TABLE OF CONTENTS

ABBREVIATIONS USED (in alphabetical order)

AHA	α-amylase from *Pseudoalteromonas haloplanktis*
AmyD	α-amylase from *Drosophila melanogaster*
EC	Enzyme Classification
EPS or Et-G7-pNP	4−nitrophenyl-α-D-maltoheptaoside-4,6-O-2-ethylidene
GAPDH	D-glyceraldehyde-3-phosphate dehydrogenase
GH	Glycoside Hydrolases
HPA	Human Pancreatic α-amylase
HSA	Human Salivary α-amylase
ITC	Isothermal Titration Calorimetry
IUBMB	International Union of Biochemistry and Molecular Biology
LBHB	Low Barrier Hydrogen Bond
NTF	Nuclear Transport Factor
PDB	Research Collaboratory for Structural Bioinformatics Protein Databank
PPA	Pig Pancreatic α-amylase
PVDF	Polyvinylidene fluoride
Q-TOF	Quadrupole-Time Of Flight
Rmsd	Root Mean Square (Deviation)
SH2	Src Homology 2
Src	Sarcomavirus
TMA	α-amylase from *Tenebrio molitor*
TMAO	Trimethylamine *N*-oxide

1.1 STARCH

1.1.1 Source

Starch is the major carbohydrate reserve in plant tubers and seed endosperm where it is found as granules typically containing several million amylopectin molecules accompanied by a much larger number of smaller amylose molecules. By far the largest source of starch is corn (maize) with other common sources being wheat, potato, tapioca and rice.

1.1.2 Structure

Starch consists of two types of molecules, amylose (normally 20-30%) and amylopectin (normally 70-80%). Both consist of polymers of α-D-glucose units: in amylose (Fig. 1) these are α(1-4) linked, with all the ring oxygen atoms being located on the same side, whereas in amylopectin (Fig. 2) about one residue in every twenty is also α(1-6) linked, thereby forming branch-points. The relative proportions of amylose to amylopectin and α(1-6) branch-points both depend on the source of the starch, *e.g.* amylomaizes contain over 50% amylose whereas "waxy" maize has almost none (~3%).

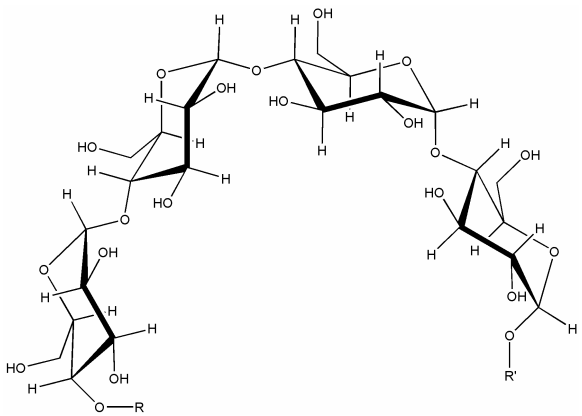


Fig. 1: Partial structure of amylose.

Amylose and amylopectin are fundamentally different molecules, amylose having a low molecular weight and a relatively extended shape (hydrodynamic radius 7-22 nm) whereas amylopectin has large but compact molecules (hydrodynamic radius 21-75 nm). Although the α(1-4) links are capable of relatively free rotation around the (φ) phi and (ψ) psi torsion angles, hydrogen bonding between the O3' and O2 oxygen atoms of sequential residues tends to encourage a left-handed helical conformation.

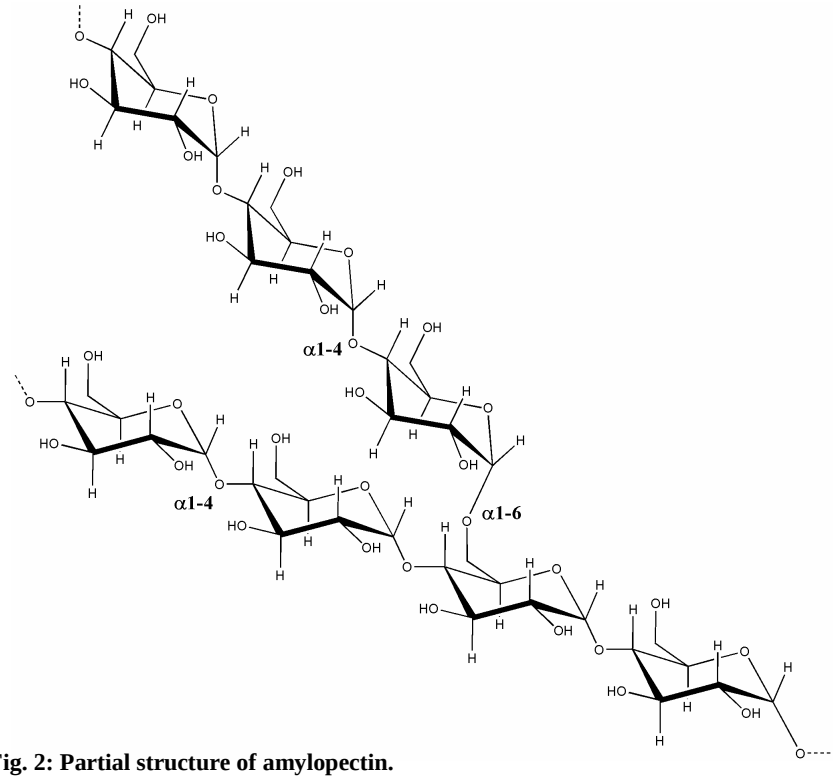
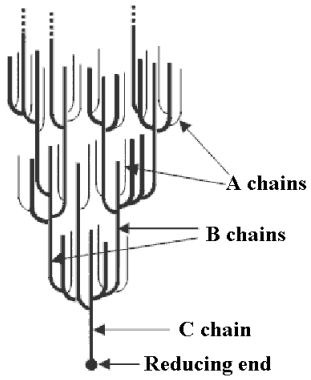


Fig. 2: Partial structure of amylopectin.

Amylose molecules consist of single, mainly unbranched chains with from 500 to 20,000 glucose units, depending on the source. In addition, a small number of α(1-6) branches and linked phosphate groups may also be found. The single helix amylose possesses a relatively hydrophobic inner surface that holds a spiral of water molecules which are relatively easily lost and replaced by hydrophobic molecules. This is also responsible for the characteristic binding of amylose to chains of charged iodine molecules (*e.g.* the polyiodides; chains of I_3^- and I_5^- forming structures such as I_9^{3-} and I_{15}^{3-}) where each turn of the helix holds about two iodine atoms and a blue color is produced due to donor-acceptor interactions between water and the electron deficient polyiodides. This reaction allows for an easy detection of the presence of starch if exposed to iodine, as well as of amylolytic trace activity if starch is used as an immobilized substrate.

Amylopectin contains up to two million glucose residues, displays a compact structure and is formed by non-random α(1-6) branching of the amylose-type α(1-4)-D-glucose structure. The branching is determined by branching enzymes which leave each chain with up to 30 glucose residues.



There are usually slightly more "outer" unbranched chains (A-chains) than "inner" branched chains (B-chains). There is only one chain (C-chain) containing the single reducing end (Fig. 3).

Fig. 3: Macromolecular structure of amylopectin.

1.1.3 Hydrolysis

The complete degradation of starch requires a number of different enzymes; the most important of which are summarized in Table 1.

Tab. 1: Characteristics of main amylolytic enzymes.

Enzyme (EC number)	Cleaved bonds	Characteristics
α-amylase (3.2.1.1)	α(1-4)	• endoenzyme capable of bypassing α(1-6) links • produces maltose and maltotriose
β-amylase (3.2.1.2)	α(1-4)	• exoenzyme unable to bypass α(1-6) links • requires a non-reducing end • produces maltose
Glucoamylase (3.2.1.3)	α(1-4) + α(1-6)	• produces glucose
Isoamylase (3.2.1.68) Pullulanase (3.2.1.41)	α(1-6)	• debranching enzymes

These enzymes are part of the large group of carbohydrate-degrading enzymes that are the glycoside hydrolases (GH). One possibility for classifying enzymes is shown in the preceding table, *i.e.* as a function of their substrate specificity. This classification is recommended by the International Union of Biochemistry and Molecular Biology (IUBMB) and attributes an EC number to each enzyme: EC 3.2.1.X for glycoside hydrolases where X represents the substrate specificity.

There are several problems associated with this classification system: firstly, multi-substrate enzymes cannot be taken into account. Secondly, highly homologous enzymes may be split into different and distant EC groups, even if they share evolutionary and structural features, as well as a common catalytic mechanism. Thirdly, this IUBMB classification system does not take into account stereochemical features, *e.g.* all GHs display a mechanism of action that either retains or inverts the anomeric configuration of the carbohydrate. The spatial position of the catalytic residues is of importance for determining the mechanism of action of the enzyme. The active site is generally constituted by two essential Asp or Glu residues typically separated by about 5 Å in retaining enzymes or about 10 Å in inverting enzymes. Fourthly, this classification system does not provide any information about whether enzymes act at the extremity (type *exo*) or internal (type *endo*) of the polysaccharide chain.

In 1991, Henrissat (Henrissat, 1991) proposed a new classification of glycoside hydrolases based on amino acid sequence similarities (http://afmb.cnrs-mrs.fr/CAZY). As sequence and structure of a protein are linked, relative information on the structure and the mechanism of action can be deduced from the primary structure alone, with the assumption that the stereospecificity and the molecular mechanism are conserved for a given family of enzymes. Some advantages of this classification are that (i) it reflects similar folds for the members of a family, (ii) it provides some information about evolutionary aspects. Indeed, examples to illustrate the latter statement are that the members of a family may hydrolyze different substrates but share a similar fold, thus reflecting a divergent evolution, or that enzymes displaying different folds but similar specificities indicate a convergent evolution.

The Carbohydrate-Active Enzymes database provides a continuously updated list of the glycoside hydrolase families. Because the fold of proteins is better conserved than their sequences, some of the families can be grouped in “clans” (i) when new sequences are found to be related to more than one family, (ii) when the sensitivity of sequence comparison methods is increased or (iii) when structural determinations demonstrate the resemblance between members of different families.

