FEBS Letters 354 (1994) 41–44

FEBS 14697

A proposed structure for 'Family 18' chitinases A possible function for narbonin

Andrew F.W. Coulson*

Biocomputing Research Unit, Institute for Cell and Molecular Biology, University of Edinburgh King's Buildings, Mayfield Road, Edinburgh EH9 3JR, UK

Received 21 August 1994; revised version received 20 September 1994

Abstract The sequence of narbonin, a leguminous seed protein with the TIM barrel structure but of unknown function, is significantly similar to endo-β-*N*-acetylglucosaminidase H from *Streptomyces plicatus*. This protein is a member of a family of chitinases, 'Family 18' of the glycosyl hydrolases. It is proposed that the catalytic domain of this family has the TIM barrel structure. It is proposed that narbonin has chitinase activity, or has been derived from a chitinase by loss of function.

Key words: Narbonin; Chitinase; Glycosyl hydrolase; TIM barrel

1. Introduction

Narbonin [1] is a 2S seed protein from a European vetch, *Vicia narbonensis*. No function has been attributed to the protein, but its structure has been determined by X-ray crystallography [2]. Its fold is a monomeric, 8-stranded α/β barrel ('TIM barrel'), the commonest of all globular domain structures [3].

Hennig et al. [2] reported that 'no significant homology to any other proteins' was found in a search of the sequence databank, without describing the method which was used. I have repeated this search, using an implementation [4] of the Smith and Waterman 'Best Local Similarity' algorithm [5] on a high-performance computing system, the Maspar MP-1 [6].

These searches revealed a weak but clear similarity to the sequence of endo-β-*N*-acetylglucosaminidase H from *Streptomyces plicatus*. This paper describes these results, and explores their implications for the function of narbonin and the stucture of a group of glycosyl hydrolases.

2. Materials and methods

2.1. Sequence of narbonin

Hennig et al. [2] give a tentative sequence of 288 residues derived from the electron density in their Table II. The entry in the Brookhaven Data Bank recently deposited by these authors ('pdb1nar.ent') contains a revised sequence of 290 residues:

```
PKPIFREYIG VKPNSTTLHD FPTEIINTET LEFHYILGFA IESYYESGKG
TGTFEESWDV ELFGPEKVKN LKRRHPEVKV VISIGGRGVN TPFDPAEENV
WVSNAKESLK LIIQKYSDDS GNLIDGIDIH YEHIRSDEPF ATLMGQLITE
LKKDDDLNIN VVSIAPSENN SSHYQKLYNA KKDYINWVDY QFSNQQKPVS
TDDAFVEIFK SLEKDYHPHK VLPGFSTDPL DTKHNKITRD IFIGGCTRLV
QTFSLPGVFF WNANDSVIPK RDGDKPFIVE LTLQQLLAAR
```

This sequence was used for all the database searches reported here. No sequence for narbonin appears in the PIR [7] or Swissprot [8] databases. 'OWL' [9] has four sequences derived from DNA data. Three of these, with minor differences between them, are from *Vicia narbonensis*; the fourth from *Vicia pannonica*. None is identical to the PDB sequence; all end with LAKR instead of LAAR. The closest to the PDB sequence is 'VNNAF6NB', with one other difference (QFGN for QFSN at position 91).

2.2. Database searching

All searches were made using the program 'MPSRCH', which implements an exhaustive search method [4] on MasPar machines. This method compares the query individually with every member of the database (in this case Swissprot Release 29) using the Smith and Waterman [5] algorithm. Scoring tables were constructed according to the Dayhoff prescription [10], which generates a table appropriate to any overall evolutionary distance. Distances are expressed in 'PAMs', where 1 PAM corresponds to 1 accepted point mutation per 100 residues of sequence. The statistical significance of alignments reported by the search program is assessed by a semi-empirical method [11] which takes account of the observed distribution of similarity scores in comparisons with unrelated sequences.

