Did cyclodextrin glycosyltransferases evolve from α-amylases?

Gabriel del-Rio, Enrique Morett, Xavier Soberon*

Instituto de Biotecnologia/UNAM, Apdo. Postal 510-3, Cuernavaca, Morelos 62271, Mexico

Received 3 June 1997; revised version received 8 September 1997

Abstract The hydrolytic enzymes, α-amylases, and the cyclodextrin glycosyltransferases (CGTases) are key enzymes in the depolymerization of starch. These two groups of enzymes are evolutionarily related. We propose that the transferase activity is likely to have evolved from an ancestral hydrolase. Sequence analysis provides support for this hypothesis. Consequently, we have conducted an experimental study to test the possible adaptive value for evolving a CGTase. We found that when an α-amylase and a CGTase are combined more glucose is generated from starch than would be expected from the independent action of either of these enzymes. Thus, we propose that the biological role of CGTases is to work in concert with α-amylases for the efficient saccharification of starch. This observation can be useful in industrial processes aimed at producing syrups with high contents of glucose or maltose.

Key words: α-Amylase; Cyclodextrin glycosyltransferase; Evolution; Divergence

1. Introduction

Starch is one of the most abundant biopolymers on earth and it is the main source of energy for a wide variety of living organisms [1]. This molecule is built of units of glucose, linked by either α(1,4) or α(1,6) glycosidic bonds. To obtain glucose from starch different enzymes are used, most of which are grouped in the α-amylase family [2] (see Table 1). This group includes enzymes that display specificity for either of the two types of bond, have either a retaining or an inverting mechanism (depending on whether they retain or invert the anomeric structure of the substrate [3]), and whose activities classify them as either hydrolases (EC 3.2.1) or transferases (EC 4.2.1). By combining these properties, there would be eight theoretically possible types of hydrolases and four of transferases. However, these numbers must be reduced considering the nature of the reactions involved [3], so only three types of hydrolyses and two types of transferases would be expected to exist in nature (not including dextrinases). All these five activities have been found (see Table 1). Thus, the α-amylase family is well suited for studying the way different enzyme activities give rise to each other; particularly the retaining enzymes (α-amylases and cyclodextrin glycosyltransferases (CGTases)), since they are biochemically and structurally well characterized (see below).

Through the analysis of the primary structure of the retaining enzymes, it has been found that although there is limited conservation of their overall primary structure, both hydrolases and transferases share sequence motifs which contain the catalytic residues. Thus, it has been assumed that all of them use the same catalytic machinery [4]. Additionally, the three-dimensional structure of six α-amylases and three CGTases (see Section 2) has already been determined (see legend in Table 1), showing that they share a $(\beta/\alpha)_8$-barrel domain [5]. Such evidence led different authors to suggest that these enzymes probably diverged from a common ancestor [2,5,6]. Taking advantage pf the fact that this is one of the few examples of two different, well characterized, enzyme activities related by the same catalytic machinery and fold [7,8], it is possible to carry out a detailed molecular phylogenetic analysis that can shed some light on the evolution of their activities.

Many successful examples of in vitro alteration of enzyme specificity have recently been reported [9–11], but few of a change to a new activity [12,13]. This may reflect the scarcity of our knowledge of the latter subject and that is more difficult to evolve a new enzyme activity than a new specificity [11]. In this study we propose an evolutionary model to account for the divergence of two enzyme activities in the α-amylase family: hydrolase and transferase. We also propose a possible biological role for CGTases.

