In Silico Analysis of the Thermodynamic Stability Changes of Psychrophilic and Mesophilic α-Amylases upon Exhaustive Single-Site Mutations

Dimitri Gilis*

Genomic and Structural Bioinformatics, Université Libre de Bruxelles, Avenue F. Roosevelt 50 CP 165/61, 1050 Brussels, Belgium

Received October 27, 2005

Identifying sequence modifications that distinguish psychrophilic from mesophilic proteins is important for designing enzymes with different thermodynamic stabilities and to understand the underlying mechanisms. The PoPMuSiC algorithm is used to introduce, in silico, all the single-site mutations in four mesophilic and one psychrophilic chloride-dependent α-amylases and to evaluate the changes in thermodynamic stability. The analysis of the distribution of the sequence positions that could be stabilized upon mutation shows a clear difference between the three domains of psychrophilic and mesophilic α-amylases. Most of the mutations stabilizing the psychrophilic enzyme are found in domains B and C, contrary to the mesophilic proteins where they are preferentially situated in the catalytic domain A. Moreover, the calculations show that the environment of some residues responsible for the activity of the psychrophilic protein has evolved to reinforce favorable interactions with these residues. In the second part, these results are exploited to propose rationally designed mutations that are predicted to confer to the psychrophilic enzyme mesophilic-like thermodynamic properties. Interestingly, most of the mutations found in domain C strengthen the interactions with domain A, in agreement with suggestions made on the basis of structural analyses. Although this study focuses on single-site mutations, the thermodynamic effects of the recommended mutations should be additive if the mutated residues are not close in space.

INTRODUCTION

α-Amylases belong to family 13 of glycosidase.[1] These enzymes are involved in the degradation of starch, a carbohydrate polymer that constitutes an important energy source for plants, animals, and bacteria. It is thus not surprising to find them in all three domains of life. Moreover, they are extensively used in several industries and are one of the most used enzymes in biotechnology.[2] This broad distribution among the different types of organisms, living in very different temperature conditions, implies that selective pressures acted to produce α-amylases that are active in a large range of temperature: this protein family is characterized by melting temperatures comprised between 40 °C and more than 100 °C. Despite a low degree of sequence similarity between α-amylases coming from the entire spectrum of sources, their overall topology is extremely well conserved, and their structure is always organized around an $(\beta/\alpha)_8$ barrel.[3,4] In general, α-amylases contain a calcium ion which is important for the stability of the enzyme.[4] The conserved sequence portions correspond to the catalytic residues, to the active site and to amino acids responsible for the binding to calcium. This fact leads to the conclusion that the activity at extreme temperatures of some members of the family is not due to the modification of the reaction mechanism. For these different reasons and given the high structural conservation within this protein family, α-amylases are good candidates to study and to understand the mechanisms of adaptation to extreme environments.[5−9]

Several structural factors have been proposed to explain the activity and the stability at very low and high temperatures.[9−14] First, an increase in the number and strength of weak interactions and structural factors responsible for protein stability have been noticed from psychrophiles to mesophiles and to thermophiles. Furthermore, the structure of a psychrophilic α-amylase has been analyzed and compared to mesophilic counterparts,[15] showing a lower percentage of proline and arginine, weaker interdomains interactions, and a larger exposure of hydrophobic residues in the psychrophilic α-amylase. These different factors are assumed to provoke a lower rigidity and stability of this enzyme. Recently, the conformational stability and kinetic parameters of three α-amylases active at low, medium, and high temperature have been studied in the same conditions in order to investigate the relationship between stability, activity, flexibility, and the adaptation to temperature.[16] The results support the concept of "corresponding states",[10,17] which assumes that psychrophilic, mesophilic, and thermophilic enzymes show more or less the same conformational flexibility and catalytic activity at their individual physiological temperatures. Note that some experiments performed on two thermostable α-amylases do not corroborate this hypothesis.[18] It was proposed to explain this result that the mechanisms related to the corresponding state hypothesis do not play a role to achieve adaptation to very high temperature.

Berezovsky et al.[13] proposed recently a physical explanation of the different strategies that are used to adapt to extreme temperatures. Their study is based on an analysis of a large number of sequences and structures from various (hyper)thermophilic and mesophilic organisms. Their results

* Corresponding author phone: +32 2 650 36 15; fax: +32 2 650 35 75; e-mail: dgilis@ulb.ac.be.

10.1021/ci050473v CCC: $33.50
Published on Web 02/18/2006

suggest that the proteins from organisms that developed originally in an extreme environment present particular structures that evolved *concomitantly* with their sequence and that are more compact and more hydrophobic than their mesophilic equivalents. On the contrary, proteins from organisms that evolved in a mesophilic environment and that adapted later to a hot environment adopted a different strategy: they were subjected to sequence modifications in view of strengthening some interactions and allowing a better resistance to higher temperatures.

To our knowledge, the previous studies aiming at locating regions and mutations responsible for the psychrophilic or thermophilic character of some α-amylases and at proposing sequence modifications to increase the melting temperature were performed experimentally or were based on sequence alignments.[5,19−28] We tackle here this problem by using an in silico method that was developed to estimate the changes in folding free energy of *all* possible point mutation in a protein structure.[29,30] In particular, we focus on psychrophilic and mesophilic chloride-dependent α-amylases and investigate the effects of *all* possible single-site mutations on their thermodynamic stability. Mutations that are predicted to stabilize the psychrophilic enzyme are identified, and the thermodynamic behavior of the three α-amylase domains upon mutations is analyzed.

The relative strength of the different types of interactions depends on the temperature,[31,32] and our in silico approach relies on database-derived potentials describing interactions at room temperature.[29,33−35] Applying these potentials to compute the energy changes upon mutation in psychrophilic, mesophilic, and thermophilic α-amylases requires the assumption that the interactions are temperature-independent. Given the melting temperature (T_m) of the chloride-dependent α-amylases considered here (see Material and Methods), the temperature range over which the assumption must be made is about 20 °C to compare the psychrophilic to the mesophilic α-amylases, whereas it is larger than 56 °C if we include in our study the thermophilic α-amylase from *Bacillus licheniformis* (T_m larger than 100 °C[36]). Taking into account the thermophilic α-amylase in this study broadens hugely the temperature range over which we must assume that the interactions do not depend on temperature. The case of thermophilic enzymes is therefore not considered here.