The present work is focused on α-amylases which are glycoside hydrolases that catalyze hydrolysis with retention of the anomeric configuration and belong to family GH13 of this classification. GH13 is constituted by a few thousand primary structures

of enzymes grouping 21 different substrate specificities (http://afmb.cnrs-mrs.fr/CAZY/fam/GH13.html). In this family, a total of approximately 200 3D structures have been elucidated: their common structural feature is a $(\beta/\alpha)_8$ barrel and thus, family GH13 is part of the GH-H clan. The members of this family share only weak sequence similarities, approximately 10 %, and only 10 residues (out of a few hundred) are strictly conserved. However, 7 regions are found to be highly conserved (Janecek, 1997).

1.2 α-AMYLASES

1.2.1 General structure

α-amylases display similar dimensions (40 Å x 50 Å x 80 Å) and molecular weights (≈ 50 kDa) and share a common general structure organized in three domains (Fig. 4).

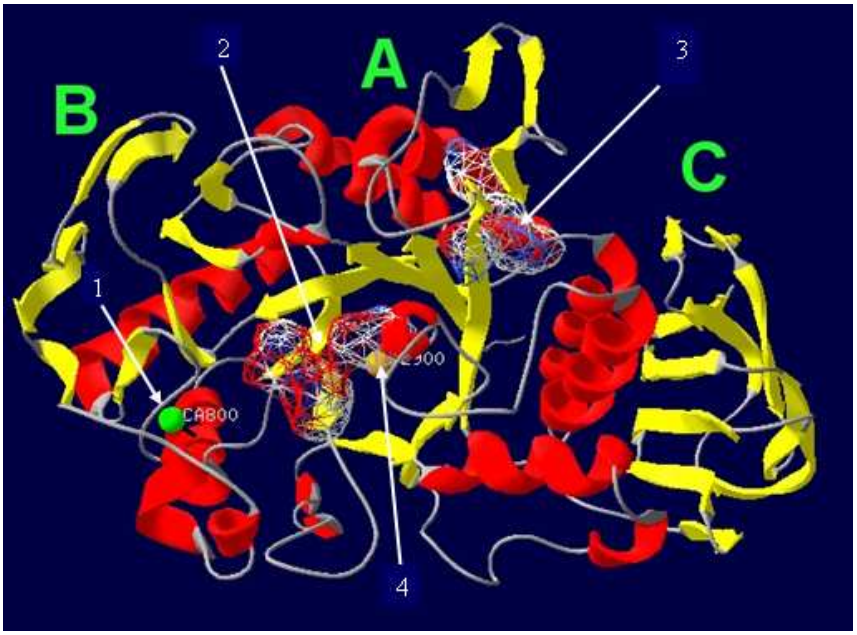


Fig. 4: General structure of α-amylases. 1) calcium ion; 2) active site; 3) non-catalytic triad and 4) chloride ion. Based on the PDB accession code 1AQH: α-amylase from *Pseudoalteromonas haloplanktis*.

The N-terminal extremity is located in the central domain **A** (residues 1-86 and 130-356) that adopts the fold of a $(\beta/\alpha)_8$ barrel and binds a chloride ion in its centre in all animal and some bacterial α-amylases (see below). Domain **B** (residues 87-129) is relatively small, formed only by β-strands and is inserted between the β3-strand and the α3-helix of domain A. This domain, with the exception of the region around the calcium-binding Asp residue, is less conserved in family GH13 in terms of size and sequence (Janecek, 1997). The active site is located in the large cleft between domains A and B, and therefore domain B determines some functional and structural properties of the enzymes: adaptation of activity to pH, efficiency of inhibitors, correct folding and substrate specificity. Domain **C** (residues 357-448) is the C-terminal domain; its 10 β-strands are organized in a Greek key motif and adopt a globular shape. The function of domain C remains unclear, but mutagenesis and deletion experiments have shown domain C to be essential for catalysis.

For their structural integrity as well as for enzymatic activity, all known α-amylases require octo-coordinated binding of at least one Ca^{2+} ion (Qian *et al.*, 1993) between domains A and B, *i.e.* close to the active site (≈ 12 Å). This coordinated calcium ion has no direct function during catalysis but it is responsible for holding domains A and B together, thus shaping the catalytic cleft. In some cases, additional calcium ions and their binding sites have been observed, such as for the α-amylase from *Aspergillus niger*. For information, in the latter case the binding of a second calcium ion has been shown to involve two catalytic residues and this might explain the inhibitory effect of high calcium concentrations on this enzyme (Boel *et al.*, 1990).

In addition to calcium, one particular group of α-amylases requires the binding of a chloride ion as allosteric activator (Levitzki and Steer, 1974) (see §1.2.4 Catalytic mechanism). The binding site of this Cl^- ion is located in the center of domain A, close to the C-terminal end of the $(\beta/\alpha)_8$ barrel and at ≈ 6 Å from the active site. The allosteric activation by chloride is typical for all animal α-amylases and some bacterial α-amylases.

Approximately 300 primary structures (August 2006) of α-amylases are known, including those from Archaea, Bacteria and Eukaryota, around 80 of which are from chloride-dependent α-amylases. Of a total of 19 available and non-redundant α-amylase 3D structures, 5 are of chloride-dependent enzymes: (i) the bacterial and

psychrophilic α-amylase from *Pseudoalteromonas haloplanktis* (AHA), formerly known as *Alteromonas haloplanctis* (Aghajari *et al.*, 1998a, Aghajari *et al.*, 1998b), (ii) the larval α-amylase from insect *Tenebrio molitor* (TMA) (Strobl *et al.*, 1998), (iii) the porcine (*Sus scrofa*) pancreatic α-amylase (PPA) (Buisson *et al.*, 1987), and (iv) the human salivary (HSA) (Ramasubbu *et al.*, 1996) and (v) pancreatic (HPA) α-amylases (Brayer *et al.*, 1995).

1.2.2 Phylogeny

Amino acid sequence alignments of α-amylases reveal the existence of three compact groups (Janecek, 1994) (Fig. 5A): a) α-amylases from fungi and yeast, b) from plants and c) from animals and streptomycetes. The primary structures of bacterial α-amylases are more diversified.

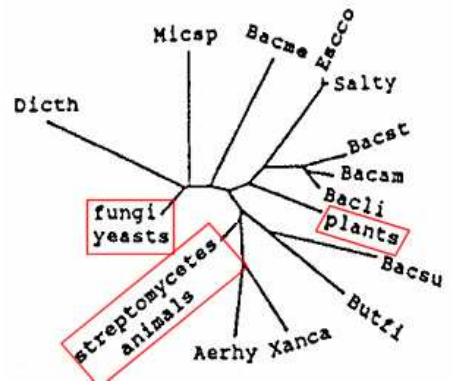


Fig. 5A: Rootless phylogenetic tree of α-amylases.

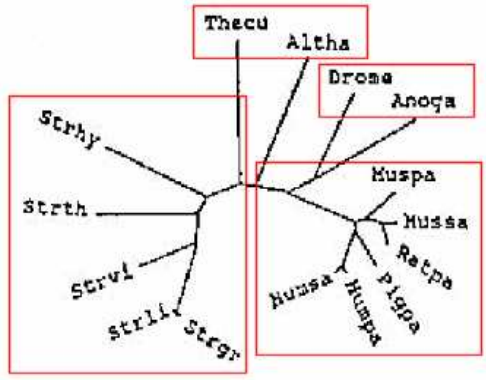


Fig. 5B: Rootless phylogenetic tree of α-amylases from streptomycetes, mammals, insects and two bacterial species.

Figure 5B illustrates the relationships between the members of group c (Fig. 5A: α-amylases from streptomycetes, mammals and insects) and α-amylases from the bacteria *Thermomonospora curvata* and *Pseudoalteromonas haloplanktis*. In this group, the α-amylases from the streptomycetes differ from the others by their independence of chloride.

1.2.3 Phylogeny of Cl^--dependent α-amylases

Allosteric activation of α-amylases by chloride was for a long time thought to be restricted to those from mammals (Brayer *et al.*, 1995), until the discovery of this feature for the α-amylase from *Pseudoalteromonas haloplanktis*. D'Amico *et al.* (D'Amico *et al.*, 2000) have aligned and compared α-amylase sequences in order to analyze structural similarities and evolutionary relationships of Cl^--dependent α-amylases. The identification of such enzymes was based on the presence of the chloride coordinating residues Arg195, Asn298 and Arg/Lys337 (PPA numbering), and led to the recognition of 38 potentially Cl^--dependent α-amylases out of the 200 sequences known at that time. The authors proposed the unrooted evolutionary tree shown in Figure 6.

The multiple sequence alignment revealed a remarkable homology, even between distant organisms such as bacteria and vertebrates (≈ 40 %). Indeed, among the 470-500 residues corresponding to an average total number of residues for α-amylases, 60 are strictly conserved.