2.3. Multiple sequence alignment

Multiple sequence alignment used a program (A.F.W. Coulson, unpublished) which attempts to identify strongly related subsequences which can be unequivocally and uniquely identified in all the sequences under examination [12]. These subsequences are pinned together, and alignments are performed between the 'nodes' using an extension of the Corpet 'MULTALIN' method [13].

3. Results and discussion

3.1. Database searches with the narbonin sequence

The problem in reliably recognising distant sequence similarities in database searches is to distinguish them from the higher-scoring tail of the distribution of chance similarities. The nature of these 'noise' similarities depends on the scoring table which is used. A table derived for a low PAM value (which stresses the significance of matches) will produce oligomers with a high proportion of matches, and few or no gaps. The 'flatter' tables at high PAM values make a smaller distinction between matches and conservative substitutions, and the high scoring 'chance' results are longer and contain more gaps. It is unlikely that any given sequence, unrelated to the query, will contain both types of chance similarity, and it is usually found that quite different sets of unrelated proteins are found at the tops of output lists of searches performed with widely differing PAM values. Hits on related sequences are more likely to appear across the whole range of PAM values.

*Corresponding author. Fax: (44) (31) 665 3870;
E-mail: a.coulson@ed.ac.uk

SSDI 0014-5793(94)01084-6

	Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.	
*	1	142	7.7	313	1	EBAG_STRPL	ENDO-BETA-N-ACETYLGLU	4.74e-05	*
*	2	123	6.6	619	1	CHIT_STRLI	CHITINASE C PRECURSOR	9.12e-03	*
	3	115	6.2	365	1	VSI3_REOVL	SIGMA 3 PROTEIN (MAJO	7.39e-02	
	4	113	6.1	365	1	VSI3_REOVJ	SIGMA 3 PROTEIN (MAJO	1.23e-01	
*	5	113	6.1	610	1	CHIT_STRPL	CHITINASE 63 PRECURSO	1.23e-01	*
	6	112	6.0	281	1	XYLF_PSEPU	2-HYDROXYMUCONIC SEMI	1.59e-01	
	7	111	6.0	283	1	DMPD_PSEPU	2-HYDROXYMUCONIC SEMI	2.04e-01	
	8	110	5.9	404	1	NANH_CLOSO	SIALIDASE PRECURSOR (	2.62e-01	
	9	108	5.8	652	1	MX1_RAT	INTERFERON-INDUCED GT	4.30e-01	
	10	108	5.8	383	1	GP39_HUMAN	CARTILAGE GLYCOPROTEI	4.30e-01	
	11	106	5.7	365	1	VSI3_REOVD	SIGMA 3 PROTEIN (MAJO	7.02e-01	
	12	105	5.7	502	1	SYFB_YEAST	PHENYLALANYL-TRNA SYN	8.95e-01	
	13	105	5.7	562	1	EST1_CAEBR	GUT ESTERASE PRECURSO	8.95e-01	
*	14	104	5.6	699	1	CHI1_BACCI	CHITINASE A1 PRECURSO	1.14e+00	*
	15	104	5.6	805	1	AMPN_RABIT	AMINOPEPTIDASE N (EC	1.14e+00	
	16	101	5.4	462	1	VSI1_REOVJ	SIGMA 1 PROTEIN PRECU	2.32e+00	
	17	101	5.4	473	1	DLDH_ECOLI	DIHYDROLIPOAMIDE DEHY	2.32e+00	
	18	101	5.4	1528	1	KEM1_YEAST	STRAND EXCHANGE PROTE	2.32e+00	
	19	100	5.4	812	1	TOP1_SCHPO	DNA TOPOISOMERASE I (	2.94e+00	
	20	99	5.3	1003	1	PUR2_CHICK	PHOSPHORIBOSYLAMINE--	3.71e+00	
*	21	99	5.3	562	1	CHIA_SERMA	CHITINASE A PRECURSOR	3.71e+00	*