MATERIAL AND METHODS

Structure Data. Only chloride-dependent α-amylases for which the structure has been resolved are considered. The atomic coordinates are retrieved from the Brookhaven Protein Data Bank[37] for pig pancreatic α-amylase[38] (PDB code 1PIF), human pancreatic α-amylase[3] (PDB code 1HNY), human salivary α-amylase[39] (PDB code 1SMD), *Tenebrio molitor* α-amylase[40] (PDB code 1JAE), and *Alteromonas haloplanktis* α-amylase[41] (PDB code 1AQH). The first four proteins are obtained from mesophilic organisms and present melting temperatures comprised between 60.7 °C and 67.4 °C,[36] whereas the latter is psychrophilic, with a melting temperature of 43.7 °C.[36] These melting temperatures have been obtained by differential scanning calorimetry and correspond to the temperature at which 50% of the proteins are unfolded.

Sequence and Structure Alignments. A sequence alignment is needed to compare the thermodynamic effect of the mutations at each sequence position of the different α-amylases. In a first step, a multiple sequence alignment is generated by the ClustalW program[42] of the European Bioinformatics Institute. A pairwise superimposition of the 3D structures is also performed in view of verifying the sequence alignment obtained by ClustalW. The SoFiSt program[43] is used for that purpose. This program has been designed to yield optimal superimposition with respect to the root-mean-square deviation of heavy main chain atoms after coordinate superimposition. In a final step, the sequence alignment is tuned with respect to the structure alignments. This procedure consisting of building the multiple sequence alignment first and then correcting it with a pairwise structure alignment is less time-consuming than obtaining it directly from the structure superimposition. The final sequence alignment is slightly different from that presented in Aghajari et al.,[15] due to the tuning on the basis of the structure superimpositions.

Computation of the Changes in Folding Free Energy upon Mutation. The PoPMuSiC program[29,30] is used to introduce, in silico, all the single site mutations and to calculate the resulting changes in folding free energy between the wild-type and the mutated proteins ($\Delta\Delta G$). It assumes that the mutations do not perturb significantly the backbone structure. The proteins are represented by the main chain atoms N, C, O, C^α and a centroid defined as the geometric average of all heavy side-chains atoms of a given amino acid type in a data set of known structures.[33] These simplifications allow fast computations of the $\Delta\Delta G$'s of a large number of mutations. The evaluation of the $\Delta\Delta G$'s relies on knowledge-based potentials. They are derived from observed frequencies of sequence and structure patterns in a data set of 141 high-resolution X-ray structures (for a list see ref 44). These proteins are mesophilic, and the energy parameters are valid for this range of temperature. The performances of PoPMuSiC are quite good, as attested by correlation coefficients between computed and measured $\Delta\Delta G$'s comprised between 0.80 and 0.87 on a test set of 296 mutations[29] and blind predictions on a protein of the serpin family.[45] Our program does not evaluate the $\Delta\Delta G$ of mutations presenting a solvent accessibility comprised between 40% and 50%.[29]

For each sequence position of each protein included in this study, the sum of the $\Delta\Delta G$'s of stabilizing mutations ($\Gamma^{(-)}$) and the sum of the $\Delta\Delta G$'s of destabilizing mutations ($\Gamma^{(+)}$) is calculated. $\Gamma^{(-)}$ is always negative and $\Gamma^{(+)}$ is positive with the convention that stabilizing mutations have negative $\Delta\Delta G$'s. Note that mutations into prolines are not included in these sums, because they could provoke a structural modification of the backbone conformation, as observed in a previous study.[46] This structural rearrangement is not taken into account in PoPMuSiC, and the computed $\Delta\Delta G$'s for these mutations are therefore less accurate. Mutations of prolines are also excluded from our analysis; the proportion of sequence positions occupied by a proline in the α-amylases studied here is comprised between 2.9% and 4.6%. A large $\Gamma^{(-)}$, in absolute value, denotes a sequence position that is not optimized with respect to thermodynamic stability, unlike positions characterized by a large $\Gamma^{(+)}$. In addition, the density of sequence positions of a protein or a protein region that presents a high rate of stabilizing mutations, $\rho^{(-)}$, is defined as $\rho^{(-)} = N^{(-)}/N$, where $N^{(-)}$ is the number of sequence positions with $\Gamma^{(-)}$ smaller than or

equal to −5 kcal/mol and N is the number of residues in the protein or in the protein region. The density of sequence position characterized by a high rate of destabilizing mutations, $\rho^{(+)}$, is equal to $\rho^{(+)} = N^{(+)}/N$, where $N^{(+)}$ is the number of sequence positions with $\Gamma^{(+)}$ larger than or equal to 30 kcal/mol. The cutoff of −5 kcal/mol corresponds approximately to the average $\Gamma^{(-)}$ minus one standard deviation and the cutoff of 30 kcal/mol to the average $\Gamma^{(+)}$ plus one standard deviation. The average of $\Gamma^{(-)}$ is about −2 kcal/mol (excluding sequence positions with $\Gamma^{(-)} = 0$ kcal/mol), with a standard deviation about 3−4 kcal/mol, depending on the considered protein. The average of $\Gamma^{(+)}$ is about 23 kcal/mol, with a standard deviation about 14−15 kcal/mol. There is a considerable difference between the average $\Gamma^{(+)}$ and $\Gamma^{(-)}$ because the number of destabilizing mutations is larger than the number of stabilizing mutations.

RESULTS AND DISCUSSION

Comparison of the Thermodynamic Effects of Mutations in the Five α-Amylases. The effects of point mutations on the stability of five α-amylases from pig pancreas, human pancreas, human salivary, *Tenebrio molitor*, and *Alteromonas haloplanktis* are analyzed in silico with PoPMuSiC.[29,30] The $\Delta\Delta G$'s as well as the different measures that are used in this work to compare the different proteins are calculated at room temperature. These measures represent the thermodynamic impact of travelling in the sequence space from the wild-type sequence to a region in its vicinity.

The structure of α-amylase is composed of three domains.[3,15,38−41] Domain A (alignment positions 1-99 and 173-410, Figure 1) contains the active site residues and those involved in ligand binding. Domain B (alignment positions 100−172, Figure 1) forms a cavity against domain A, in which the calcium ion is bound and is a part of the substrate binding cleft. Domain C (alignment positions 411-510, Figure 1) is located opposite to domain B.

