

Considering all the conserved structural motifs (TruB, PUA and NLSs) and the high homology with yeast Cbf5p, which is known to be Ψ -synthase, we think that AtNAP57 is also a functional pseudouridine synthase. This can be proved by complementation test of yeast *cbf5* null mutation with AtNAP57. Such experiments were carried out for the *Drosophila melanogaster* homologue of CBF5–Nop60B gene. Nop60B encodes an essential nucleolar protein that complements a *cbf5* null mutation [31]. On the other hand, rat NAP57 does not complement the *cbf5* null phenotype in yeast [31]. It has to be checked whether AtNAP57 does function in yeast or not. The complementation experiments are in progress in our laboratory.

1	AGCTACAACITTTTAACGCGCTCTCTCTCATTCTTTTCMAAAACCCCTAATCCGACGGGAGCC	<u>TruB</u>	GCG GAG GTC GAC	
			<u>M A R V D</u>	5
75	ATC TCA CAC TCG AAG AAG AAG AAG CAA GAC AAA ACG GAG AAC GAT GCC GCC GAC ACC GGA			
	I S H S K K K K Q D K T E N D A A D T G			25
135	GAC TAC ATG ATC AAG CCA CAG AGT TTC ACT CCA GCC ATC GAC ACT TCT CAG TGG CCC ATC			
	D Y M I K P Q S F T P A I D T S Q W P I			45
195	CTC CTC AAG AAC TAC GAC GCG CTC AAC GTC CGT ACC GGT CAC TAC ACT CCA ATC AGC GCC			
	L L K N Y D R L M V R T G H Y T P I S A			65
255	GGT CAC TCT CCT CTG AAA CGT CCT CTT CAA GAG TAT ATC AGG TAC GGT GTC ATC AAT CTC			
	G H S P L K R P L Q E Y I R Y G V I N L			85
315	GAT AAA CCC GCG AAC CCT TCT TCT CAC GAG GTC GTC GCT TGG ATT AAG CGT ATC CTC CGT			
	D K P A N P S S H E V V A W I K R I L R			105
375	GTT GAG AAA ACC GGT CAC AGT GGT ACT CTT GAC CCA AAA GTC ACT GGA AAC CTC ATT GTC			
	V E K T G H S G T L D P K V T G N L I V			125
435	TGT ATT CAG CGA GCC ACA CGG CTT GTG AAA TCG CAG CAA GGT GCG GGT AAA GAG TAC GTT			
	C I D R A T R L V K S Q Q G A G K E Y V			145
495	TGT GTT GCT CCG CTT CAC TCA GCT GTC CCT GAT GTT GCT AAA GTA GCT AGA GCT CTT GAA			
	C V A R L H S A V P D V A K V A R A L E			165
555	TCG CTC ACT GGA GCT GTG TTT CAG AGG CCT CCT TTG ATC TCT GCT GTT AAA AGA CAG CTT			
	S L T G A V F Q R P P L I S A V K R Q L			185
615	AGG ATT AGG ACA ATC TAT GAA AGC AAG CTG CTT GAG TAT GAT GCT GAT AGG CAT TTG GTT			
	R I R T I Y E S K L L E Y D A D R H L V			205
675	GTT TTC TGG GTT TCT TGT GAG GCG GGT ACT TAC ATT AGG ACT ATG TGT GTT CAC TTG GGT			
	V F N W V S C E A G T Y I R T M C V H L G			225
735	TTA CTT CTT GGG GTT GGT GGG CAT ATG CAG GAG CTT AGG AGG GTT AGG TCT GGG ATT TTG			