	22	98	5.3	324	1	YACH_ECOLI	HYPOTHETICAL 36.2 KD	4.67e+00	
	23	97	5.2	764	1	PA_BACAN	PROTECTIVE ANTIGEN PR	5.87e+00	
	24	97	5.2	375	1	NUEM_NEUCR	NADH-UBIQUINONE OXIDO	5.87e+00	
	25	96	5.2	786	1	EXOP_RHIME	SUCCINOGLYCAN BIOSYNT	7.37e+00	
*	26	96	5.2	820	1	CHIA_ALTSO	CHITINASE A PRECURSOR	7.37e+00	*
	27	96	5.2	428	1	MTBA_BACAR	MODIFICATION METHYLAS	7.37e+00	
	28	96	5.2	220	1	VM02_VACCC	PROTEIN M2.	7.37e+00	
	29	95	5.1	180	1	ARF1_HUMAN	ADP-RIBOSYLATION FACT	9.24e+00	
	30	95	5.1	104	1	YJCH_ECOLI	HYPOTHETICAL 11.7 KD	9.24e+00	
	31	95	5.1	180	1	ARF3_HUMAN	ADP-RIBOSYLATION FACT	9.24e+00	
	32	95	5.1	382	1	NANH_CLOPE	SIALIDASE (EC 3.2.1.1	9.24e+00	
	33	94	5.1	700	1	CH60_PLAFG	MITOCHONDRIAL CHAPERO	1.16e+01	
	34	94	5.1	877	1	DPO1_BACCA	DNA POLYMERASE I (EC	1.16e+01	
	35	94	5.1	494	1	PRE_STRAG	PLASMID RECOMBINATION	1.16e+01	
	36	94	5.1	659	1	MX3_RAT	INTERFERON-INDUCED GT	1.16e+01	
*	37	94	5.1	267	1	EBAG_FLASP	ENDO-BETA-N-ACETYLGLU	1.16e+01	*
	38	94	5.1	659	1	MX2_RAT	INTERFERON-INDUCED GT	1.16e+01	
	39	93	5.0	270	1	RFA2_HUMAN	REPLICATION PROTEIN A	1.44e+01	
	40	93	5.0	181	1	ARF1_DROME	ADP-RIBOSYLATION FACT	1.44e+01	
	41	93	5.0	656	1	TOP3_YEAST	DNA TOPOISOMERASE III	1.44e+01	
	42	93	5.0	2233	1	RRPL_PI3H4	RNA POLYMERASE BETA S	1.44e+01	
	43	92	5.0	690	1	VTER_EBV	PROBABLE DNA PACKAGIN	1.80e+01	
	44	92	5.0	271	1	ATBP_STAAU	POTENTIAL ATP-BINDING	1.80e+01	
	45	92	5.0	562	1	EST1_CAEEL	GUT ESTERASE PRECURSO	1.80e+01	
	46	92	5.0	2339	1	RPC1_PLAFA	DNA-DIRECTED RNA POLY	1.80e+01	
	47	92	5.0	253	1	ADH2_DROMO	ALCOHOL DEHYDROGENASE	1.80e+01	
	48	92	5.0	539	1	CH61_STRAL	60 KD CHAPERONIN 1 (P	1.80e+01	
	49	91	4.9	703	1	ARYB_MANSE	ARYLPHORIN BETA SUBUN	2.23e+01	
	50	91	4.9	540	1	CH60_MYCTU	60 KD CHAPERONIN (PRO	2.23e+01	

```
RESULT    1
ID   EBAG_STRPL     STANDARD;      PRT;   313 AA.
DE   ENDO-BETA-N-ACETYLGLUCOSAMINIDASE H PRECURSOR (EC 3.2.1.96) (MANNOSYL-
DE   GLYCOPROTEIN ENDO-BETA-N-ACETYL-GLUCOSAMINIDASE H) (DI-N-
DE   ACETYLCHITOBIOSYL BETA-N-ACETYLGLUCOSAMINIDASE H).

  DB  1;  Score      142;  Match 21.9%;  Predicted No. 4.74e-05;
  Matches     43;  Conservative    56;  Mismatches  85;  Indels  12;  Gaps  12;
           *. *. *.. *.   .  .      *  .  ** .  *..*  * . .  ..*.    .
Db      56 YVEVN-NNSMLNVGKYTLADGGGNAFDVAVIFAANINYDTGTKTAYLHFNENVQRVLDNA 114
Qy       8 YIGVKPNSTTLHDFPTEIINTETLEFHYILGFAIESYYESGKGTGTFEESWDVELFGPEK 67

           *  ..  ..  .**..*. *    . * . *...    . **. *    .**. *. ..
Db     115 VTQIRPLQQQGIKVLLSVLGNHQGAGFANFPSQQAA-SAFAKQ-LSDAVAKYGLDGVDFD 172
Qy      68 VKNLKR-RHPEVKVVISIGGRGVNTPF-D-PAEENVWVSNAKESLKLIIQKYSDDSGNLI 124

           *     *.  . . *  .   .*.* *. .  *  *.. .  *. .  * *  .  . *
Db     173 DEYA-EYGNNGTAQPNDSSFVHLVTALRANMPDKIISLYNIGPAASRLS-YGGVDVSDKF 230
Qy     125 DGIDIHYEHIRSDEPFATLMGQLITELKKD-DDLNINVVSIAPSENNSSHYQKLYNAKK- 182

           **  *  * ... * *
Db     231 DY-AWNPY-YGTWQVP 244
Qy     183 DYINWVDYQFSNQQKP 198
```