Table 1 contains $\rho^{(-)}$, $\rho^{(+)}$, $<\Gamma^{(-)}>$, and $<\Gamma^{(+)}>$ for the three domains of each of the considered α-amylases. The two first measures are defined in Material and Methods. $<\Gamma^{(-)}>$ and $<\Gamma^{(+)}>$ are the average of $\Gamma^{(-)}$ and $\Gamma^{(+)}$ in a protein domain. Note that $\Gamma^{(-)}$ and $\Gamma^{(+)}$ are the sum of negative and positive $\Delta\Delta G$'s at a sequence position (see Material and Methods). Positions with a large $\Gamma^{(-)}$, in absolute value, can be easily stabilized upon mutation and are thus not optimized with respect to thermodynamic stability. In contrast, most of the mutations at positions presenting a large $\Gamma^{(+)}$ destabilize the protein structure. We would like to stress that these different measures are computed from the differences between the folding free energy of the mutated and the wild-type proteins. They do not provide an estimation of the absolute stability of a protein but rather the stabilization that could be gained or lost by this protein upon single-site mutations.

The value of $\rho^{(-)}$ computed for domain A of the five proteins (Table 1) shows that the psychrophilic enzyme (1AQH) presents a lower density of sequence positions that could be stabilized. Moreover, $\rho^{(+)}$ is higher for 1AQH, indicating that a larger proportion of mutations will destabilize domain A of this protein. The values of $<\Gamma^{(-)}>$ and $<\Gamma^{(+)}>$ follow the same trend, although the differences between 1AQH and the other proteins is less marked. Note that domain A of *Tenebrio molitor* α-amylase (1JAE) presents a range of possibilities to improve its thermodynamic stability, as denoted by its large $<\Gamma^{(-)}>$.

Figure 1 shows the sequence alignment of the considered proteins and the results of the PoPMuSiC computations per sequence position. This figure clearly stresses the positions in domain A that are well optimized in 1AQH and not in the mesophilic enzymes. Residue E237 (alignment numbering) is very interesting. It belongs to the catalytic triad, it is of course conserved in all α-amylases, and it is characterized by a $\Gamma^{(-)}$ equal to −3.1 kcal/mol in 1AQH and comprised between −7.5 and −19.5 kcal/mol in the other four proteins. It is not surprising to find a low $\Gamma^{(-)}$ for a catalytic residue as selective pressures first optimize the activity of the protein, its stability being possibly refined in a second step. To explain this difference between the psychrophilic and the mesophilic α-amylases, the sequences of 1AQH and pig pancreatic α-amylase (1PIF) are compared. Residues within a sphere of 6 Å around E237 and that are not conserved in 1PIF are I199 and R256 (Figure 1), which correspond to a Phe and a Leu residue, respectively, in 1AQH. The active site of the enzymes being highly conserved, it is not surprising to find only two differences between 1AQH and 1PIF. Introducing the mutation I199F into the pig pancreatic α-amylase (1PIF) is predicted to destabilize the protein by 1.78 kcal/mol, whereas mutating residue R256 into Leu increases the stability by 1.11 kcal/mol. The Pymol program (http://pymol.sourceforge.net) was used to substitute Ile and Arg at positions 199 and 256 of 1PIF by Phe and Leu, respectively, and to place their side chains correctly. This mutant structure is referenced as 1PIF_FL. PoPMuSiC was run on 1PIF_FL, and $\Gamma^{(-)}$ was computed for each sequence position. This double mutation provokes an increase of $\Gamma^{(-)}$ at position 237: it is equal to −10.1 kcal/mol in 1PIF and to −5.8 kcal/mol in 1PIF_FL; for the sake of comparison, it is equal to −3.1 kcal/mol in 1AQH. Introducing the mutations I199F and R256L in 1PIF generates a new sequence that is better thermodynamically adapted to the catalytic glutamic acid at position 237. Moreover, $\Gamma^{(-)}$ increases from −8.59 in 1PIF to 0 in 1PIF_FL at position 256 and remains equal to 0 at position 199.

The behavior of domains B and C upon mutations is completely different, compared to domain A. Indeed, Table 2 clearly shows that $\rho^{(-)}$ is larger for both domains of 1AQH, compared to that of the other proteins. This indicates a possibility to increase their thermodynamic stability by mutations. Domain B presents a low $<\Gamma^{(-)}>$ as well as a slightly lower $<\Gamma^{(+)}>$ in 1AQH and *Tenebrio molitor* α-amylase (1JAE). In the case of domain C, $<\Gamma^{(-)}>$ is small in 1AQH. Note also that $\rho^{(+)}$ is markedly lower in 1AQH. These results mean that 1AQH could be stabilized by introducing mutations in domains B and C and that a large part of possible mutations destabilizes the structure in a lesser extent than in the other proteins.

Note that positions corresponding to residues belonging to the β-barrel (domain A) and to the Greek-key motif in domain C present in general high $\Gamma^{(+)}$'s in the mesophilic proteins (Figure 1).

We would like to stress that according to structural considerations, domains A and B are merged in CATH.[49] The subdivision into three domains has been proposed according to sequence and functional issues.[50] Our calcula-