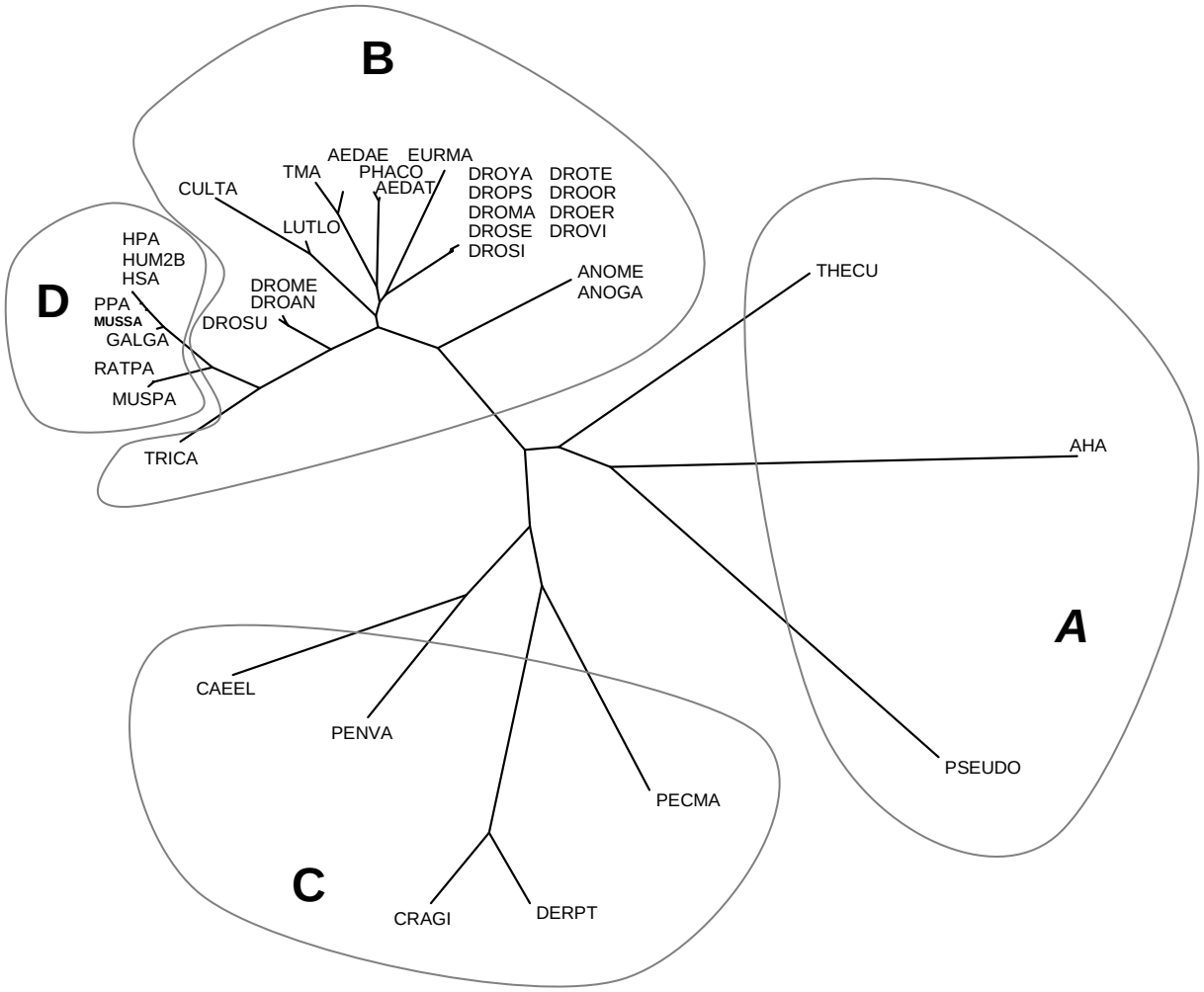


Fig. 6: Unrooted evolutionary tree of chloride-dependent α-amylases. For abbreviations, refer to addendum 2. The branch lengths are proportional to the sequence divergence. Cluster *A*: bacteria; cluster *B*: insects and acari; cluster *C*: mollusks, crustaceans and nematodes; cluster *D*: mammals and birds. Source: (D'Amico *et al.*, 2000).

1.2.4 Catalytic mechanism

The catalytic residues Asp195, Glu223 and Asp300 (PPA numbering) are located on the C-terminal side of the β-strands of domain A and their side chains are oriented towards the catalytic cleft between domains A and B. The first two residues act respectively as nucleophile and acid/base catalyst while the role of Asp300 is less certain. It has been proposed that this residue is implicated either in the stabilization of the oxocarbenium intermediate state (Uitdehaag *et al.*, 1999) or in maintaining the correct protonation state of the catalytic Glu.

As mentioned, α-amylases or α(1-4)-D-glucan-4-glucanohydrolases, hydrolyze α(1-4) O-glycosidic bonds with retention of the anomeric configuration. The catalytic mechanism has been proposed to be divided in two steps (Fig. 7) (McCarter and Withers, 1994): glycosylation and deglycosylation. First, the carboxylic group of Asp195 carries out a nucleophilic attack on the anomeric carbon of the substrate leading to the first intermediate oxocarbenium state. This step is facilitated by a general acid/base catalysis implicating the carboxylic function(s) of Glu223 alone or of Glu223 and Asp300, leading to a weakened C1-O bond and the formation of the covalent β-glycosyl-enzyme intermediate. The hydrolysis of the covalent intermediate state starts with the attack of a water molecule (R''-OH), which is activated by the carboxylate group of Glu223, on the anomeric carbon. Again, an intermediate state is formed, leading finally to product formation with retention of the anomeric

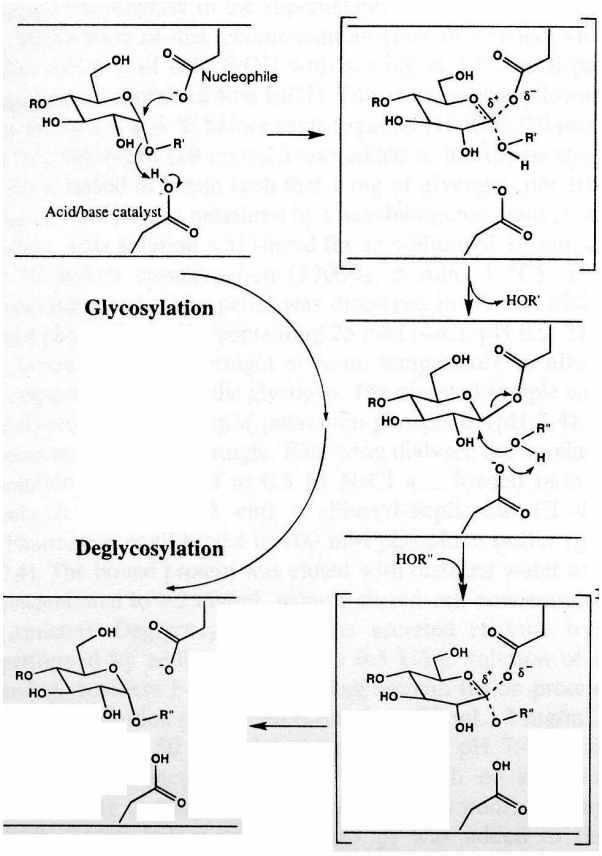


Fig. 7: Catalytic mechanism of α-amylases.

configuration and the initial protonation state of the catalytic residues. If R' represents a glucose monomer and R'' a hydrogen atom, we indeed observe the hydrolysis of starch with formation of a new reducing end. The presence of glucidic residues R and R' illustrates the fact that α-amylases are endoenzymes.

The role of chloride as an allosteric activator was first discussed by Levitzky and Steer in 1974 (Levitzki and Steer, 1974) and has since been extensively analyzed for AHA (Feller *et al.*, 1996, Aghajari *et al.*, 2002), HPA (Numao *et al.*, 2002) and PPA (Qian *et al.*, 2005). All authors agree that chloride is not directly implicated in substrate binding, although the chloride binding site is located close to the center of the central barrel, *i.e.* close to the catalytic and the calcium binding sites. Moreover, all authors agree on the fact that the chloride anion is necessary to increase the pK_a value of the catalytic residue Glu223 so as to maintain the side chain of this residue in the protonated state required for catalysis. Qian *et al.* (2005) have demonstrated that in addition to the induced pK_a shift, the direct effect of chloride binding is to orient the glutamic side chain so as to optimize the catalytic process.

1.2.5 α-amylase from *Pseudoalteromonas haloplanktis*

Pseudoalteromonas haloplanktis, formerly known as *Alteromonas haloplanctis*, is an aerobic, Gram-negative coccobacillus strain that has been isolated from sea water near the Antarctic research base Dumont D'Urville in Terre Adélie. The Cl^--dependent α-amylase secreted by *Pseudoalteromonas haloplanktis* has first been described by (Feller *et al.*, 1992): the mature primary structure counts 453 amino acid residues with a molecular weight of 49,340 Da and the functional enzyme displays an optimal pH for activity around neutrality and a pI of 5,5. Interestingly, its closest structural homologue is the porcine pancreatic α-amylase (PPA) with 53 % identity and this surprising fact has been attributed to horizontal gene transfer from animal to bacterium (Da Lage *et al.*, 2004).

Moreover, in agreement with the activity-stability-flexibility relationship[1], AHA presents the general characteristics of psychrophilic enzymes (Feller *et al.*, 1994): (i) a catalytic constant 7 times higher at low temperatures than for PPA, the mesophilic counterpart, (ii) a reduced conformational stability of approximately 30 kJ/mole when compared to PPA, (iii) an apparent T_{opt} for activity of 30 °C in contrast to 50 °C for PPA, and (iv) highly increased dissociation constants for calcium (2,000 times higher than PPA) and chloride (20 times higher than PPA) ions. The primary structure of AHA contains 13 proline residues instead of 21 for PPA, 13 arginine residues versus 29 for PPA and the folded enzyme displays 4 disulfide bridges instead of 5 for PPA.

AHA was the first psychrophilic enzyme to be crystallized (Aghajari *et al.*, 1996) and the 3D structure was published in 1998 (Aghajari *et al.*, 1998a) with the Protein Data Bank accession code 1AQH.

1.2.6 α-amylase from *Drosophila melanogaster*

The α-amylase from *Drosophila melanogaster* is a highly polymorphic enzyme at both the allozyme and specific activity level. The enzyme we used in our study is coded for by the gene *amy-p* with a length of 2141 bp for the complete coding sequence. The sequence has been deposited in the EMBL data bank with the accession code AB042870 and the protein is catalogued in the SwissProt data bank under the name AmyA_Drome with the accession code P08144. The protein has been produced, purified and crystallized, but the 3D structure has not yet been elucidated. Based on sequence similarities, this mesophilic α-amylase is thought to be chloride-dependent. It shares 41 % identity with AHA, 54 % with PPA and 62 % with TMA. Similarities are 56 % with AHA, 67 % with PPA and 75 % with TMA.