Fig. 1. Highest scoring 50 local similarities when the sequence of narbonin was compared with Swissprot Release 29, with a scoring table corresponding to 178 PAMs, and a gap penalty of −10. Asterisks indicate the similarities to identified chitinases. 'Pred.No.' is the expected number of similarities achieving the observed score by chance. In the alignment, the database sequence is in the upper rows.

Database searches were performed with PAM values ranging between 40 and 300; members of a family of chitinases were found amongst the top 50 search results across the whole range. Values of the search parameters (PAM value and gap penalty score) were then chosen iteratively to maximise the significance of the strongest local similarity detected. The optimum values were found to be 178 PAM, and a gap penalty of −10. Fig. 1 shows the top of the search output, and the highest scoring alignment under these conditions. The 'expected frequency' of 4.5×10^{-5} (odds of about 20,000 to 1 against this occurring by

chance) indicates a weak but significant similarity to the *Streptomyces* chitinase [14], and this is reinforced by the appearance of other chitinases lower in the output list. Similarities to other glycosyl-transferring enzymes (such as the sialidases at positions 8 and 32 in the output list) may also represent a real relationship between the proteins, but other apparent similarities presumably arise by chance.

3.2. Multiple alignment of related sequences

Examination of the protein sequence database showed that the sequence of the corresponding chitinase from a *Flavobacterium* was also known [15]. In order to examine the sequence conservation a multiple sequence alignment was performed with both chitinases, and the narbonins from *Vicia narbonenesis* and from *Vicia pannonica*. This alignment is shown as Fig. 2. The program identifies two short subsequences with a high proportion of matches as 'nodes'; the strongest sequence conservation is around and between these nodes, but the sequence similarity extends to include the major part of all four sequences.

The TIM barrel in narbonin is not completely regular. The connection between the first two β-strands is non-helical, and extra secondary structure elements (two β-hairpins and two short helices; see Fig. 2) are inserted into strands at the C-terminal end of the barrel. The sequence similarity in the alignment is strongest in the first two thirds of the proteins, and here the most conserved regions in the alignment correlate with the presence of secondary structure, and especially with strands of β-sheet. The proteins are about the same length, but it is possible that the pattern of insertions of loops into the core structure differs between the two proteins sufficiently to disrupt the alignment in the C-terminal third of the sequence.