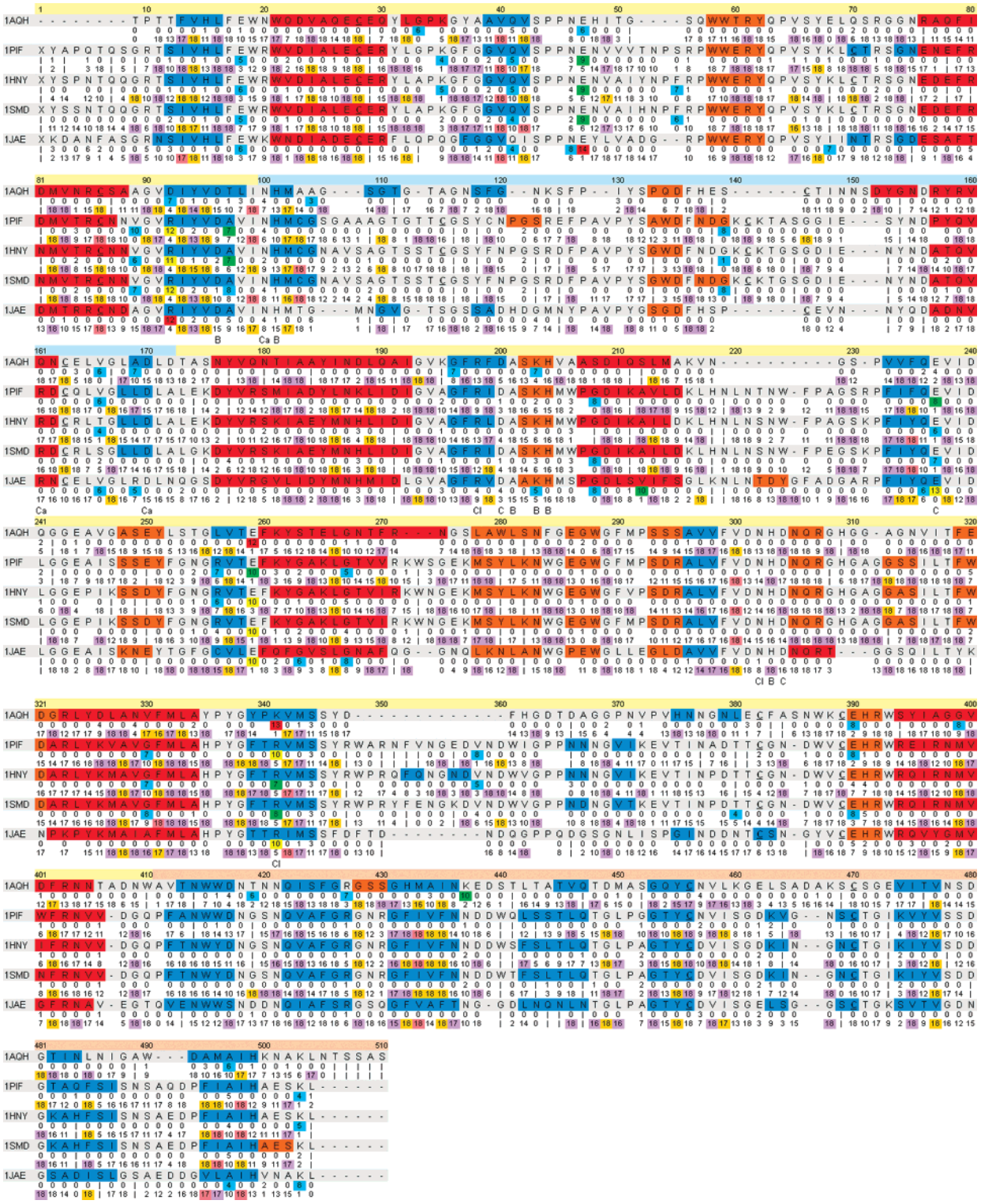

Figure 1. sequence alignment of α-amylases. The alignment numbering is colored according to the domain limits: yellow corresponds to domain A, blue to domain B, and salmon to domain C. The sequences are colored according to the secondary structures: blue corresponds to β-strands, red to α-helices, and orange to 3−10 helices. The secondary structures have been assigned by PROMOTIF[47] and retrieved from PDBSum.[48] Cysteine residues involved in disulfide bonds are underlined and in bold. The first line under each sequence contains the number of mutations per sequence position with a $\Delta\Delta G$ lower than or equal to −0.5 kcal/mol. It is colored in blue if -10 kcal/mol $\leq \Gamma^{(-)} < -5$ kcal/mol, in green if -15 kcal/mol $\leq \Gamma^{(-)} < -10$ kcal/mol, in yellow if -20 kcal/mol $\leq \Gamma^{(-)} < -15$ kcal/mol, and in red if $\Gamma^{(-)} < -20$ kcal/mol. The second line under each sequence contains the number of mutations per sequence position with a $\Delta\Delta G$ larger than or equal to 0.5 kcal/mol. It is colored in violet if 30 kcal/mol $< \Gamma^{(+)} \leq 45$ kcal/mol, in orange if 45 kcal/mol $< \Gamma^{(+)} \leq 60$ kcal/mol, and in pink if $\Gamma^{(+)} > 60$ kcal/mol. '|' indicates a sequence position with a solvent accessibility comprised between 40 and 50% and for which no stability measure can be computed by PoPMuSiC (see Material and Methods). We did not mutate prolines (see Material and Methods), and the two first lines under the sequences are left blank in this case. The line under each alignment group contains information about residues responsible for the binding and the catalytic activity and that are conserved in the 5 sequences. 'Ca' and 'Cl' indicate residues involved in the binding of calcium and chloride ions, respectively. 'B' and 'C' correspond to residues responsible for binding to the substrate and for the catalytic activity, respectively.

Table 1. Stability Measures per Protein Domain[a]

	domain A					domain B					domain C				
	1AQH	1PIF	1HNY	1SMD	1JAE	1AQH	1PIF	1HNY	1SMD	1JAE	1AQH	1PIF	1HNY	1SMD	1JAE
$\rho^{(-)}$ (%)	3.4	5.5	4.8	4.6	5.1	5.8	3.3	3.4	1.7	3.8	4.7	1.2	1.2	0.0	2.4
$\rho^{(+)}$ (%)	35.6	32.7	34.3	32.2	32.4	19.2	22.9	20.4	18.7	19.2	23.6	30.3	32.0	31.7	29.2
$<\Gamma^{(-)}>$ (kcal/mol)	−0.71	−0.81	−0.74	−0.77	−0.98	−0.75	−0.49	−0.47	−0.49	−0.74	−0.78	−0.41	−0.48	−0.43	−0.40
$<\Gamma^{(+)}>$ (kcal/mol)	23.4	24.1	24.4	24.2	23.9	19.4	21.5	20.7	20.6	18.3	20.7	23.3	23.8	23.3	24.1

[a] Stability measures for the three domains of *Alteromonas haloplanktis* (1AQH), pig pancreatic (1PIF), human pancreatic (1HNY), human salivary (1SMD), and *Tenebrio molitor* (1JAE) α-amylases. See Material and Methods for the definition of $\rho^{(-)}$ and $\rho^{(+)}$. $<\Gamma(-)>$ and $<\Gamma^{(+)}>$ are the average $\Gamma^{(-)}$ and $\Gamma^{(+)}$ values in the corresponding protein domain. 1AQH is the psychrophilic enzyme. See Figure 1 for the domain limits.