[1] At low temperatures: the higher the activity, the more flexible an enzyme needs to be, which is achieved by lowering the stability. This relationship is the generally accepted hypothesis for cold adaptation of enzymes.

1.2.7 The non-catalytic triad

The occurrence of a triad mimicking the active site of serine proteases and lipases is certainly the most surprising feature of chloride-dependent α-amylases. The first one to report the presence of such a motif in α-amylases was David Blow in 1990 (Blow, 1990), who screened the PDB for "lurking" triads and identified together with Janet Thornton (European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK) the TAKA-amylase (α-amylase from *Aspergillus oryzae*) as a candidate, albeit that this enzyme bears a strongly distorted triad. No function has been attributed to this motif. Aghajari *et al.* (Aghajari *et al.*, 1998a) reported the presence of a triad in AHA, and after multiple sequence alignments of chloride-dependent α-amylases, it became clear that this triad was actually conserved in all known chloride-dependent α-amylases, from bacterium to human (D'Amico *et al.*, 2000). In 1998 (Aghajari *et al.*, 1998a), the authors proposed the triad to be responsible for autoproteolysis of α-amylases, as in certain buffers devoid of chloride and calcium, these enzymes are rapidly cleaved into two peptides. We now know that this hypothesis was incorrect.

In AHA, the motif is situated at the surface of the $(\beta/\alpha)_8$ barrel in close proximity to the C-terminal domain and at 22 Å from the active site (Fig. 4), and has low solvent accessibility. It is formed by residues Ser303, His337 and Glu19 (amino acid numbering based on AHA is used throughout the document, unless stated otherwise) and can be superimposed with the catalytic triads of serine proteases and lipases (Fig. 8A), as well as with the triads of the five chloride-dependent α-amylase crystal structures (Fig. 8B), demonstrating a high degree of conservation in the stereochemistry of the three side chains.

However, differences between the non-catalytic triad of α-amylases and the catalytic triad of proteases and lipases do exist and are (i) a Glu residue replacing Asp and (ii) a flipped position of the His residue (easily observable in Fig. 8A) the ND1 atom of which now faces the Ser OG atom instead of the carboxylate group. Throughout this document, triads with the His residue in its flipped position will be designated as being in an amylase-like configuration, in contrast with the protease-like orientation.

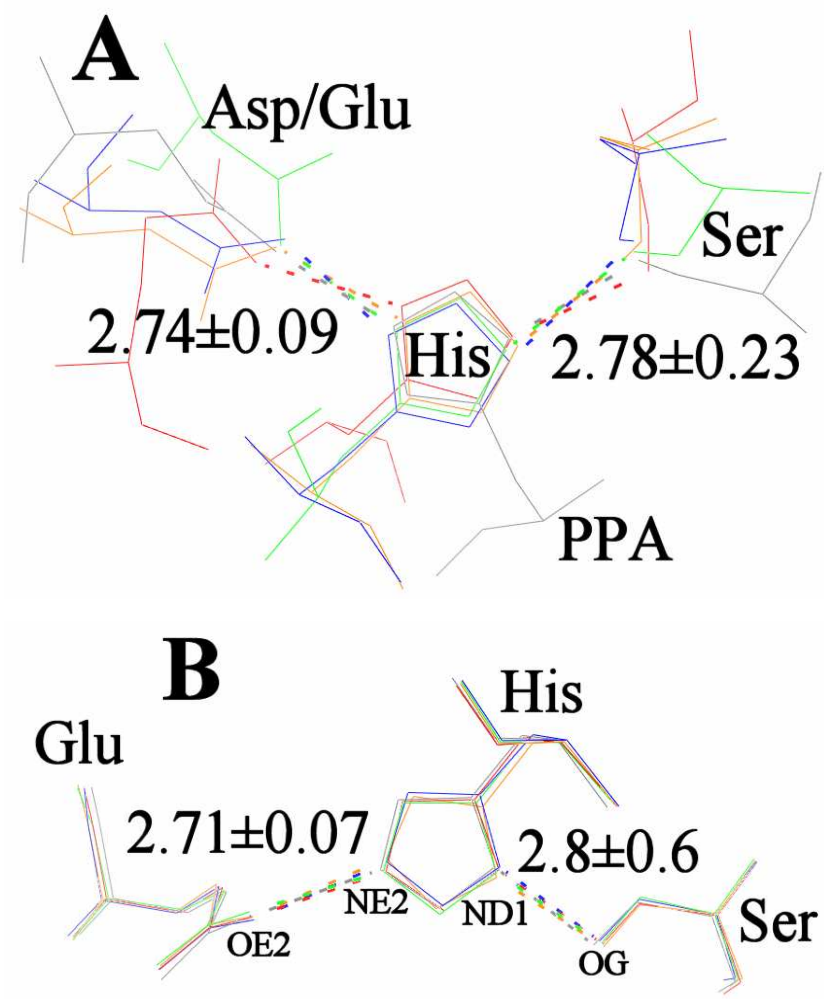
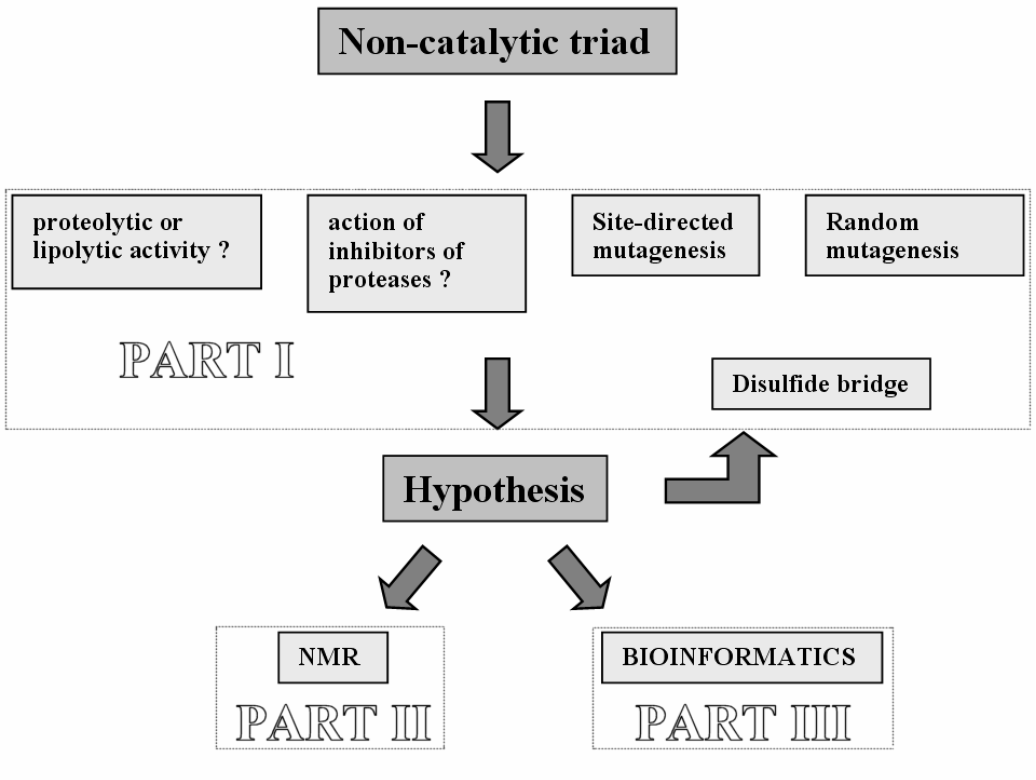


Fig. 8: Structural superimposition of catalytic and non-catalytic triads. *A*, the non-catalytic triad of pig pancreatic α-amylase (PPA, black, labeled) is superimposed with the active site catalytic triads of trypsin (blue), chymotrypsin (orange), subtilisin (green) and lipase (red). *B*, superimposition of the non-catalytic triads in the five crystal structures of chloride-dependent α-amylases: porcine pancreatic (orange), human pancreatic (green), human salivary (blue), insect (red) and bacterial (black) α-amylases. Interatomic distances are given in Å and the four functional atoms are indicated in panel B.

1.2.8 Aim of this study

To our knowledge, the three publications cited above are the only publications that ever mentioned the non-catalytic triad in α-amylases; this motif has never been reported to be present in other proteins. Hence, its role remains unknown and the aim of this study is to unravel the function of this highly conserved motif by means of mutational, biophysical and biochemical approaches. The structure of this work can be schematized as follows with a flow-chart.



The first part focuses on the mutational study of the triad in several α-amylases and led to our working hypothesis. In part II, we then test the validity of the working hypothesis by NMR spectroscopy and finally, we investigate the presence of such a triad in other proteins by bioinformatic means.

2.1 INTRODUCTION

This chapter corresponds to Part I in the flow-chart.

At the beginning of this project, we knew that the triad was highly conserved at the level of both sequence and structure among the chloride-dependent α-amylases. On the other hand, the role that this triad might fulfill was unknown. In order to unravel its function, we firstly tested an α-amylase for proteolytic and lipolytic activity. We then investigated the effects of mutating the triad on the behavior of several α-amylases.

2.2 MATERIALS AND METHODS

2.2.1 Bacterial strains

***Escherichia coli* RR1**

This strain was used because of its high growth rate and its capacity to produce high amounts of recombinant α-amylase at 18 °C.

***Escherichia coli* BL21(DE3)**

E. coli BL21(DE3) is an all-purpose strain for high level protein expression, commercialized by Stratagene. We used the strain for the production of genes that are cloned in a pET vector, thus allowing induction by IPTG. The designation DE3 indicates that this host carries in its genomic DNA the sequence coding for the prophage lambda DE3 containing phage T7 RNA polymerase gene under the control of a strong promoter *lacUV5*.

***Epicurian Coli*® XL1-Blue**

The advantage of this strain is its high transformation efficiency. These supercompetent cells are commercialized by Stratagene.