	L L L G V G G H M Q E L R R V R S G I L			245
795	GGT GAG AAT AAT AAC ATG GTT ACT ATG CAT GAT GTG ATG GAT GCC CAA TTT GTG TAT GAC			
	G E H N H N H V T M H D V M D A Q F V Y D			265
855	AAC TCC AGA GAT GAG TCT TAT CTT AGG AGG GTG ATT ATG CCT CTT GAG ATG ATA TTG ACG			
	N S R D E S Y L R R V I M P L E M I L T			285
915	AGT TAC AAG AGA CTT GTT GTC AAG GAT TCT GCT GTG AAT GCT ATC TGT TAT GGT GCT AAG			
	S Y K R L V V K D S A V N A I C Y G A K			305
975	TTG ATG ATT CCT GGG TTG TTG AGA TTC GAG AAT GAC ATT GAT GTT GGT ACT GAA GTC GTT			
	L M I P G L L R F E N D I D V G T E V V			325
1035	CTT ATG ACT ACC AAG GGT GAG GCG ATT GCC GTT GGG ATT GCT GAG ATG ACA ACA TCT GTG			
	L M T T T K G E A I A V G I A E M T T S Y			345
1095	ATG GCT ACT TGT GAT CAT GGT GTG GTA GCT AAG ATC AAG CGT GTG GTG ATG GAT AGA GAT			
	M A T T C D H G V V A K I K R V V M D H D			365
1155	ACC TAC CCG AGG AAG TGG GGA TTG GGA CCA AAG GCT TCT ATG AAG AAG AAG CTT ATT GCT			
	T Y F R K W G L G P R A S M K K K L I A			385
1215	GAT GGG AAA TTG GAT AAG CAT GGG AAG CCT AAT GAG AAG ACT CCG GTT GAG TGG AGC AGG			
	D G K L D C K H G K P H E K T P V E W S R			405
1275	AAT GTG GTT TTG CCT ACC GGT GGA GAT GCT ATA ATT GCT GGT GCT GCT GCT CCT GAG			
	N V V L P T G G D A I I A G A A A P E			425
1335	GAG ATA AAG GCT GAC GCT GAG AAT GGA GAA GCA GGG GAA GCG CGT AAG CGT AAG CAT GAT			
	E I K A D A E N G E A G E A R K R K H D			445
1395	GAT AGC AGT GAT AGC CCT GCT OCT GTA ACA ACC AAG AAA TCT AAA ACC AAA GAA GTT GAA			
	D S S D S P A P V T T K K S K T K E V E			465
1455	GGA GAA GAG GCT GAA GAG AAG GTG AAG TCT TCT AAG AAG AAG AAG AAG GAT AAG GAA			
	G E E A E E K V K S K K K K K D K E			485
1515	GAG GAG AAA GAA GAG GAA GCC GGG TCT GAG AAG AAG GAA AAG AAG AAG AAG AAA GAT AAG			
	E E K E E E A G S E K K E K K K K D			505
1575	AAG GAG GAG GTT ATA GAA GAG GTA GCT TCA CCA AAG TCT GAG AAG AAG AAG AAA AAG AAG			
	<u>E N V I E V A S P K S E</u> <u>K K K</u>			525
1635	AGC AAA GAC ACC GAA GCT GCT GTT GAC GCA GAA GAT GAA TCA GCA GCA GAG AAG AGT GAG			
	S K D T E A A V D A E D E S A A E K S E			545
1695	AAG AAA AAG AAA AAG AAG GAT AAG AAG AAG AAG AAG AAG AAG AAG AAG AAG AAG AAG AAG			
	K K K K K K D K K K N K D S E D E E			565
1755	<u>TruB</u> GATGGGTTTGGGTTATGGTCTTTTAAAGTTTGAACCCITTTGATTTTGGCGGTATCTATCTACTCTGGAACAAA			
1833	GTGAGACAAAAGGCCAITTTTATGCTTGTATTCTAAGTCTCGCTTACGTTACAGATTTTATGTTTTGTAGGTTAGTTTC			
1912	AGTTCTAAGCAGATTTCTCCCTTTA			