2. Materials and methods

2.1. Sequence analysis

A multiple sequence alignment including 49 sequences of α-amylases and 22 of CGTases was obtained from the FSSP+HSSP databases (http://www.embl-heidelberg.de/dali/dali.html). This alignment was kindly provided by Dr. Lissa Holm and it is based on six structures of α-amylases (from fungi (2aaa), bacteria (1amg and 1bpl), mammals (1smd and 1ppi) and plants (1amy)) and three of CGTases (from bacillus (1cdg, 1ciu, 1cyg)). Known sequences homologous to these structures (but having less than 25% identity) were included. The SwissProt names, and the corresponding accession numbers indicated in parentheses, of α-amylase sequences employed are: amy2_horvu (P04063), amy3_horvu (P04747), amy4_horvu (P04748), amy5_horvu (P04749), amy6_horvu (P04750), amy1_horvu (P00693) (*Hordeum vulgare*); amy1_orysa (P17654), am2a_orysa (P27935), amc2_orysa (P27941), am3d_orysa (P27933), am3b_orysa (P27937), am3c_orysa (P27939), am3e_orysa (P27934), am3a_orysa (P27932), amc1_orysa (P27940) (*Oryza sativa*); amy3_wheat (P08117), (*Triticum aestivum*); amya_vigmu (P17859) (*Vigna mungo*); amy2_ecoli (P26612) (*Escherichia coli*); amt4_psest (P13507) (*Pseudomonas stutzeri*); amt4_psesa (P22963) (*Pseudomonas Saccharophila*); amyp_pig (P00690) (*Sus scrofa*); amyc_human (P19961), amyp_human (P04746), amys_human (P04745) (*Homo sapiens*); amyp_mouse (P00688), amys_mouse (P00687) (*Mus musculus*); amyp_rat (P00689) (*Rattus norvegicus*); amya_drome (P08144) (*Drosophila melanogaster*); amy_trica (P09107) (*Tribolium castaneum*); amy_altha (P29957) (*Alteromonas haloplanktis*); amya_aerhy (P41131) (*Aeromonas hydrophila*); amy_thecu (P29750) (*Thermonospora curvata*); amy_strtl (P27350), amy_strvl (P22998) (*Streptomyces thermoviolaceus*); amy_strgr (P30270) (*Streptomyces griceus*); amy_strlm (P09794) (*Streptomyces limosus*); amy_strhy (P08486) (*Streptomyces hygroscopicus*); amy_strli (Q05884) (*Streptomyces lividans*); amya_aspor (P10529) (*Aspergillus oryzae*); amy_aspsh (P30292) (*Aspergillus shirousami*); amy1_schoc (P19269) (*Schwanniomyces occidentalis*); amy1_sacfi (P21567) (*Sac-*

*Corresponding author. Departamento de Reconocimiento Molecular y Bioestructura, Instituto de Biotecnologia/UNAM, Apdo. Postal 510-3, Cuernavaca, Morelos 62271, Mexico. Fax: (52) (73) 172388. E-mail: soberon@ibt.unam.mx

PII S0014-5793(97)01192-7

charomycopsis fibuligera); amy1_schpo (Q09840) (*Schizosaccharomyces pombe*); amyb_bacpo (P21543) (*Bacillus polymixa*); amy_bacli (P06278) (*Bacillus licheniformis*); amy_bacam (P00692) (*Bacillus amyloliquefaciens*); amt6_bacs7 (P19571) (*Bacillus* sp. strain 707); amy_bacst (P06279) (*Bacillus stearothermophilus*); amy2_salty (P26613) (*Salmonella typhimurium*). For CGTases: cdgu_bacci (P43379), amy_bacci (P08137), cdgt_bacci (P30920) (*Bacillus circulans*); amyr_bacs8 (P17692) (*Bacillus* sp. strain B1018); cdgt_bacsp (P30921) (*Bacillus* sp. strain 17-1); cdgt_bacs0 (P05618) (*Bacillus* sp. strain 1011); cdgt_bacs3 (P09121) (*Bacillus* sp. strain 38-2); cdgt_bacli (P14014) (*Bacillus licheniformis*); cdgt_bacss (P31747) (*Bacillus* sp. strain 6.6.3); amy_thetu (P26827) (*Thermoanaerobacter thermosulfurogenes*); cdg2_bacma (P31835), cdg1_bacma (P04830) (*Bacillus macerans*); cdgt_bacst (P31797), amym_bacst (P19531) (*Bacillus stearothermophilus*); cdgt_bacoh (P27036) (*Bacillus ohbensis*); cdgt_bacs2 (P31746) (*Bacillus* sp. strain 1-1); cdgt_klepn (P08704) (*Klebsiella pneumoniae*).