***Epicurian Coli*® XL1-Red**

These mutagenic competent cells, developed by Stratagene, were used for random mutagenesis experiments. The strain lacks three DNA repair systems, *mutS*, *mutD* and *mutT*, which multiplies the error frequency by 5,000 compared to the wild-type *E. coli* strain.

2.2.2 Plasmids

All plasmids confer ampicillin resistance and were purified on a Nucleospin column (Macherey-Nagel).

pαH12wt*

A stop codon was introduced after the last residue of the mature enzyme in order to delete the C-terminal domain of the precursor of the psychrophilic α-amylase from *Pseudoalteromonas haloplanktis*. The gene and its peptide signal were then cloned into a pUC12 plasmid, after the promoter *lacZ*, by ligation of the *SmaI* restriction site of the polylinker of the plasmid with the *HpaI* restriction site located 60 bp upstream of the start codon of the peptide signal. pαH12wt* (4.23 kb) constitutively expresses the enzyme in *E. coli* RR1.

pET-22b(+)

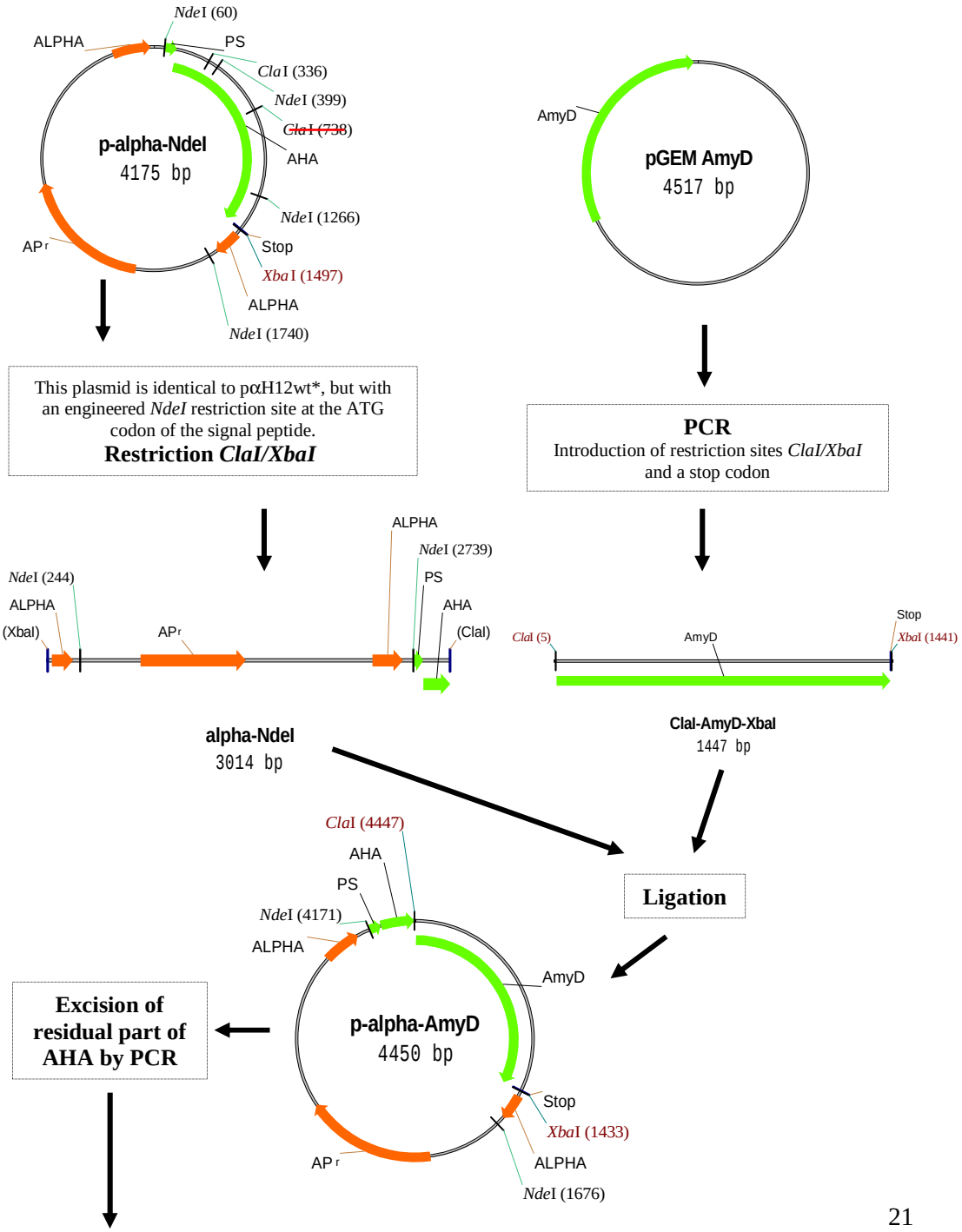
This plasmid is part of the pET System vectors developed by Novagen. It is a powerful vector for the expression of recombinant proteins in *E. coli*. Target genes are under control of strong bacteriophage T7 transcription and translation signals. Expression is induced by providing a source of T7 RNA polymerase in the host cell. Thus, we used this vector in combination with the *E. coli* BL21(DE3) strain. Another important benefit of this system is its ability to maintain target genes transcriptionally silent in the uninduced state.

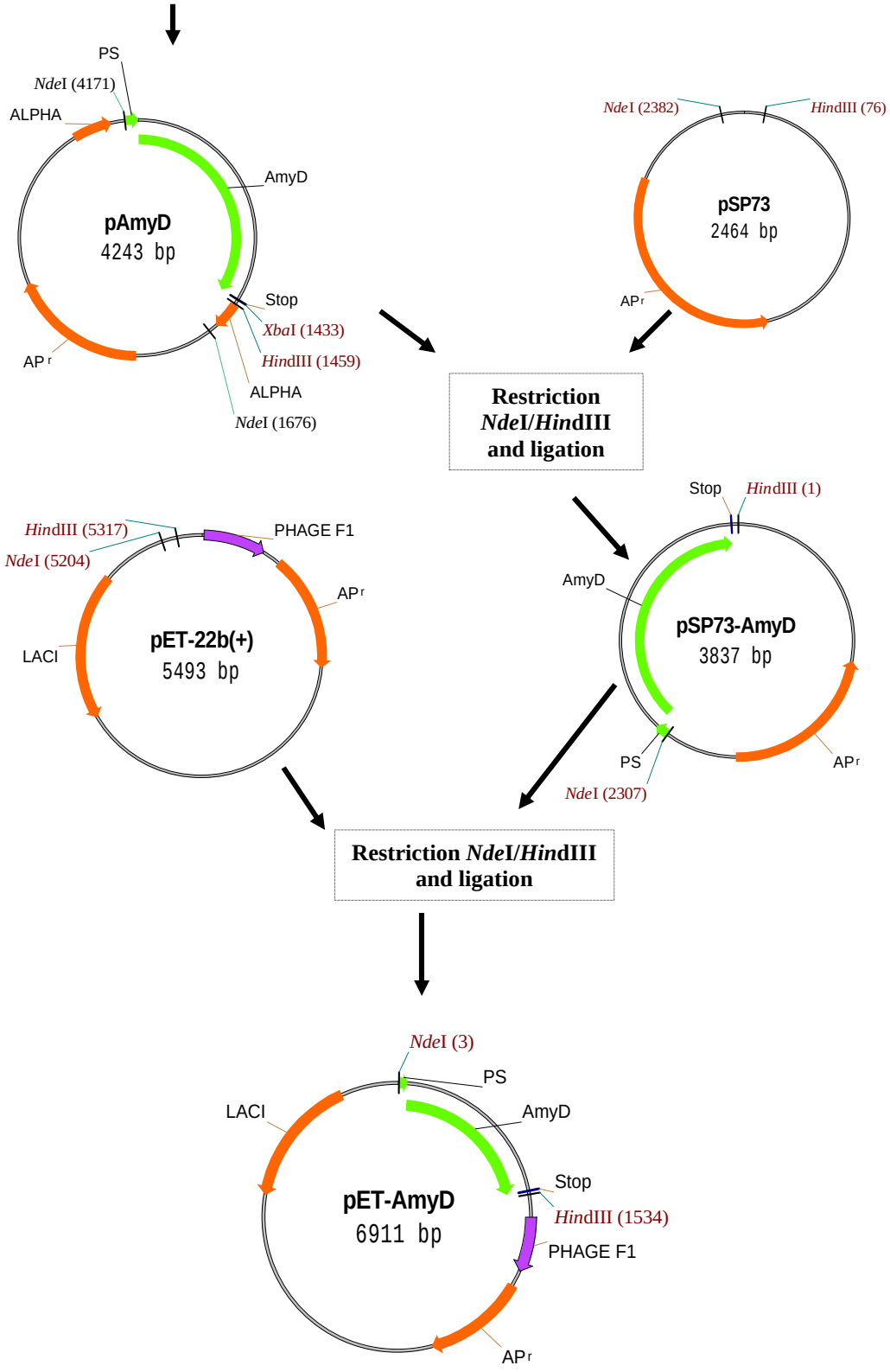
pSP73

The pSP73 vector is a small sized plasmid that offers a wide range of restriction sites and thus, was used for the subcloning of genes before their final transfer into the pET vector.

pET-AmyD

This plasmid was constructed according to the following strategy. The *amy-p* gene (coding for AmyD) in the pGEM-T Easy vector was kindly provided by Dr. Jean-Luc Da Lage from UPR 9034 Laboratoire des Populations, Génétique et Evolution, CNRS, 91198 Gif-sur-Yvette cedex, France.