Table 2. Stability Measures for the Whole Proteins[a]

	1AQH	1PIF	1HNY	1SMD	1JAE
$\rho^{(-)}$ (%)	4.0	4.4	3.9	3.4	4.5
$\rho^{(+)}$ (%)	26.5	31.0	32.0	30.3	30.2
$<\Gamma^{(-)}>$ (kcal/mol)	−0.73	−0.69	−0.66	−0.67	−0.84
$<\Gamma^{(+)}>$ (kcal/mol)	22.3	23.6	23.8	23.6	23.2

[a] Stability measures for the whole proteins. See legend of Table 1 for a description of the measures.

tions dealing with the thermodynamic effects of points mutations stem in favor of the separation into 3 domains, the behavior of domains A and B upon mutation being very different.

The values of $\rho^{(-)}$, $\rho^{(+)}$, $<\Gamma^{(-)}>$, and $<\Gamma^{(+)}>$ computed for the *whole* proteins are provided in Table 2. They can be compared to experimental measures made on the *whole* proteins. First, $\rho^{(-)}$ is almost the same for the different proteins, the human salivary (1SMD) and *Tenebrio molitor* (1JAE) α-amylases presenting slightly smaller and larger values, respectively, and the same trend is followed by the absolute value of $<\Gamma^{(-)}>$. On the other hand, $\rho^{(+)}$ and $<\Gamma^{(+)}>$ of the psychrophilic protein (1AQH) is markedly smaller than that of the others. The fact that $\rho^{(+)}$ and $<\Gamma^{(+)}>$ are lower in 1AQH suggests that all the possible single-site mutations are on the average less destabilizing in the psychrophilic enzyme than in the mesophilic ones. This observation is in agreement with the in vitro result showing that 1AQH has reached a state close to the lowest possible stability without disruption of its native state.[36] On the other hand, it is surprising that the density of sequence positions that could be stabilized upon mutation is not higher in 1AQH as this enzyme has lost in overall stability to adapt to low temperature. A large rate of stabilizing mutations could be expected as well as a noticeably smaller $<\Gamma^{(-)}>$. But we have to keep in mind that the measures presented here reflect the impact of *single-site* mutations. The fact that the density of stabilizing mutations is equivalent in the different proteins is in agreement with an experimental study suggesting that single-site mutations are not sufficient to drastically stabilize 1AQH.[27]

The greater proportion of charged residues in (hyper)-thermophiles is a well-known feature. Surprisingly, the increase in the amount of positively charged residues is mainly due to lysine. This amino acid seems to play a particular role, as it is over-represented compared to arginine in (hyper)thermophiles. Berezovsky et al.[14] proposed a physical explanation of the enrichment in lysine residues. They performed unfolding simulations and showed that lysines present a larger number of accessible rotamers than arginines of similar solvent accessibility in protein structures. Therefore, it appears that substituting arginine by lysine stabilizes the protein entropically. According to this result, psychrophilic genomes and proteins, which are less stable than their mesophilic counterparts, should not display a larger amount in lysines. We first computed the amino acid composition of the psychrophilic α-amylase to check this assumption. As expected, we do not obtain any bias: the content of lysine and arginine is both equal to 2.9%. We evaluated also the amino acid composition of four genomes from psychrophilic organisms, and we found that the lysine and arginine contents are equivalent (*Colwellia* psychrerythraea 34H: K: 6.0%, R: 3.8%; *Desulfotalea psychrophila LSv54*: K: 5.7%, R: 5.0%; *Photobacterium* profundum SS9: K: 5.4%, R: 4.5% and *Psychrobacter arcticus 273-4*: K: 5.1%, R: 4.5%).

Mutations Stabilizing the Psychrophilic Enzyme. Modifying the thermostability of α-amylases was the central theme of several experimental studies, which focused on 2 systems: the psychrophilic α-amylase 1AQH[24,27] and the thermophilic α-amylase from *Bacillus* licheniformis.[6,21,23,25,26] In both cases, the authors aimed at increasing the melting temperature of the enzymes and at understanding the factors governing adaptation to extreme temperatures. In this section, we analyze the sequence alignment presented in Figure 1 in order to identify the positions that present a higher rate of stabilizing mutations in 1AQH (large $\Gamma^{(-)}$ in absolute value) than in the four mesophilic enzymes. We propose mutations that are expected to increase the thermodynamic stability of 1AQH, and we comment our calculations on the mutations that have been experimentally characterized.

The single-site substitutions in 1AQH that were experimentally studied and for which the melting temperature has been measured are N12R, R64E, N150D, Q164I, V196F, L219R, T232V, E279W, N288V, K300R, and M379V (1AQH numbering).[24,27] Mutations Q164I, T232V, N288V, and M379V present two calorimetric domains, each of them being characterized by a melting temperature. Our computed $\Delta\Delta G$'s represent the change in stability of the *whole* protein and can therefore not be compared to melting temperatures of protein domains. Position 64 has a solvent accessibility comprised between 40 and 50%, and our program does not evaluate the $\Delta\Delta G$'s of mutations in this solvent accessibility range. Table 3 summarizes the measured T_m and the calculated $\Delta\Delta G$ of the mutations for which these two quantities can be compared. The correlation coefficient between the predicted $\Delta\Delta G$'s and the measured $\Delta T_m = T_m(\text{wild-type}) - T_m(\text{mutant})$ is equal to 0.75. Note that we have already discussed the mutation L219R (L303R according to the alignment numbering) in the previous section. Indeed, we showed that this position is well optimized with

Table 3. Measured Thermodynamic Parameters of Mutations Introduced in 1AQH and Their Predicted ΔΔ*G*'s[a]

	alignment position	ΔΔ*G* (kcal/mol)	T_m (°C)
Mutations Measured as Stabilizing			
wild-type			44.0
N150D	218	0.02	44.8
V196F	276	0.16	45.4
K300R	388	−0.28	45.2
Mutations Measured as Destabilizing			
N12R	27	1.00	43.3
L219R	303	0.99	42.5
E279W	367	0.61	41.4

[a] Measured thermodynamic parameters of some mutations introduced in 1AQH and their predicted ΔΔ*G*'s. The second column is the wild-type residue, its position along the sequence, and the mutant residue. The third column is the position along the alignment (Figure 1). ΔΔ*G* is the computed change in folding free energy due to the mutation (see Material and Methods). T_m is the melting temperature and has been measured in refs 24 and 27.

respect to thermodynamic stability in 1AQH and not in 1PIF. An Arg residue is found at this position of 1PIF. Mutating it into Leu in 1PIF is predicted as stabilizing, contrary to the mutation Leu → Arg in 1AQH.