On the basis of this protein sequence similarity, I propose that the chitinases also have the TIM barrel structure as the main structural and functional unit.

3.3. 'Family 18 glycosyl hydrolases'

The glycosyl hydrolases have been classified by Henrissat [16,17] on the basis of sequence similarity, and in this classification the Streptomyces chitinase is assigned to Family 18. No structure has been reported for any member of the family. The proteins are very diverse in length, and the narbonin-related chitinases are amongst the shortest. In some cases it is known that the length diversity arises from the presence of domains with additional functions – for example, many of the sequences contain protein sequence motifs characteristic of binding sites for relatively small molecules. Thus, the 'chitinase D' from *Bacillus circulans* [18] and the 'chitinase 63' from *Streptomyces*

```
Narbonin    .......... .......... .......... .......... ........PK
Narbonin2   .......... .......... .......... .......... .......MPK
ebagstrpl   MFTPVRRRVR TAALALSAAA ALVLGSTAAS GASATPSPAP APAPAPVKQG
ebagflasp   .......... .......... .......... .......... ...ATPTKSG

            *    *  *   *  *         *       *     **     *  ** *
Narbonin    PIFREYIGVK PNSTTLHDFP TEIINTETLE FHYILGFAIE SYYESGKGTG
Narbonin2   PIFREYIGV. PNSETLHDFP HEIIDTENLE FHFILGFATE SYYESGKSTG
ebagstrpl   PTSVAYVEV. NNNSMLNVGK YTLADGGGNA FDVAVIFAAN INYDTGTKTA
ebagflasp   PTSIAYVEV. NNDQLANVGR YQLANGA.NA FDVAIIFAAN INWNGSK..A
Structure    ssssss  I III        IIII     s sssssrrrrr rrIIII rrr

                  *        *          ---- --- *       *
Narbonin    TFEESWDVEL FGPEKVKNLK RRHP.EVKVV ISIGGRGVNT PFDPAEENVW
Narbonin2   NFEESWDVEL FGPENVKNLK TKHP.EVKVV ISIRGHDDKT PFDPDEENIW
ebagstrpl   YLHFNENVQR VLDNAVTQIR PLQQQGIKVL LSVLGNHQGA GFANFPSQQA
ebagflasp   VLYNNENVQA TLDDAATQIR PLQAKGIKVS LSILGNHQGA GIANFPTQAA
Structure   rrrr   KKK K hhhhhhhh hhII IIsss sssss IIII      IIIIh

               * *      **        -----     *            *    **
Narbonin    VSNAKESLKL IIQKYSDDSG NLIDGID..I HYEHIRSD.. .EPFATLMGQ
Narbonin2   VWKAVKSLKQ IIKKYRNESG NMIDGID..I NYEHINSDD. .ELFVNCIGQ
ebagstrpl   ASAFAKQLSD AVAKY..... .GLDGVDFDD EYAEYGNNGT AQPNDSSFVH
ebagflasp   AEDFAAQVSA TVSKY..... .GLDGVDLDD EYSDYGTNGT PQPNQQSIGG
Structure   hhhhhhhhhh hhhh qqIIqq    sss  s s             hhhhhhh

            *** *      *  *    *  *     **       * **   *  *    *
Narbonin    LITELKKD.D DLNINVVSIA PSENNSSHYQ KLYNAKKDYI NWVDYQFSNQ
Narbonin2   VIRELKKD.D DLNIDVVSIA PSENNQSSNQ KLYNANTDYI NWVDYQFSNQ
ebagstrpl   LVTALRANMP DKIISLYNIG PAASRLSYGG VDVSDKFDY. AWNPYYGTWQ
ebagflasp   LISALRADVP GKLISFYDIG PASSALSSSS STIGSKLDY. AWNPYYGTYS
Structure   hhhhhhh         ssss   KKKKhhhhh hhhhhKKKK   ssssIIII

            *      *         *          *      *          *
Narbonin    QKPVSTDDAF VEIFKSLEKD YHPHKVLPGF STDPLDTKHN KITRDIFIGG
Narbonin2   VKPVTTVDAF VDIYNSLVKD YDAGKVLPGF NTEPLDIKDT KTTRDTFIRG
ebagstrpl   VPGIA.LPK. ....AQLSPA AVEIGRTSRS TVADLARRTV DEGYGVYL..
ebagflasp   APSIPGLDK. ....SRLSAA AVDVQNTPQS TAVSLAQRTK ADGYGVFM..
Structure         hhhh hhhhhhhhhh  JJJJssssss s  hhhhhh     hhhhhhh

               * * **          *      *
Narbonin    CTRLVQTFSL PGVFFWNAND SVIPKRDGDK PFIVELTLQQ LLAAR
Narbonin2   CTKLLQTSSL PGVFIWNAND SVIPQRDDDT PFIVELKLQQ LLAKR
ebagstrpl   ......TYNL DGGDRTADVS AFTRELYGSE AVRT....P. .....
ebagflasp   ......TYNL PDGDVSPYVS SMTKVLYGQA ATY.....H. .....
Structure   hhhhhh      sssss hhh h    JJJJ    hhhhhhhh hhh
```