PS
NdeI (4171)
ALPHA
AmyD
pAmyD
4243 bp
Stop
XbaI (1433)
HindIII (1459)
ALPHA
NdeI (1676)
APr
NdeI (2382)
HindIII (76)
pSP73
2464 bp
APr
Restriction NdeI/HindIII and ligation
HindIII (5317)
PHAGE F1
NdeI (5204)
APr
LACI
pET-22b(+)
5493 bp
Stop
HindIII (1)
AmyD
pSP73-AmyD
3837 bp
PS
APr
NdeI (2307)
Restriction NdeI/HindIII and ligation
NdeI (3)
PS
LACI
AmyD
Stop
pET-AmyD
6911 bp
HindIII (1534)
PHAGE F1
APr

2.2.3 Culture media

Luria Bertani (LB) medium was prepared with 5 g/L yeast extract (Difco); 10 g/L tryptone (Difco); 10 g/L NaCl and 1 mM $CaCl_2.2H_2O$; pH 7.2. If agar (16 g/L) and ampicillin are added, the medium is designated by the abbreviation LBAA. We used LB in the random mutagenesis experiment and for the preparation of Petri dishes, except if stated otherwise.

Tryptone-Yeast-Phosphate (TYP) medium contained 16 g/L yeast extract (Difco); 16 g/L tryptone (Difco); 5 g/L NaCl; 0.1 µM $CaCl_2.2H_2O$ and 2.5 g/L $K_2HPO_4.2H_2O$; pH 7.0. This rich culture medium was used for the production of bacterial α-amylases.

Terrific broth (TB) was composed of 24 g/L yeast extract (Difco); 12 g/L tryptone (Difco); 5.04 g/L glycerol; 2.31 g/L KH_2PO_4 and 12.54 g/L $K_2HPO_4.2H_2O$; pH 7.0. This rich culture medium was used for the production of insect α-amylases.

SOC broth was composed of 5 g/L yeast extract (Difco); 20 g/L tryptone (Difco); 10 mM NaCl; 2.5 mM KCl; 10 mM $MgCl_2.6H_2O$; 10 mM $MgSO_4.7H_2O$ and 20 mM glucose; pH 7.2. This medium was essentially used in the transformation protocol.

Ampicillin was filtered (0.22 µm) and added to the respective media after autoclaving and cooling to 50 °C.

2.2.4 Activity tests

(i) Lipolytic and proteolytic activities of the α-amylase were assayed on immobilized substrates. In both cases, the standard medium for Petri dishes was 50 mM HEPES buffer at pH 7.2 to which 15 g/L of agar was added. Before autoclaving, either 0.5 % tributyrine, previously emulsified by sonication, for detection of lipolytic activity, or 1 % casein for detection of proteolytic activity was added. 5 µL of sample, previously sterilized by filtration on a 0.22 µm Millipore filter and containing purified 0.35 mM *Pseudoalteromonas haloplanktis* α-amylase (in 10 mM Tris buffer; 25 mM

NaCl; 1 mM $CaCl_2.2H_2O$; 1 mM NaN_3, pH 8) was deposited on a Petri dish which was then incubated at room temperature for one month.

(ii) Amylolytic activity was measured using 3.5 mM 4−nitrophenyl-α-D-maltoheptaoside-4,6-O-2-ethylidene (Et-G7-pNP or EPS) as substrate (Boehringer Mannheim). The reaction medium contains an excess of α-glucosidase as coupling enzyme in 50 mM HEPES, 50 mM NaCl, 10 mM $MgCl_2$, pH 7.2. Activities were recorded in a thermostated Uvikon 860 spectrophotometer (Kontron) and calculated on the basis of an absorption coefficient for para-nitrophenolate of 8980 $M^{-1}cm^{-1}$ at 405 nm and a stoichiometric factor of 1.25 was applied.

2.2.5 Site directed mutagenesis

Two PCR-based techniques were used for the introduction of point mutations: the Inverse Polymerase Chain Reaction (IPCR) and the Quikchange Mutagenesis kit by Stratagene. The former was used in the introduction of the mutations in the pαH12wt* plasmid while the latter kit was used for mutating the pSP-AmyD vector.

(i) Concerning IPCR, this technique allows the amplification of plasmids and the introduction of a mutation, carried by only one of the two primers. The particularity of this technique is that primers hybridize adjacently, as shown in Fig. 9, leading to the amplification of the entire linear plasmid.

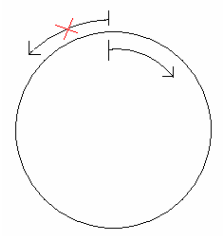


Fig. 9: Principle of IPCR. The red cross indicates the mutation.

The "sense" or mutated primer binds to the non-coding strand of the template DNA and is composed of 27 nucleotides, with 3 codons at the 5' end of the mutated codon and 5 codons at the 3' end. The "antisense" or silent primer is complementary to the coding strand of the template DNA and is composed of 24 nucleotides.

The reaction mixture was composed of 1 µL template DNA purified with a Nucleospin column; 3 µL dNTP 2mM; 5 µL of each primer at a concentration of

20 µM; 10 µL polymerase buffer 10x; 2 µL $MgSO_4$ 100 mM; 1 µL of Vent polymerase 10 units/µL and 73 µL of H_2O milliQ.

The mixture was heat denatured at 95 °C for 3 minutes, followed by 23 cycles composed of 1 minute of denaturation at 95 °C, 30 s at 55 °C to allow hybridization of the primers with the single stranded template DNA and 4 min, 15 s of elongation time (at 1 kbase/minute) at 72 °C. A final cycle with an extended elongation time (9 min, 15 s) completed the reaction.

PCR products (10 µL) were analyzed by electrophoresis on 1 % agarose gels and purified (90 µL) with the Qiaquick® PCR purification kit (Qiagen) according to the recommended protocol. Linear fragments were phospholigated with T4 kinase and ligase before transformation.

(ii) Introduction of a mutation with the Quikchange Mutagenesis kit was performed as recommended by the manufacturer. The main differences with the preceding technique are that both primers carry the desired mutation, they hybridize with complementary sequences of the template (Fig. 10), and the nicked DNA is repaired by the competent cells (*E. coli* XL1-Blue) after transformation.

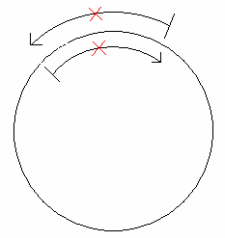


Fig. 10: Principle of Quikchange Mutagenesis kit. The red crosses indicate the mutations.

The nucleotide sequence of the constructs was checked with an Amersham Pharmacia Biotech ALF DNA sequencer.

2.2.6 Stabilizers

The stabilization assays of the α-amylase mutant S303A were performed directly on the supernatant. For each assay, sample volume was 1 mL and consisted of 0.86 mL supernatant and 0.14 mL "complete^{TM}Mini, EDTA-free Protease Inhibitor Cocktail" (Roche Applied Science) in phosphate buffer 100 mM, pH 7, so as to prevent α-amylase degradation by proteases present in the raw supernatant. To these samples were added the following stabilizing agents: $MgSO_4.7H_2O$, $Na_2SO_4.10H_2O$,

trimethylamine *N*-oxide (TMAO), sucrose, trehalose, glycerol, alanine and polyethyleneglycol (PEG) 4000. Final concentrations were 1 M for each stabilizer, except for PEG4000 which we used at a concentration of 0.1 M. A first blank consisted of only supernatant, the second of supernatant and "complete™Mini, EDTA-free Protease Inhibitor Cocktail". Assay temperature was maintained at 4 °C.

2.2.7 Kinetic parameters

Relative substrate specificity was measured by Isothermal Titration Calorimetry (ITC) as described by Lonhienne *et al.* (Lonhienne *et al.*, 2000) for the macromolecular substrate starch (100 % activity) and for shorter oligosaccharides such as maltooligosaccharides from corn syrup, maltopentaose and EPS. The substrates were prepared in 50 mM HEPES, 50 mM NaCl, 1 mM $CaCl_2.2H_2O$, pH 7.2 at a concentration of 2 % for starch and 5 mM for the oligosaccharides. The purified AHA and a sample of the culture supernatant of S303A were dialyzed overnight against the same buffer.

K_M values were determined for EPS by varying the substrate concentration in the above mentioned reaction medium (see activity tests). The initial reaction rates were fitted on the Michaelis-Menten equation for K_M calculation.

2.2.8 Random mutagenesis

XL1-Red competent cells were transformed with S303A mutant plasmid and incubated on LBAA plates for 36 hours at 37 °C. 5 mL of LB medium were then inoculated with about 200 colonies and incubated for 24 hours at 37 °C. Under these conditions, mutation frequency was estimated to be 1-6 mutations per 2 kb. Mutated plasmids were extracted and purified with the Nucleospin kit and the size of the extracted plasmids was controlled by 1 % agarose gel electrophoresis after digestion.

To screen mutant enzymes for restored activity, *E. coli* RR1 competent cells were transformed with 1 µL of purified mutant plasmids and incubated on LBAA plates for 36 hours at 25 °C. Each isolated colony was transferred into 200 µL of LB ampicillin medium in a 96 wells uncoated plate (Greiner). In addition, a number of

wells were inoculated with *E. coli* RR1 cells previously transformed with wild-type plasmid so as to serve as a blank. After incubation of the plates for 48 hours at 37 °C, the EPS reaction mixture was directly added to the microcultures and absorbance variation at 405 nm was recorded after 15 minutes. Under these conditions, cells expressing the wild-type α-amylase display activity while those expressing the S303A mutant do not.

2.2.9 Western Blot

After separation of the sample proteins on a biphasic SDS gel and transfer onto a PVDF membrane, the presence of α-amylases was revealed by immunoblotting in 3 steps. Firstly, α-amylases were coupled to specific anti-amylase IgG antibodies from rabbit. These primary antibodies were then recognized by secondary anti-rabbit IgG antibodies coupled to alkaline phosphatase. Finally, the membrane was treated with the alkaline phosphatase substrate to reveal the presence of the complexes.

2.2.10 Production of α-amylases

(i) AHA: *E. coli* RR1 competent cells were transformed with pαH12wt* or with one of the mutant plasmids, plated onto LBAA Petri dishes and incubated for 24 hours at 37 °C. The colonies were then transferred into 1 L erlenmeyer flasks, each containing 200 mL of TYP culture medium and 20 mg of ampicillin, and were incubated at 18 °C. When required, the culture medium was centrifuged at 13,000 g for 50 minutes at 4 °C in order to eliminate the cell debris.