The alignment was manually edited in regions where ambiguity was observed. In order to include the neopollulanase sequence from *Bacillus stearothermophilus*, we used the alignments provided by two different fold recognition algorithms, TOPITS (http://www.embl-heidelberg.de/predictprotein) and UCLA-DOE (http://www.mbi.ucla.edu/people/frsvr/frsvr.html). This neopollulanase was chosen because it presents both specificities and activities found in the α-amylase family [14].

2.2. Phylogenetic analysis

Phylogenies were reconstructed using distance-based and parsimony methods. The phylogenetic inference programs employed are part of the PHYLIP package, version 3.5, for UNIX (J. Felsentein, 1993, Department of Genetics, University of Washington, Seattle. Distributed by the author). We used the programs PROTDIST with the PAM250 matrix, and FITCH to produce a tree of genetic distance. Global rearrangements and 10 jumbles were used to minimize the effect of sequence order. To generate the most parsimonious tree we used the program PROTPARS. When several equally parsimonious trees were found, a consensus tree was proposed using the program CONSENSE. Bootstrapping analysis was performed using the SEQBOOT program, generating 1000 permutation resamplings of our sequence alignment. Global rearrangements and 10 jumbles were also performed for each of the 1000 bootstrapped data sets. The above mentioned programs were run on a Silicon Graphics workstation. For phylogenetic tree representations, we used the program TreeView version 1.4 [15] on a Macintosh computer.

2.3. Enzyme activity

The depolymerization of starch was followed by either determination of reducing sugars using the reagent 3,5-dinitrosalicylic acid [16], or thin layer chromatography on silica gel aminated plaques, using the solvent system water/ethanol/butanol (2:5:3, v/v). The glucose concentration was determined with a kit obtained from Boehringer Mannheim (USA), which includes two enzymes, hexokinase and glucose 6-phosphate dehydrogenase.

α-Amylase from *Bacillus licheniformis* was obtained from BIOTECSA S.A. C.V. (Mexico), and the CGTase from *Bacillus macerans* was obtained from AMANO International Enzyme Co., Inc. (USA). Soluble starch used in this study was from Sigma (USA). Protein was determined by the Bradford method (Bio-Rad laboratories, USA). The reactions of starch depolymerization with α-amylase and/or CGTase were conducted at pH 6.0 (acetate buffer 30 mM, $CaCl_2$ 3.0 mM), and at 60°C. These conditions were optimal for both enzymes. The reactions were carried out for 5 h, and samples were taken every hour. Since the α-amylase and CGTase activities are not easily comparable, we used different enzyme mixtures for our study: α-amylase alone, 142 μg/ml; CGTase alone, 70 μg/ml; α-amylase(1):CGTase(1) mix, 142 μg/ml:70 μg/ml; α-amylase(1):CGTase(2) mix, 142 μg/ml:140 μg/ml; and α-amylase(2):CGTase(1) mix, 142 μg/ml:35 μg/ml. The final reaction volume was 10 ml, and contained 1.2 mg/ml of soluble starch.

3. Results and discussion

3.1. Phylogeny

The distance tree generated with all the α-amylase and CGTase sequences resulted in two well defined groups comprising each type of enzyme (data not shown), except for three putative α-amylases which were included with CGTases: amyr_bacs8, amy_thetu and amy_bacci. Two of them (amyr_bacs8 and amy_thetu) may actually be CGTases since they harbor a signature of CGTases [17], and have sequence lengths similar to CGTases. On the other hand, amy_bacci (528 aa) has the sequence length of α-amylases (around 500 aa). This amylase aligns only with the three domains shared by α-amylases and CGTases (this includes the $(\beta/\alpha)_8$-barrel domain, B domain, and the C-terminal domain in α-amylases [18]). It has been shown that the two additional domains found in CGTases determine their ability to make cyclodextrins [19,20]. Accordingly, the amylase from *Bacillus circulans* (amy_bacci) does not form cyclodextrins. Hence, we believe that the sequence signature criterion might not be sufficient to distinguish between α-amylases and CGTases.