Figure 2. Nucleotide and amino-acid sequence of AtNAP57.

The predicted amino-acid sequence is shown below the DNA sequence. Nucleotides are numbered on the left, and amino acids on the right. Start and stop codons are grey-shaded. The conserved TruB pseudouridine synthase domain is double line underlined. The PUA (RNA binding domain) motif is single line underlined. The NLS BP sequence is indicated by a box, critical amino acids are grey-shaded. The sequence has been submitted to GenBank (accession number AF234984).

A possible role of yeast Cbf5p in centromere function, indicated by binding of centromeric DNA, was not confirmed in other organisms. In *D. melanogaster* there is no apparent association between Nop60B protein and the chromosomes [31]. It is possible that the role of Cbf5p in centromere function is unique to yeast. Alternatively, it may be that a conserved centromeric function exists but has not been detected yet in other organisms, or that the association of Cbf5p with the yeast centromere is non-functional.

One of the best characterized CBF5 homologues is probably the human gene DKC1. This is because the DKC1 gene and its product dyskerin are believed to play a role in development of a human disease *dyskeratosis congenita* [32, 33]. Dyskeratosis is a rare inherited bone marrow-failure syndrome characterized by abnormal skin pigmentation, nail dystrophy, and mucosal leukoplakia. More than 80% of patients develop bone-marrow failure towards the end of the first decade of life, and this is the major cause of premature death. The X-linked (develops only in males) form of the disease has been shown to be caused by mutations in the DKC1 gene. This single-copy gene is localized on the X chromosome. It comprises 15 exons spanning at least 16 kb and is transcribed into a widely expressed 2.6 kb message [34]. Numerous missense mutations and one 3' deletion [35] were detected in DKC1 gene sequence in patients who suffered dyskeratosis. These mutations result in production of non-functional dyskerin that is thought to cause the disease. Dyskerin is a 514 amino-acids protein with a predicted molecular mass of 57.6 kDa [32]. Dyskerin localizes to the nucleolus – it contains multiple putative NLSs at the N- and C-ends [36, 37]. It is a highly charged peptide with some conserved sequence motifs like the TruB Ψ synthase motif, multiple phosphorylation sites, and a carboxy-terminal lysine-rich repeat domain [32]. The whole amino-acid sequence of dyskerin shows high homology with other pro-

teins from the TruB family of pseudouridine synthases (Fig. 3). Interestingly, all missense mutations characterized so far, are located in the most conserved regions of dyskerin. By analogy to the function of known dyskerin orthologues, involvement in the cell cycle, nucleolar function in ribosome biogenesis and Ψ synthesis have been predicted for the human protein. The molecular mechanism leading to *dyskeratosis congenita* has not yet been elucidated. On the basis of early experimental data it was thought to involve the rRNA biogenesis failure due to predicted RNA binding function of dyskerin in the nucleolus. Missense mutations in the DKC1 gene could modify the function of dyskerin affecting snoRNP assembly and stability. Recent data suggest, however, that dyskerin has another function the disturbance of which can lead to dyskeratosis. It was proposed that dyskeratosis may not be caused by the deficiency in rRNA, but rather by a defect in the maintenance of telomeres [38, 39]. It was previously shown that the 3' end of the RNA component of human telomerase (hTR) is structurally and functionally similar to the H/ACA family of snoRNAs. Dyskerin and its homologues interact with the H/ACA motif of snoRNAs, creating functional particles involved in rRNA biogenesis. It has been shown recently that dyskerin associates with the H/ACA portion of hTR and is a part of the human telomerase RNP complex. This interaction seems to be important for biogenesis, processing or turnover of the telomerase RNP. Decreased accumulation of hTR, reduced telomerase activity and abnormally short tracts of telomeric DNA were detected in the cells expressing mutated forms of dyskerin. On the other hand, the mutant dyskerins were still able to carry out snoRNP functions, as the mutation had no discernible impact on rRNA processing. Dyskeratosis is a disease affecting strongly dividing tissues like bone marrow and epithelium, and the risk to the patients of having some form of cancer is increasing with their age. This is consistent

with telomere shortening that leads to chromosomal instability, telomeric rearrangements and cancer progression.

The authors would like to thank Ms Kamila Jankowiak and Ms Katarzyna Ober for great help with DNA sequencing.