(ii) AmyD: *E. coli* BL21(DE3) competent cells were transformed with plasmid pET-AmyD, plated onto LBAA Petri dishes and incubated for 16 hours at 37 °C. The colonies were then transferred into 1 L erlenmeyer flasks containing each 200 mL of TB culture medium and 20 mg of ampicillin. After ±24 hours incubation at 18 °C (A_{550} between 4 and 5) the production was induced by adding 0.5 mM IPTG. The production was at its maximum after 24 hours induction and the culture medium was centrifuged at 13,000 g for 50 minutes at 4 °C.

2.3 RESULTS

2.3.1 Activity Assays

As mentioned in the materials and methods, we first probed the α-amylase from *P. haloplanktis* for proteolytic and lipolytic activity on immobilized substrates, so as to allow for the detection of trace activity not recorded by usual spectrophotometric methods. The complete absence of any sign of hydrolysis of the substrates for both the wild-type enzyme and the recombinant α-amylase demonstrates that these enzymes are devoid of proteolytic and lipolytic activity, thus ruling out a possible accessory hydrolytic activity in this enzyme, as already suggested by the relatively low solvent accessibility of the triad.

2.3.2 Site-directed mutagenesis

In order to evaluate the function of the non-catalytic triad, the mutations Glu19Gln, Ser303Ala and His337Asn were introduced separately in the AHA gene. The three combinations of double mutations and the triple mutant were also constructed. These mutations eliminate the respective chemical function of the original residues while maintaining a similar steric hindrance, thus they should allow the enzyme to maintain a correct fold and allow us to evaluate the impact of the mutations.

The mutants are readily expressed in *E. coli*, as demonstrated by Western blotting, but they display a strongly reduced stability (Fig. 11). Moreover, they display a very low amylolytic activity (less than 0.5% of the wild-type α-amylase expressed under the same conditions) (Fig. 12), with half-lives of activity in cell-free extracts of one or two days whereas the wild-type α-amylase is stable for weeks under these conditions.

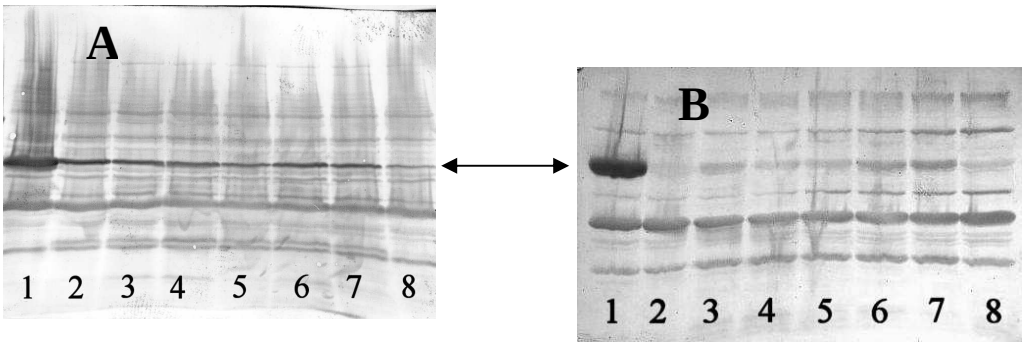


Fig. 11: Immunoblots of wild-type and mutants of AHA after 38 hours of culture in panel A, and after 132 hours of culture in panel B. 1: wild-type; 2: S303A; 3: H337N; 4: E19Q; 5: E/S double mutant; 6: E/H; 7: H/S; 8: E/H/S triple mutant. Double arrow indicates the native forms.

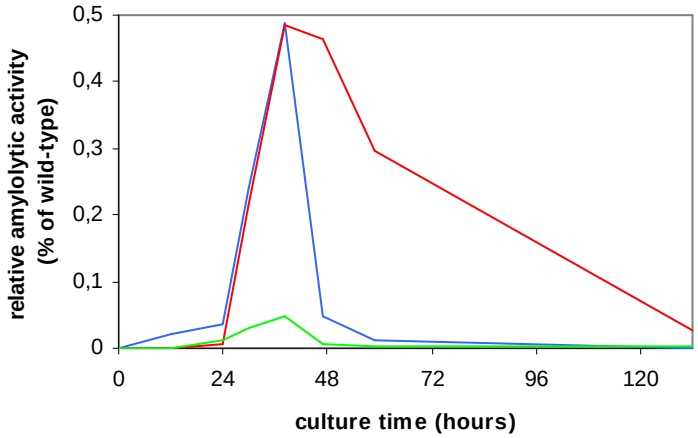


Fig. 12: Evolution of relative amylolytic activity (in % of wild-type activity) measured in the supernatant (red) and in the cell lysate (blue) of the S303A culture, as well as in the cell lysate (green) of the H337N culture.

The least stable and active mutants (activity only detectable on immobilized starch) are those bearing the Glu19Gln mutation, either alone or in combination.

In an attempt to produce a stable mutant containing an altered triad, the following mutations have also been engineered in *P. haloplanktis* α-amylase. His337, the central residue of the triad, has been subjected to saturation mutagenesis and replaced by all other amino acid residues. The reason why we chose to subject only the

His residue to saturation mutagenesis resides in the fact that a stable triad-His mutant would have been helpful for the analysis by NMR (see next chapter). Furthermore, two double mutants, Glu19Cys-His337Cys and Ser303Cys-His337Cys, were designed to introduce a covalent disulfide bridge within the triad. This is the opposite strategy if compared to all the other described mutations in this chapter: instead of destabilizing the non-catalytic triad by the elimination of putative hydrogen bonds, we introduced a covalent interaction to increase the bond energy between the residues of the (altered) triad. Finally, the three basic mutations Glu19Gln, Ser303Ala and His337Asn were introduced separately in Mut5, a strongly stabilized variant of the *P. haloplanktis* α-amylase (D'Amico *et al.*, 2003), and in the closely related insect α-amylase from *Drosophila melanogaster* (AmyD) that is one of the most stable chloride-dependent α-amylases according to microcalorimetric studies (Feller *et al.*, 1999). The mutation of the triad in Mut5 and AmyD should have less impact on the overall stability of these more stable enzymes, thus allowing for the characterization of such a mutant.

Each of the described mutants was expressed in *E. coli* but all were found to display the same weak residual activity and pronounced instability. As a result, attempts to purify these mutants were unsuccessful, even using fast chromatographic techniques. These results demonstrate that any of the tested mutations in the non-catalytic triad leads to a drastic destabilization of the protein structure in these α-amylases, even in the case of more stable enzymes such as AmyD and Mut5. These two cases underline that the low stability of AHA due to its psychrophilic character cannot account for the observed pronounced instability of the triad mutants.

2.3.3 Mutant Ser303Ala characterization

The triad mutant Ser303Ala is slightly more stable as far as the half-life of activity in cell free extracts is concerned. The stability of this mutant cannot be improved in cell free extracts, neither by inhibition of serine-, thio-, metallo-, or acid proteases nor by the use of the usual protein stabilizers such as $MgSO_4$, Na_2SO_4, sucrose, trehalose, glycerol, alanine or polyethylene glycol. Trimethylamine *N*-oxide (TMAO), a protecting osmolyte that has a strong ability to force proteins to fold (Baskakov and Bolen, 1998) is also ineffective in stabilizing the Ser303Ala mutant. By contrast, concentration-independent kinetic parameters such as the K_M value for the

substrate Et-G7-pNP (EPS) (K_M≈150 µM) and the relative substrate specificity for the macromolecular polymer (starch, 100% activity) and shorter oligosaccharides (maltooligosaccharides from corn syrup, 64%; maltopentaose, 69%; Et-G7-pNP, 113%) are unchanged with respect to the wild-type enzyme (Fig. 13 and 14). Maintenance of these kinetic parameters suggests that mutants of the non-catalytic triad are produced in a native state followed by a fast and irreversible unfolding, leading to proteolysis of the unfolded state.

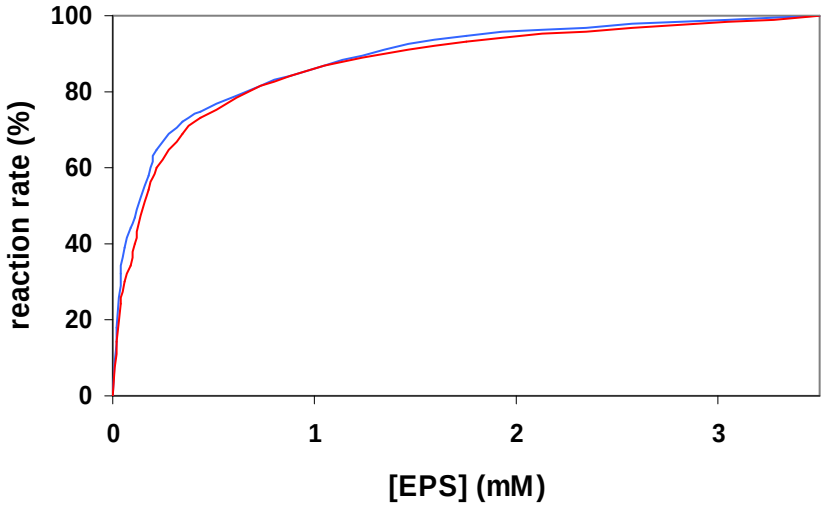


Fig. 13: Determination of the Michaelis-Menten constant for the synthetic substrate Et-G7-pNP at 25 °C: wild-type α-amylase (blue) compared to mutant S303A (red).