To perform a more exhaustive phylogenetic analysis, we selected as representative members those for which structures were available (see Section 2). Based on a published thermodynamic study of enzyme activity evolution [21], which considers the independent evolution of K_m and k_{cat}, and applying it to the case of α-amylases, we concluded that the ancestor of this group must have been an enzyme with low specificity and low activity. These are characteristics present in neopollulanases. These enzymes, based on other arguments, had already been proposed as the ancestor for all of the activities in the α-amylase family [14]. The topologies of phylogenetic trees obtained through genetic distance and parsimony analyses were similar, and with both methods α-amylases grouped apart from CGTases (data not shown). As shown in Fig. 1A, the neopollulanase from *B. stearothermophilus* (neop_bacst) is grouped with α-amylases, from which it appears to be a relatively recent derivation. Thus, instead of neopollulanases being the ancestor of both α-amylases and CGTases as previously proposed, it seems that neop_bacst comes from an α-amylase. On the other hand, our phylogenetic analysis suggests that the α-amylase group is more diverse, and thus older than the CGTase group. This is consistent with there being α-amylases in a much broader range of taxonomic groups. This notion would imply that CGTases, too, evolved from α-amylases.

Additionally, we carried out a similar phylogenetic analysis using an alignment previously published for α-amylases and CGTases [17], which proposed a couple of amylases as *intermediates* between these two groups. According to the fold recognition algorithms used (see Section 2), both *intermediate* enzymes present at their N-terminus the three domains shared by α-amylases and CGTases (data not shown), but posses sequence lengths similar to or larger than CGTases (anc_dth,

Table 1
Starch processing enzymes

Class of enzyme activity	Specificity	Mechanism involved in amylases found in nature[a]
Hydrolases	α(1,4)	Inverting (β-amylase) Retaining (α-amylase, neopollulanase)
Hydrolases	α(1,6)	Retaining (oligo α(1,6)-glycosidase, neopollulanase)
Transferases	α(1,4)	Retaining (CGTase, neopollulanase)
Transferases	α(1,6)	Retaining (neopollulanase)

[a]The name of a representative enzyme with a known structure [29] is given in parentheses. The type of mechanism, inverting or retaining, is also indicated.

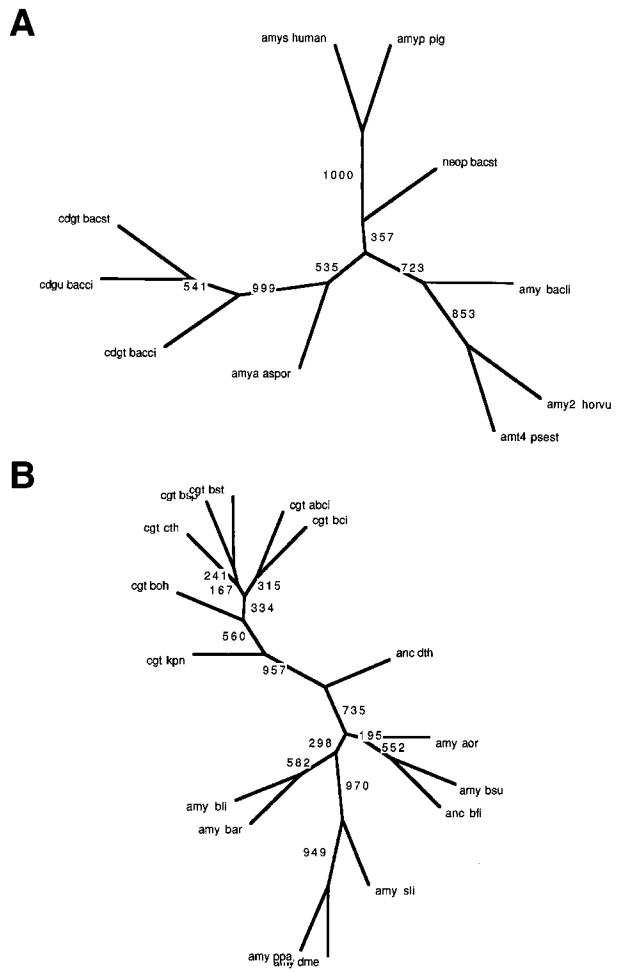

Fig. 1. Phylogenetic three (genetic distance tree) for α-amylase, CGTases. A: Genetic distance tree that includes a neopollulanase. The acronyms presented in this tree are explain in Section 2. B: Genetic distance tree that includes α-amylases (amy_*) and CGTases (cgt_*) previously aligned by Svensson et al. [17] (see text).