REFERENCES

1. Sollner-Webb, B., Tycowski, K.T. & Steitz, J.A. (1996) Ribosomal RNA processing in eukaryotes; in *Ribosomal RNA: Structure, Evolution, Processing and Function in Protein Biosynthesis* (Zimmermann, R.A. & Dahlberg, A.E., eds.) pp. 469-490, CRC Press, Boca Raton.
2. Scheer, U. & Hock, R. (1999) Structure and function of the nucleolus. *Curr. Opin. Cell Biol.* **11**, 385-390.
3. Charette, M. & Gray, M.W. (2000) Pseudouridine in RNA: What, where, how, and why. *Life* **49**, 341-351.
4. Ofengand, J., Bakin, A., Wrzesinski, J., Nurse, K. & Lane, B.G. (1995) The pseudouridine residues of ribosomal RNA. *Biochem. Cell Biol.* **73**, 915-924.
5. Massenet, S., Mougins, A. & Branlant, C. (1998) Posttranscriptional modifications in the U small nuclear RNAs; in *Modification and Editing of RNA* (Grosjean, H. & Benne, R., eds.) pp. 201-227, ASM Press, Washington, DC.
6. Ofengand, J. & Fournier, M.J. (1998) The pseudouridine residues of rRNA: Number, location, biosynthesis, and function; in *Modification and Editing of RNA* (Grosjean, H. & Benne, R., eds.) pp. 229-253, ASM Press, Washington, DC.
7. Gu, X., Liu, Y. & Santi, D.V. (1999) The mechanism of pseudouridine synthase I as deduced from its interaction with 5-fluorouracil-tRNA. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 14270-14275.
8. Koonin, E.V. (1996) Pseudouridine synthases: Four families of enzymes containing a putative uridine-binding motif also conserved in dUTPases and dCTP deaminases. *Nucleic Acids Res.* **24**, 2411-2415.
9. Samuelsson, T. & Olsson, M. (1990) Transfer RNA pseudouridine synthases in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **265**, 8782-8787.
10. Motorin, Y., Keith, G., Simon, C., Foiret, D., Simos, G., Hurt, E. & Grosjean, H. (1998) The yeast tRNA: Pseudouridine synthase Pus1p displays a multisite substrate specificity. *RNA*. **4**, 856-869.
11. Wrzesinski, J., Nurse, K., Bakin, A., Lane, B.G. & Ofengand, J. (1995) A dual-specificity pseudouridine synthase: An *Escherichia coli* synthase purified and cloned on the basis of its specificity for Ψ 746 in 23S RNA is also specific for Ψ 32 in tRNA^{Phe}. *RNA* **1**, 437-448.
12. Massenet, S., Motorin, Y., Lafontaine, D.L.J., Hurt, E.C., Grosjean, H. & Branlant, C. (1999) Pseudouridine mapping in the *Saccharomyces cerevisiae* spliceosomal U small nuclear RNAs (snRNAs) reveals that pseudouridine synthase Pus1p exhibits a dual substrate specificity for U2 snRNA and tRNA. *Mol. Cell Biol.* **19**, 2142-2154.
13. Becker, H.F., Motorin, Y., Planta, R.J. & Grosjean, H. (1997) The yeast gene YNL292w encodes a pseudouridine synthase (Pus4) catalyzing the formation of Ψ 55 in both mitochondrial and cytoplasmic tRNAs. *Nucleic Acids Res.* **25**, 4493-4499.
14. Ni, J., Tien, A.L. & Fournier, M.J. (1997) Small nucleolar RNAs direct site-specific synthesis of pseudouridine in ribosomal RNA. *Cell* **89**, 565-573.
15. Balakin, A.G., Smith, L. & Fournier, M.J. (1996) The RNA world of the nucleolus: Two major families of small RNAs defined by dif-