Fig. 2. Multiple sequence alignment of narbonins (from *Vicia narbonensis* and from *Vicia pannonica*) and closely related chitinases (from *Streptomyces plicatus* and *Flavobacterium sp.*); hyphens indicate the two regions in which significant similarities can be unequivocally aligned across all the proteins, and asterisks other locations at which there is at least one match between a narbonin and a chitinase. 'Structure' indicates the secondary structure assignment in narbonin: h, helix; s, β strands forming part of the TIM barrel; r and q, β-hairpins; I,J,K, turns of Types I,II and III.

plicatus [19] both have a fibronectin Type III domain on the N-terminal side of the catalytic region; the second of these proteins also has a cellulose binding domain still further upstream. Corresponding domains have not been identified in the chitinase from *Brugia malayi* [20], but this protein contains a threefold approximate tandem repeat of a Glu- and Tyr-rich 14-mer on the C-terminal side of the catalytic region. (These three proteins are the closest found in database searches to the narbonin-related chitinases.)

The presence of these diverse additional domains makes multiple sequence alignments of the complete sequences unconvincing. Limiting the aligned regions to the presumed catalytic domains was more satisfactory (data not shown) and showed the highest sequence conservation in a central portion of about 60 residues. This region includes the highly conserved nodes and the intervening sequence in Fig. 2 which represent two successive strands of the TIM barrel and the α-helix which links them. I propose that the catalytic core structure of the Family 18 glycosyl hydrolases is an 8-stranded α/β barrel, to which additional domains and perhaps extra loops are added to confer substrate-binding and other properties.

3.4. Narbonin function

A significant similarity between the sequences of two proteins generally indicates a functional as well as a structural relationship between them. It is therefore plausible to suggest that narbonin may have chitinase or a closely related glycosyl hydrolase activity. Chitinases are found in seeds, where they exert an anti-fungal protective role. However, Henrissat [16] assigned all these seed chitinases to 'Family 19'. One X-ray crystal structure of a Family 19 protein, from barley [21], is known. This structure is largely α-helical (10 helices and 47% of the sequence), and these authors recognised no similarity in either the structure or the sequence to any protein outside Family 19. Subsequently Holm and Sander [22] pointed out that many of the secondary structural elements could be superimposed on those in T4 lysozyme. Proteins from two other families of glycosyl hydrolases are now known to have the TIM barrel structure – α-amylases and cyclodextrin gluconotransferases from Family 13 [18], and a xylanase from Family 10 (Jenkins, J., personal communication). Cellobiohydrolase II (Family 6) has a related fold, in which however the 'barrel' appears to be open-sided [23] (only the α-carbon co-ordinates have been published).