685 aa; anc_bfi, 976 aa). As shown in Fig. 1B, one of the *intermediate* amylases (anc_dth, from *Dictyoglomus thermophilum*) is indeed placed apart from α-amylases and CGTases, but the other one (anc_bfi, from *Butyrivibrio fibrisolvens*) has close relatives within the α-amylases group. This indicates that anc_bfi and anc_dth originated independently. Since the most likely ancestor of anc_bfi was an α-amylase, by analogy we believe that the ancestor of anc_dth was also an α-amylase-like enzyme. This is consistent with the proposal that α-

Table 2
Glucose produced from starch by different depolymerization mixtures

Enzymatic mixture	Glucose (mg/ml)		Glucose produced (mg/ml)
	at 0 h	at 5 h	
α-Amylase(1)	1.74	3.19	1.45
CGTase(1)	0.016	0.076	0.050
α-Amylase(1):CGTase(1)	1.49	5.71	4.22
α-Amylase(1):CGTase(2)	1.37	6.69	5.32
α-Amylase(2):CGTase(1)	1.46	5.09	3.63

α-Amylase from *B. licheniformis*, and CGTase from *B. macerans*. The concentration of each enzyme is given in parentheses as a multiple of standarized concentrations: 142 μg/ml α-amylase and 70 μg/ml CGTase.