Mutations computed with a ΔΔ*G* lower than or equal to −1.0 kcal/mol by PoPMuSiC were selected, with the aim of engineering a modified *Alteromonas haloplanktis* α-amylase (1AQH) presenting mesophilic-like thermodynamic properties. This cutoff was chosen to catch the most stabilizing mutations and to ensure that the proposed mutations are really stabilizing given the error on the calculations. The 72 mutations that fulfill this criterion are given in Table 1 of Supporting Information. A large number of these mutations occurs at positions that are important for the activity of the protein, close to the active site or crucial for the binding to the chloride and calcium ions. The most interesting mutations that are outside these regions and that present the largest stabilization effect are provided in Table 4, and their location in the structure is shown on Figure 2. The 19 mutations listed in Table 4 are expected to stabilize the enzyme and to increase its melting temperature, assuming a relationship between thermodynamic and thermal stabilities.

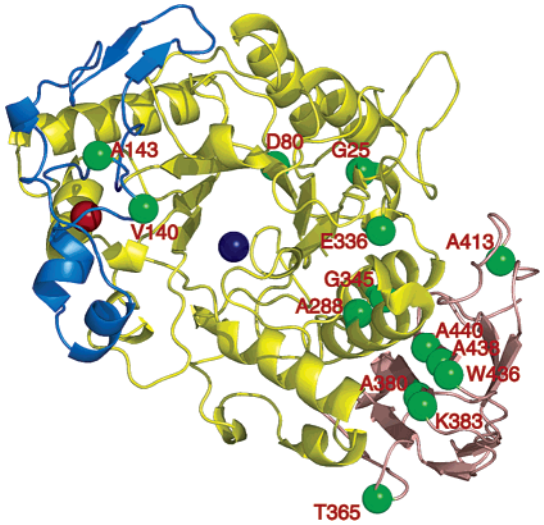

Figure 2. 3D structure of 1AQH. Sequence positions of the mutations proposed in Table 4 are represented by a green sphere. The one letter code of the amino acid and its sequence position is provided (1AQH numbering). Domains A, B, and C are colored according to the color code used for Figure 1, i.e., in yellow, blue, and salmon, respectively. The calcium and chloride ions are symbolized by a red and a blue sphere, respectively. The figure has been generated with Pymol (http://pymol.sourceforge.net).

Residue G25 is positioned in a helix, and the mutant residue proposed here, an alanine, is found in the mesophilic α-amylases 1HNY and 1SMD. Mutating G25 into Ala in 1AQH is predicted to stabilize this helix. The other mutations that are predicted to stabilize domain A are located at positions 80, 288, 336, and 345 along the sequence. Two positions of domain B are proposed to be mutated. The first, V140, is fully accessible to the solvent, whereas the second, A143, is buried. Replacing A143 by a Phe will reinforce the interactions with the neighboring aromatic residues (F108, F116, and to a lesser extent Y111). It has been suggested previously that the association between domains A and C is lower in the psychrophilic α-amylase compared to the mesophilic counterparts.[15] Our energy calculations confirm this structural analysis: most of the mutations of residues from domain C that are predicted by PoPMuSiC as stabilizing

Table 4. Mutations Proposed To Increase the Stability of 1AQH[a]

alignment position	sequence position	wild-type residue	mutant residue	ΔΔ*G* (kcal/mol)	localization in the structure
33	25	G	A	−1.12	domain A
92	80	D	C	−1.27	domain A
166	140	V	G	−1.50	domain B
166	140	V	N	−1.47	domain B
169	143	A	F	−1.60	domain B
329	288	N	A	−1.21	domain A
390	336	E	W	−1.09	domain A
390	336	E	F	−1.06	domain A
399	345	G	L	−1.49	domain A
399	345	G	M	−1.34	domain A
419	365	T	G	−1.95	domain C
434	380	A	V	−1.01	domain C
437	383	K	F	−1.95	domain C, close to domain A
437	383	K	L	−1.45	domain C, close to domain A
437	383	K	Y	−1.44	domain C, close to domain A
467	413	A	G	−1.08	domain C
490	436	W	G	−1.65	domain C, close to domain A
495	438	A	V	−1.01	domain C, close to domain A
497	440	A	F	−1.90	domain C, close to domain A

[a] ΔΔ*G*'s of the most interesting stabilizing mutations in 1AQH, as computed by PoPMuSiC.

are found at the interface between domains A and C (Figure 2). The mutations at the buried positions 380, 383, 438, and 440 increase the number of hydrophobic or aromatic−aromatic contacts between both domains. In the case of mutations T365G, A413G, and W436G, the computed $\Delta\Delta G$'s are dominated by the term describing local interactions along the sequence. They are thus predicted to mainly stabilize the secondary structure of domain C. Interestingly, one of the mutant residues proposed at positions 25, 345, 365, 380, and 438 (alignment positions 33, 399, 419, 434, and 495) corresponds to the native residue of at least one of the mesophilic protein.

CONCLUSIONS

Understanding the mechanisms underlying the adaptation of enzymes to low temperature is the central theme of several works as well as engineering proteins with modified melting temperature.[7−9] For that purpose, structural and sequence analyses are usually coupled to experimental methods, with the aim of characterizing the thermodynamic properties of wild-type and mutated psychrophilic and mesophilic proteins. The study presented in this paper is intermediate between these two approaches. Indeed, structural analyses generally amount the effects of a mutation to the formation or the breakage of some interactions. The energy calculations performed in silico take into account the whole environment of the mutated residue as well as the effects of the mutation on the stability of the secondary structure.

In silico means have been used here to evaluate the impact of *all* the single-site mutations on the thermodynamic stability of one psychrophilic and four mesophilic α-amylases. The program used for that purpose, PoPMuSiC,[29,30] has been previously proven to lead to accurate predictions[29] and has been successfully tested in blind predictions.[45] The density of sequence positions presenting a high rate of stabilizing mutations has been shown to be very different in the three domains of the analyzed α-amylases. Domains B and C of the psychrophilic enzyme, 1AQH, present a larger range of possibilities to be stabilized than the four mesophilic equivalents. This indicates that the stabilization of 1AQH should be reached by introducing mutations in both of these domains. Interestingly, the behavior upon mutation of the catalytic domain A of the psychrophilic enzyme is very different from the four mesophilic equivalents. The catalytic domain A of the psychrophilic enzyme presents less possibilities to be stabilized. Our calculations show that the environment of some key residues for the activity of the enzyme evolved to minimize the thermodynamic strain in domain A. A "thermodynamic strain" is detected when a residue that does not contribute favorably to the stability of the protein cannot be mutated without compromising the catalytic activity of the protein.