- ferent box elements with related functions. *Cell* **86**, 823-834.
- 16.** Ganot, P., Bortolin, M.-L. & Kiss, T. (1997) Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs. *Cell* **89**, 799-809.
- 17.** Ganot, P., Caizergues-Ferrer, M. & Kiss, T. (1997) The family of box ACA small nucleolar RNAs is defined by an evolutionarily conserved secondary structure and ubiquitous sequence elements essential for RNA accumulation. *Genes Dev.* **11**, 941-956.
- 18.** Lafontaine, D.L.J., Bousquet-Antonelli, C., Henry, Y., Caizergues-Ferrer, M. & Tollervey, D. (1998) The box H+ACA snoRNAs carry Cbf5p, the putative rRNA pseudouridine synthase. *Genes Dev.* **12**, 527-537.
- 19.** Meier, U.T. & Blobel, G. (1994) NAP57, a mammalian nucleolar protein with a putative homolog in yeast and bacteria. *J. Cell Biol.* **127**, 1505-1514.
- 20.** Jiang, W., Middleton, K., Yoon, H.-J., Fouquet, C. & Carbon, J. (1993) An essential yeast protein, Cbf5p, binds *in vitro* to centromeres and microtubules. *Mol. Cell Biol.* **13**, 4884-4893.
- 21.** Nurse, K., Wrzesinski, J., Bakin, A., Lane, B.G. & Ofengand, J. (1995) Purification, cloning, and properties of the tRNA Ψ 55 synthase from *Escherichia coli*. *RNA* **1**, 102-112.
- 22.** Zabarjadian, Y., King, T., Fournier, M.J., Clarke, L. & Carbon, J. (1999) Point mutations in yeast *CBF5* can abolish *in vivo* pseudouridylation of rRNA. *Mol. Cell Biol.* **19**, 7461-7472.
- 23.** Cadwell, C., Yoon, H.-J., Zabarjadian, Y. & Carbon, J. (1997) The yeast nucleolar protein Cbf5p is involved in rRNA biosynthesis and interacts genetically with the RNA polymerase I transcription factor RRN3. *Mol. Cell Biol.* **17**, 6175-6183.
- 24.** Frohman, M.A. (1993) Rapid amplification of complementary DNA ends for generation of full-length complementary DNAs: Thermal RACE. *Methods Enzymol.* **218**, 340-356.
- 25.** Frohman, M.A. (1994) Cloning PCR products; in *The Polymerase Chain Reaction* (Mullis, K.B., Ferre, F. & Gibbs, R.A., eds.) pp. 14-37, Birkhäuser, Boston.
- 26.** Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**, 403-410.
- 27.** Thompson, J.D., Higgins, D.G. & Gibson, T.J. (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673-4680.
- 28.** Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389-3402.
- 29.** Aravind, L. & Koonin, E.V. (1999) Novel predicted RNA-binding domains associated with the translation machinery. *J. Mol. Evol.* **48**, 291-302.
- 30.** Dingwall, C. & Laskey, R.A. (1991) Nuclear targeting sequences - a consensus? *Trends Biochem. Sci.* **16**, 478-481.
- 31.** Phillips, B., Billin, A.N., Cadwell, C., Buchholz, R., Erickson, C., Merriam, J.R., Carbon, J. & Poole, S.J. (1998) The Nop60B gene of *Drosophila* encodes an essential nucleolar protein that functions in yeast. *Mol. Gen. Genet.* **260**, 20-29.
- 32.** Heiss, N.S., Knight, S.W., Vulliamy, T.J., Klauck, S.M., Wiemann, S., Mason, P.J., Poustka, A. & Dokal, I. (1998) X-linked dyskeratosis congenita is caused by mutations in a highly conserved gene with putative nucleolar functions. *Nat. Genet.* **19**, 32-38.

- 33.** Knight, S.W., Heiss, N.S., Vulliamy, T.J., Greschner, S., Stavrides, G., Pai, G.S., Les-tringant, G., Varma, N., Mason, P.J., Dokal, I. & Poustka, A. (1999) X-linked dyskeratosis congenita is predominantly caused by mis-sense mutations in the DKC1 gene. *Am. J. Hum. Genet.* **65**, 50-58.
- 34.** Hassock, S., Vetrie, D. & Gianelli, F. (1999) Mapping and characterization of the X-linked dyskeratosis congenita (DKC) gene. *Genomics* **55**, 21-27.
- 35.** Vulliamy, T.J., Knight, S.W., Heiss, N.S., Smith, O.P., Poustka, A., Dokal, I. & Mason, P.J. (1999) Dyskeratosis congenita caused by a 3' deletion and somatic mosaicism in a female carrier. *Blood* **94**, 1254-1260.
- 36.** Heiss, N.S., Girod, A., Salowsky, R., Wiemann, S., Papperkok, R. & Poustka, A. (1999) Dyskerin localizes to the nucleolus and its mislocalization is unlikely to play a role in the pathogenesis of dyskeratosis congenita. *Hum. Mol. Genet.* **8**, 2515-2524.
- 37.** Youssoufian, H., Gharibyan, V. & Qatanani, M. (1999) Analysis of epitope-tagged forms of the dyskeratosis congenita protein (dyskerin): Identification of a nuclear localization signal. *Blood Cells Mol. Dis.* **25**, 305-309.
- 38.** Mitchell, J.R., Wood, E. & Collins, K. (1999) A telomerase component is defective in the human disease dyskeratosis congenita. *Nature* **402**, 551-555.
- 39.** Shay, J.W. & Wright, W.E. (1999) Mutant dyskerin ends relationship with telomerase. *Science* **286**, 2284-2286.