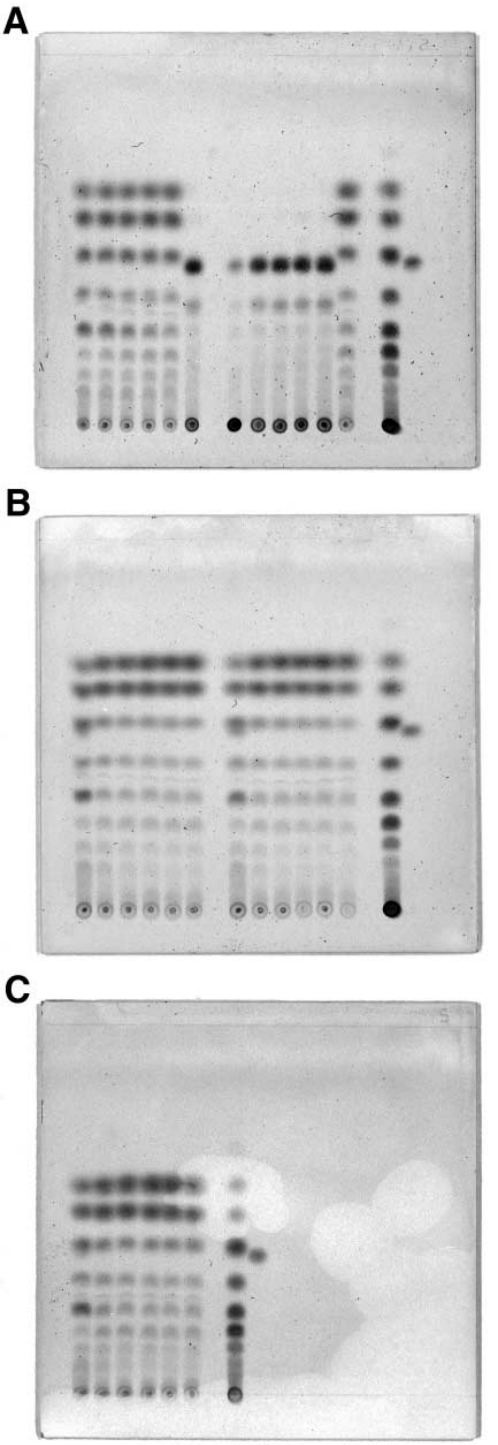

Fig. 2. Thin layer chromatography of starch depolymerization products. Six reaction times are presented for every enzymatic reaction: 0, 1, 2, 3, 4 and 5 h. A: Lanes 1–6, α-amylase hydrolysis products; lanes 7–12, CGTase transferase products. B: Lanes 1–6, α-amylase(1):CGTase(1) depolymerization products; lanes 7–12, α-amylase(1):CGTase(2) depolymerization products. C: Lanes 1–6, α-amylase(2):CGTase(1) depolymerization products. For all of the panels, the last two lanes are a ladder of glucose polymers from 1 to 9 (G1 to G9), and α-cyclodextin, respectively. See Section 2 for description of reaction condiction and nomenclature used.

amylases gave rise to CGTase by the acquisition of domains, through an intermediate that had similarity to anc_dth.

In *Escherichia coli* [22], *Hemophilus influenza* [23] and *Bacillus subtilis* [24] the operons for the synthesis of starch always include two depolymerases: glycogen phosphorylase and α-amylase. At least in *E. coli*, the starch and glycolytic operons are genetically controlled by the same regulator, csrA [25]. This indicates that α-amylases evolved as part of the intermediate metabolism. The fact that α-amylases are present in more diverse organisms than CGTases, that their sequences also show larger diversity, and that α-amylases belong to the intermediate metabolism, all point to α-amylases having appeared before CGTases.

Assuming a scenario where CGTases appeared in an organism already possessing an α-amylase, what could be the adaptive value for their emergence? We figured out that the combined action of α-amylases and CGTases may produce more glucose in less time than the simple action of either of these enzymes. This assumption is based on two observations. First, α-amylases, when depolymerizing oligosaccharides containing less than the number of glucoses they can bind at their active sites [26], tend to generate products whose further hydrolysis is inefficient. Second, CGTases produce mainly uniform size cyclodextrins [27]. Consequently, α-amylase would efficiently linearize and depolymerize the cyclodextrins, to glucose or maltose and a complementary small oligosaccharide (i.e. a 6-glucose cyclodextrin would be split to glucose and a maltopentaose, or to maltose and maltotetraose). In order to test this model, we performed the depolymerization of starch employing mixtures of an α-amylase and a CGTase (see Section 2).

3.2. Depolymerization of starch

The mono- and oligosaccharides produced by the action of α-amylase, CGTase and different combinations of them are presented in Fig. 2. As expected, the action of α-amylase on starch produced more small oligosaccharides than CGTase, while the latter produced mainly cyclodextrins. As predicted, the combined action of these enzymes always produced more glucose and small oligosaccharides (maltose, see Fig. 2) than the single action of either of them. Additionally, all of the enzymatic mixtures assayed produced more glucose (see Table 2) and reducing sugars (data not shown) than the action of either α-amylase or CGTase alone. Furthermore, the glucose (and reducing sugars, data not shown) produced by any of the enzymatic mixtures was more than the sum of the glucose produced by α-amylase and CGTase by themselves (see Table 2). This implies that these enzymes cooperate in the depolymerization of starch, as we predicted. In this regard, it is interesting to note that most of the known CGTases come from organisms possessing an α-amylase, in agreement with our results regarding the advantage of using these two activities for starch depolymerization.

The aim of this study was to test the hypothesis that CGTase activity evolved from an ancestral hydrolase (α-amylase-like). We found some evidence in our molecular phylogenetic analysis. To test the possible selective advantage of evolving a CGTase activity from an α-amylase (to accelerate the saccharification of starch), we measured the production of glucose from starch by enzymatic mixtures of α-amylase and CGTase. We found that the combined action of these enzyme activities produces glucose from starch faster and more efficiently. Besides supporting our evolutionary hypothesis, this result explains the biological role of CGTases, and could be useful in biotechnological processes aimed at producing glucose or maltose from starch [28].