Finally, we proposed new mutations that appear to be good candidates to stabilize 1AQH (Table 4). The approach used here is complementary to structural analyses and experimental approaches, as it allows for refining the assumptions derived from the former method through energy calculations. Contrary to structural analyses, these calculations evaluate exactly the consequence of possible antagonist effects. To illustrate the complementarity of these approaches, most of the mutations found in domain C are located at its interface with domain A. This result agrees with the observation that the interactions between both domains are weaker in the psychrophilic α-amylase compared to its mesophilic counterparts.[15] We would like to stress that the thermodynamic effects of the mutations proposed in Table 4 should be additive if the mutations are not coupled, i.e., if the mutated residues do not interact.

ACKNOWLEDGMENT

I thank Yves Dehouck and Marianne Rooman for their critical comments. I acknowledge support from the Communauté Française de Belgique through the Action de Recherche Concertée no. 02/07-289 and from the E.C. through the Concerted Action Qol 2001-3.8.4.

Supporting Information Available: Mutations of 1AQH computed with a$\Delta\Delta G$ lower than or equal to −1.0 kcal/mol (Table SI1). This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Coutinho, P. M.; Henrissat, B. 1999, Carbohydrate-Active Enzymes server at http://afmb.cnrs-mrs.fr/CAZY/
(2) Vihinen, M.; Mäntsälä, P. Microbial amylolytic enzyme. *Crit. Rev. Biochem. Mol. Biol*. **1989**, *24*, 329−418.
(3) Brayer, G.; Yaoguang, L.; Withers, S. The structure of human pancreatic alpha-amylase at 1.8 Å resolution and comparisons with related enzymes. *Protein Sci*. **1995**, *4*, 1730−1742.
(4) Janeček, S. alpha-Amylase family: molecular biology and evolution. *Prog. Biophys. Mol. Biol*. **1997**, *67*, 67−97.
(5) Nielsen, J. E.; Borchert, T. B. Protein engineering of bacterial alpha-amylases. *Biochim. Biophys. Acta* **2000**, *1543*, 253−274.
(6) Declerck, N.; Machius, M.; Joyet, P.; Wiegand, G.; Huber, R.; Gaillardin, C. Engineering the thermostability of *Bacillus licheniformis* alpha-amylase. *Biologia (Bratislava)* **2002**, *57 suppl. 11*, 203-211.
(7) D'Amico, S.; Claverie, P.; Collins, T.; Georlette, D.; Gratia, E.; Hoyoux, A.; Meuwis, M.-A.; Feller, G.; Gerday C. Molecular basis of cold adaptation. *Philos. Trans. R. Soc. London B* **2002**, *357*, 917−925.
(8) Feller, G.; Gerday, C. Psychrophilic enzymes: hot topics in cold adaptation. *Nat. Rev. Microbiol.* **2003**, *1*, 200−208.
(9) Feller, G. Molecular adaptations to cold in psychrophilic enzymes. *Cell Mol. Life Sci.* **2003**, *60*, 648−662.
(10) Jaenicke, R.; Böhm, G. The stability of proteins in extreme environments. *Curr. Opin. Struct. Biol*. **1998**, *8*, 738−748.
(11) Kumar, S.; Nussinov, R. How do thermophilic proteins deal with heat? *Cell. Mol. Life Sci.* **2001**, *58*, 1216−1233.
(12) Vieille, C.; Zeikus, G. J. Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. *Microbiol. Mol. Biol. Rev*. **2001**, *65*, 1−43.
(13) Berezovsky, I. N.; Shakhnovich, E. I. Physics and evolution of thermophilic adaptation. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 12742−12747.
(14) Berezovsky, I. N.; Chen, W. W.; Choi, P. J.; Shakhnovich, E. I. Entropic stabilization of proteins and its proteomic consequences. *Plos Comput. Biol.* **2005**, *1*, e47.
(15) Aghajari, N.; Feller, G.; Gerday, C.; Haser, R. Crystal structures of the psychrophilic alpha-amylase from Alteromonas haloplanctis in its native form and complexed with an inhibitor. *Structure* **1998**, *6*, 1503−1516.
(16) D'Amico, S.; Marx, J.-C.; Gerday, C.; Feller, G. Activity-stability relationships in extremophilic enzymes. *J. Biol. Chem.* **2003**, *278*, 7891−7896.
(17) Jaenicke, R. Protein stability and molecular adaptation to extreme conditions. *Eur. J. Biochem*. **1991**, *202*, 715−728.
(18) Fitter, J.; Haber-Pohlmeier, S. Structural stability and unfolding properties of thermostable bacterial alpha-amylases: a comparative study of homologous enzymes. *Biochemistry* **2004**, *43*, 9589−9599.
(19) Declerck, N.; Joyet, P.; Gaillardin, C.; Masson, J.-M. Use of amber suppressors to investigate the thermostability of *Bacillus licheniformis* alpha-amylase. Amino acid replacements at 6 histidine residues reveal a critical position at His-133. *J. Biol. Chem*. **1990**, *265*, 15481−15488.
(20) Conrad, B.; Hoang, V.; Polley, A.; Hofemeister, J. Hybrid Bacillus amyloliquefaciens X *Bacillus licheniformis* alpha-amylases. Construction, properties and sequence determinants. *Eur. J. Biochem.* **1995**, *230*, 481−490.