Hennig et al. based their suggestion that narbonin was without enzymic activity on the absence from the crystal structure of plausible active site residues at the C-terminal end of the β-barrel. If it is without catalytic activity, narbonin may show some analogy to the Drosophila vitellogenins [24]. In this case, we showed that the insect yolk proteins had been derived from lipid acylases with the loss of active site residues. The proteins are therefore devoid of enzymic activity, but retain a lipid binding site (which is used for storage and release of lipid hormones). If this analogy were correct, narbonin would show substrate binding activity, in the absence of catalysis.

References

[1] Schlesier, B., Manteuffel, R., Rudolph, A. and Behlke, J. (1978) Biochem. Physiol. Pflanz. 173, 420–428.
[2] Hennig, M., Schlesier, B., Dauter, Z., Pfeffer, S., Betzel, C., Hohne, W.E. and Wilson, K.S. (1992) FEBS Lett. 306, 80–84.
[3] Farber, G.K. (1993) Curr. Opin. Struct. Biol. 3, 409–412.
[4] Coulson, A.F.W., Collins, J.F. and Lyall, A. (1987) Comput. J. 30, 420–424.
[5] Smith, T.F. and Waterman, M.S. (1981) J. Mol. Biol. 141, 645–648
[6] MasPar Computer Corporation, Sunnyvale, California.
[7] Barker, W., George, D.G., Mewes, H.W., Pfeiffer, F. and Tsugita, A. (1993) Nucleic Acids Res. 21, 3089–3092.
[8] Bairoch, A. and Boeckmann, B. (1993) Nucleic Acids Res. 21, 3093–3096.
[9] Bleasby, A.J. and Wootton, J.C. (1990) Prot. Eng. 3, 153–159.
[10] Dayhoff, M.O., Schwarz, R.M. and Orcutt, B.C. (1978) in: Atlas of Protein Sequence and Structure, Vol. 5, supplement 3 (Dayhoff, M.O. (Ed.) NBRF, Washington, pp. 345–352.
[11] Collins, J.F. and Coulson, A.F.W. (1990) Methods Enzymol. 183, 474–486.
[12] Vingron, M. and Argos, P. (1990) Prot. Eng. 3, 565–569.
[13] Corpet, F. (1988) Nucleic Acids Res. 16, 10881–10890.
[14] Robbins, P.W., Trimble, R.B., Wirth, D.F., Hering, C., Maley, F., Maley, G.F., Das, R., Gibson, B.W., Royal, N. and Biemann, K.R. (1984) J. Biol. Chem. 259, 7577–7583.
[15] Takegawa, K., Mikami, B., Iwahara, S., Morita, Y., Yamamoto, K. and Tochikura, T. (1991) Eur. J. Biochem. 202, 175–180.
[16] Henrissat, B. (1991) Biochem. J. 280, 309–316.
[17] Henrissat, B. and Bairoch, A. (1993) Biochem. J. 293, 781–788.
[18] Watanabe, T., Oyanagi, W., Suzuki, K., Ohnishi, K. and Tanaka H., (1992) J. Bacteriol. 174, 408–414.
[19] Robbins, P.W., Overbye, K., Albright, C.,Benfield, B. and Pero, J. (1992) Gene 111, 69–76.
[20] Fuhrman, J.A., Lane, W.S., Smith, R.F., Piessens, W.F. and Perler, F.B. (1992) Proc. Natl. Acad. Sci. USA 89, 1548–1552.
[21] Hart, P.J., Monzingo, A.F., Ready, M.P., Ernst, S.R. and Robertus, J.D. (1993) J. Mol. Biol. 229, 189–193.
[22] Holm, L. and Sander, C. (1994) FEBS Lett. 340, 129–131.
[23] Rouvinen, J., Bergfors, T., Teeri, T., Knowles, J.K.C. and Jones, T.A. (1990) Science 249, 380–386.
[24] Bownes, M., Shirras, A., Blair, M., Collins, J.F. and Coulson, A.F.W. (1988) Proc. Natl. Acad. Sci. USA 85, 1554–1557.