Acknowledgements: We want to specially thank Agustin Lopez-Munguia for fruitful discussions, and Alejandro Garciarrubio for critical reading. We also thank Fernando Gonzalez for technical support and Alma Tremari for secretarial assistance. This work was funded in part by PAPIIT project number IN207293 from DGAPA/UNAM.

References

[1] Janecek, S. and Baláz, S. (1993) J. Protein Chem. 12, 509–514.
[2] Jespersen, H.M., MacGregor, E.A., Henrissat, B., Sierks, M.R. and Svensson, B. (1993) J. Protein Chem. 12, 791–805.
[3] Sinnott, M.L. (1990) Chem. Rev. 90, 1171–1202.
[4] Svensson, B. (1994) Plant Mol. Biol. 25, 141–157.
[5] Henrissat, B. (1991) Biochem. J. 280, 309–316.
[6] Holm, L. and Sander, Ch. (1994) Nucleic Acids Res. 22, 3600–3609.
[7] Knox, J.R., Moews, P.C. and Frere, J.M. (1996) Chem. Biol. 3, 937–947.
[8] Schenkman, S., Eicinger, D., Pereira, M.A.E. and Nussenzweig, V. (1994) Annu. Rev. Microbiol. 48, 499–523.
[9] Stemmer, W.P.C. (1994) Proc. Natl. Acad. Sci. USA 91, 10747–10751.
[10] Kauffman, S.A. (1992) J. Theor. Biol. 157, 1–7.
[11] Shao, Z. and Arnold, F.H. (1997) Curr. Opin. Chem. Eng. 210, 513–518.
[12] Kuroki, R., Weaver, L.H. and Matthews, B.W. (1995) Nature Struct. Biol. 2, 1007–1011.
[13] Abrahmsen, L., Tom, J., Burnier, J., Butcher, K.A., Kossiakoff, A. and Wells, J.A. (1991) Biochemistry 30, 4151–4159.
[14] Takata, H., Kuriki, T., Okada, S., Takesada, Y., Iizuka, M., Minamiura, N. and Imanaka, T. (1992) J. Biol. Chem. 267, 18447–18452.
[15] Roderic, D.M. (1996) Comput. Appl. Biosci. 12, 357–358.
[16] Summer, J.B. and Howell, S.F. (1935) J. Biol. Chem. 108, 51–54.
[17] Janecek, S., MacGregor, E.A. and Svensson, B. (1995) Biochem. J. 305, 685–688.
[18] Jespersen, H.M., MacGregor, E.A., Sierks, M.R. and Svensson, B. (1991) Biochem. J. 280, 51–55.
[19] Kaneko, T., Song, K.B., Hamamoto, T., Kudo, T. and Horikoshi, K. (1989) J. Gen. Microbiol. 135, 3447–3457.
[20] Fujiwara, S., Kakihara, H., Woo, K.B., Lejeune, A., Kanemoto, M., Sakaguchi, K. and Imanaka, T. (1992) Appl. Environ. Microbiol. 58, 4016–4025.
[21] Demetrius, L. (1995) Protein Eng. 8, 791–800.
[22] Romeo, T., Kumar, A. and Preiss, J. (1988) Gene 70, 363–376.
[23] Tatusov, R., Mushegian, A.R., Bork, P., Brown, N.P., Hayes, W.S., Borodovsky, M., Rudd, K.E. and Koonin, E.V. (1996) Curr. Biol. 6, 279–291.
[24] Kiel, J.A.K.W., Boels, J.M., Beldman, G. and Venema, G. (1994) Mol. Microbiol. 11, 203–218.
[25] Sabnis, N.A., Yang, H. and Romeo, T. (1995) J. Biol. Chem. 270, 29096–29104.
[26] Allen, J.D. and Thoma, J.A. (1997) Biochemistry (in press).
[27] Nakamura, A., Haga, K. and Yamane, K. (1994) FEBS Lett. 337, 66–70.
[28] Godfrey, T. and Reichelt, J. (1983) in: The Nature Press, pp. 375–390.
[29] Reardon, D. and Farber, G.K. (1995) FASEB J. 9, 497–503.