(21) Declerck, N.; Machius, M.; Chambert, R.; Wiegand, G.; Huber, R.; Gaillardin, C. Hyperthermostable mutants of *Bacillus licheniformis* alpha-amylase: thermodynamic studies and structural interpretation. *Protein Eng.* **1997**, *10*, 541−549.
(22) Shaw, A.; Bott, R.; Day, A. G. Protein engineering of alpha-amylase for low pH performance. *Curr. Opin. Biotech.* **1999**, *10*, 349−352.
(23) Declerck, N.; Machius, M.; Wiegand, G.; Huber, R.; Gaillardin, C. Probing structural determinants specifying high thermostability in *Bacillus licheniformis* alpha-amylase. *J. Mol. Biol.* **2000**, *301*, 1041−1057.
(24) D'Amico, S.; Gerday, C.; Feller, G. Structural determinants of cold adaptation and stability in a large protein. *J. Biol. Chem.* **2001**, *276*, 25791−25796.
(25) Declerck, N.; Machius, M.; Joyet, P.; Wiegand, G.; Huber, R.; Gaillardin, D. Hyperthermostabilization of *Bacillus licheniformis* alpha-amylase and modulation of its stability over a 50 degrees C temperature range. *Protein Eng.* **2003**, *16*, 287−293.
(26) Machius, M.; Declerck, N.; Huber, R.; Wiegand, G. Kinetic stabilization of *Bacillus licheniformis* alpha-amylase through introduction of hydrophobic residues at the surface. *J. Biol. Chem.* **2003**, *278*, 11546−11553.
(27) D'Amico, S.; Gerday, C.; Feller, G. Temperature adaptation of proteins: engineering mesophilic-like activity and stability in a cold-adapted alpha-amylase. *J. Mol. Biol.* **2003**, *332*, 981−988.
(28) Shahhoseini, M.; Ziaee, A.-A.; Pourbabai, A.-A.; Ghaemi, N.; Declerck, N. A natural variant of *Bacillus licheniformis* alpha-amylase isolated from flour mill wastewaters sheds light on the origin of high thermostability. *J. Appl. Microbiol.* **2005**, *98*, 24−32.
(29) Gilis, D.; Rooman, M. PoPMuSiC, an algorithm for predicting protein mutant stability changes: application to prion proteins. *Protein Eng.* **2000**, *13*, 849−856.
(30) Kwasigroch, J. M.; Gilis, D.; Dehouck, Y.; Rooman, M. PoPMuSiC, rationally designing point mutations in protein structures. *Bioinformatics* **2002**, *18*, 1701−1702.
(31) Elcock, A. H. The stability of salt bridges at high temperatures: implications for hyperthermophilic proteins. *J. Mol. Biol.* **1998**, *284*, 489−502.
(32) Thomas, A. S.; Elcock, A. H. Molecular simulations suggest protein salt bridges are uniquely suited to life at high temperatures. *J. Am. Chem. Soc.* **2004**, *126*, 2208−2214.
(33) Kocher, J.-P.; Rooman, M. J.; Wodak, S. J. Factors influencing the ability of knowledge-based potentials to identify native sequence-structure matches. *J. Mol. Biol.* **1994**, *235*, 1598−1613.
(34) Rooman, M.; Gilis, D. Different derivations of knowledge-based potentials and analysis of their robustness and context-dependent predictive power. *Eur. J. Biochem.* **1998**, *254*, 135−143.
(35) Dehouck, Y.; Gilis, D.; Rooman, M. Database-derived potentials dependent on protein size for in silico folding and design. *Biophys. J.* **2004**, *87*, 171−181.
(36) Feller, G.; D'Amico, S.; Gerday, C. Thermodynamic stability of a cold-active alpha-amylase from the Antarctic bacterium Alteromonas haloplanctis. *Biochemistry* **1999**, *38*, 4613−4619.
(37) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(38) Machius, M.; Vértesy, L.; Huber, R.; Wiegand, G. Carbohydrate and protein-based inhibitors of porcine pancreatic alpha-amylase: structure analysis and comparison of their binding characteristics. *J. Mol. Biol.* **1996**, *260*, 409−421.
(39) Ramasubbu, N.; Paloth, V.; Luo, Y.; Brayer G. D.; Levine M. J. Structure of Human Salivary α-Amylase at 1.6 Å Resolution: Implications for its Role in the Oral Cavity. *Acta Crystallogr.* **1996**, *D52*, 435−446.
(40) Strobl, S.; Maskos, K.; Betz, M.; Wiegand, G.; Huber, R.; Gomis-Rüth, X.; Glockshuber, R. Crystal structure of yellow meal worm alpha-amylase at 1.64 A resolution. *J. Mol. Biol.* **1998**, *278*, 617−628.
(41) Aghajari, N.; Feller, G.; Gerday, C.; Haser, R. Structures of the psychrophilic Alteromonas haloplanctis alpha-amylase give insights into cold adaptation at a molecular level. *Protein Sci.* **1998**, *7*, 564−572.
(42) Thompson, J. D.; Higgins, D. G.; Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **1994**, *22*, 4673−4680.
(43) Boutonnet, N. S.; Rooman, M. J.; Ochagavia, M. E.; Richelle, J.; Wodak, S. J. Optimal protein structure alignments by multiple linkage clustering: application to distantly related proteins. *Protein Eng.* **1995**, *8*, 647−662.
(44) Wintjens, R. T.; Rooman, M. J.; Wodak, S. J. Automatic classification and analysis of alpha alpha-turn motifs in proteins. *J. Mol. Biol.* **1996**, *255*, 235−253.
(45) Gilis, D.; McLennan, H. R.; Dehouck, Y.; Cabrita, L. D.; Rooman, M.; Bottomley, S. P. In vitro and in silico design of alpha1-antitrypsin mutants with different conformational stabilities. *J. Mol. Biol.* **2003**, *325*, 581−589.
(46) Gilis, D.; Rooman, M. Stability changes upon mutation of solvent-accessible residues in proteins evaluated by database-derived potentials. *J. Mol. Biol.* **1996**, *257*, 1112−1126.
(47) Hutchinson, E. G.; Thornton, J. M. PROMOTIF- -a program to identify and analyze structural motifs in proteins. *Protein Sci.* **1996**, *5*, 212−220.
(48) Laskowski, R. A.; Chistyakov, V. V.; Thornton, J. M. PDBsum more: new summaries and analyses of the known 3D structures of proteins and nucleic acids. *Nucleic Acids Res.* **2005**, *33*, D266−268.
(49) Orengo, C. A.; Michie, A. D.; Jones, S.; Jones, D. T.; Swindells, M. B.; Thornton, J. M. CATH- -a hierarchic classification of protein domain structures. *Structure* **1997**, *5*, 1093−1108.
(50) MacGregor E. A.; Janeèek, S.; Svensson B. Relationship of sequence and structure to specificity in the alpha-amylase family of enzymes. *Biochim. Biophys. Acta* **2001**, *1546*, 1−20.

CI050473V