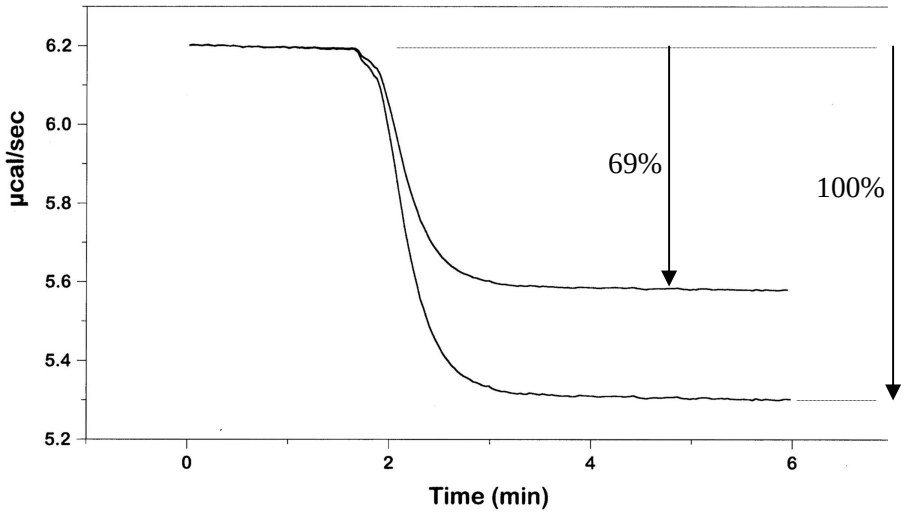


Fig. 14: Example of relative substrate specificity determination by Isothermal Titration Calorimetry for the wild-type amylase. The lower curve corresponds to the macromolecular polymer starch (100% activity) and the upper curve to maltopentaose as substrate.

2.3.4 Random mutagenesis

In an attempt to understand the destabilizing effect of these mutations, the mutant Ser303Ala was subjected to extensive random mutagenesis in the mutator strain *E. coli* XL1-Red with a programmed mutation frequency of one to six bp substitutions per gene. The objective of this experiment was to obtain a secondary mutation restoring the stability to that of the native α-amylase or at least improving the stability of the mutant Ser303Ala. However, none of the 12,500 mutants screened displayed even a slightly improved stability than the Ser303Ala mutant. In control experiments on the wild-type α-amylase gene randomly mutated in the same conditions, it was found that at a mutation frequency of 1 bp substitution per gene, about 5% of the mutants possessed a significantly improved stability. The sequence of some of these mutants also confirmed the programmed mutation frequency. These results indicate that under the described conditions, the mutational perturbations in the non-catalytic triad of the Ser303Ala α-amylase cannot be balanced by any simple mutation or weak interaction in the enzyme structure.

2.4 CONCLUSION

The triad in the tested α-amylases was found, under the conditions used, to be devoid of proteolytic and lipolytic activity, which justifies the "non-catalytic" attribute.

The mutation of any of the three residues of the non-catalytic triad in AHA leads without exception to the production of unstable enzymes with very low amylolytic activity. The most stable mutant S303A can neither be purified nor stabilized by usual protein stabilizers. Furthermore, we could not counterbalance the marked instability by any simple secondary mutation, as shown by the random mutagenesis experiment. Finally, the substrate specificity of this mutant is unchanged with respect to the wild-type enzyme, indicating that the mutant α-amylases are expressed in a native state followed by a fast and irreversible unfolding.

All these mutations, except for those of the saturation mutagenesis experiment, were chosen in order to eliminate the chemical functions of the side chains of the implicated residues while maintaining a similar steric hindrance. In other words, these mutations were designed to destabilize the triad by perturbing the H-bond network. The introduction of a disulfide bridge in the triad represents the opposite approach: it was hoped to stabilize the triad through the introduction of a covalent bond. However, this attempt was unsuccessful as the mutants display invariantly the same low stability and activity.

Knowing that psychrophilic enzymes such as AHA, in their wild-type form, are already characterized by a relatively low stability as compared to their mesophilic and thermophilic homologues, we investigated the effect of the mutations in the triad on more stable enzymes. The same mutations in α-amylases such as the stabilized mutant Mut5 of AHA or in wild-type AmyD, produced exactly the same effects, thus clearly excluding the psychrophilic character from being responsible for our results.

The role of the non-catalytic triad is therefore undeniably of a structural nature. All three residues of this motif are located in domain A that adopts a $(\beta/\alpha)_8$ barrel fold. Glu19 is located on the α1-helix, His on the loop between the β8-sheet and the α8-helix

and Ser is located on the β8-sheet. Thus, we propose that the triad stabilizes domain A by holding its N- and C-terminal ends together.

This is the first time that a triad with only minor differences when compared to the catalytic triad, and thus for believed to be a catalytic motif, is reported to be of structural importance. We therefore propose the non-catalytic triad to be a novel structural motif in the conformational stability of proteins.

At this stage, one may wonder how a combination of three residues could constitute a key determinant of the conformational stability in a large monomeric protein (453 amino acids, 49,340 Da). This α-amylase tolerates various amino acid substitutions: more than twenty mutations have already been introduced during various studies with some being at close proximity to the non-catalytic triad (*e.g.* Asn12Arg close to Glu19, Lys300Arg and Lys300Gln close to Ser303) but none had such a destructive impact on conformational stability as those implicating the residues of the triad. Moreover, the 2 putative H-bonds linking the side chains in the non-catalytic triad are *a priori* unlikely to contribute significantly to the α-amylase conformational stability (400 H-bonds have been detected in the 3D-structure (Aghajari *et al.*, 1998b)) except if at least one of these putative H-bonds is not a conventional interaction but a strong H-bond.

Hydrogen bonds can be classified as a function of the distance separating the heteroatoms implicated in the establishment of the bond and as a function of the pK_a values of the proton donor and acceptor. According to this classification, three types of H-bonds are described (distances and bond energy values do slightly vary in the literature): (i) weak or normal H-bonds display bond lengths of 2.8-3.1 Å with bond energies ranging from 10 to 50 kJ/mol. (ii) Strong or low-barrier hydrogen bonds (LBHB) display bond lengths of 2.5-2.8 Å with energies ranging from 50 to 100 kJ/mol, and (iii) for very strong or single-well H-bonds, bond lengths from 2.3-2.5 Å and energies higher than 100 kJ/mol are reported (Hibbert and Emsley, 1990). The general consensus is that the shorter the bond, the higher its bond energy and the closer the pK_a values of acceptor and donor, leading to an equidistant sharing of the proton in the strongest H-bonds. For proteins and nucleic acids, weak hydrogen bonding is generally expected: thus, strong or low-barrier hydrogen bonds in biological systems are unusual and noteworthy (Hibbert and Emsley, 1990).

The occurrence of a strong hydrogen bond in the non-catalytic triad could account for the instability of mutants of the participating residues. Indeed, the stability of AHA has been characterized by differential scanning calorimetry, yielding a calorimetric enthalpy for unfolding of approximately 800 kJ/mol (Feller *et al.*, 1999). It follows that disrupting an LBHB by mutagenesis can result in a ±10% decrease of this value.

As we will see in the next chapter, strong hydrogen bonds and especially LBHBs have been proposed as participants in the mechanisms of action of many enzymes. In biological systems, they are most readily identified and studied by NMR spectroscopy as this method permits their structural assignment.

3.1 INTRODUCTION

The story of the *catalytic* triad began in 1969 with the X-ray crystal structure of chymotrypsin (Blow *et al.*, 1969), one of the first four protein structures to be solved. Based on this diffraction data, Blow's interpretation of the underlying catalytic mechanism led to the so-called "charge-relay" model, with protons being relayed towards the Asp residue and electrons towards Ser. The net effect would be to make Ser a powerful nucleophile (Fig. 15).

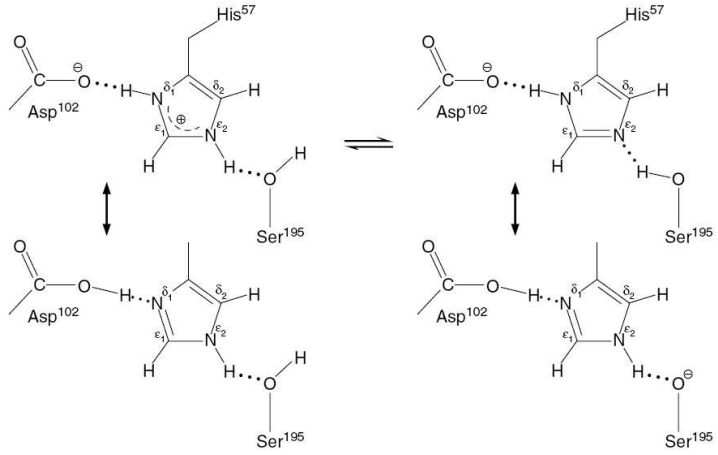


Fig. 15: The charge-relay system, as proposed by Blow *et al.* (1969) for chymotrypsin.

The proposed catalytic charge-relay system provided as many insights as it raised questions (mainly concerning the possibility of alkoxide ion formation, the reactivity of Asp, the position of protons in the hydrogen bonds and the solvent accessibility). It therefore needed additional X-ray diffraction work and spectroscopic studies, capable of examining more directly the properties of the H-bonds, to draw a clearer picture of the underlying mechanism.

Nuclear magnetic resonance has since provided precious information on the hydrogen bond network that is the key to the catalytic mechanism[2]. The first landmark

[2] For extensive review of the contributions of NMR to the study on hydrogen bonds in serine protease active sites, refer to Bachovchin, W. W. (2001): Contributions of NMR spectroscopy to the study of hydrogen bonds in serine protease active sites *Magn. Reson. Chem.,* **39,** S199-S213.

paper on this topic was published by Robillard and Shulman in 1972 (Robillard and Shulman, 1972), who first identified a lone and downfield ^{1}H-NMR signal in the spectra of chymotrypsin and chymotrypsinogen. These signals were assigned to the proton in the Asp-His hydrogen bond of the catalytic triad and its downfield position was explained by the presence of an unusually strong H-bond between the 2 residues. Over the last 35 years, it can be seen that the evolution of the knowledge on the catalytic triad has been directly dependent on the evolution of NMR technology. Thus, the currently proposed model (Fig. 16) is essentially based on observations of this resonance phenomenon. As can be seen in the red box, the NMR signal already observed by Robillard and Shulman (1972) is still part of the model with its presence limited to the imidazolium form of His.

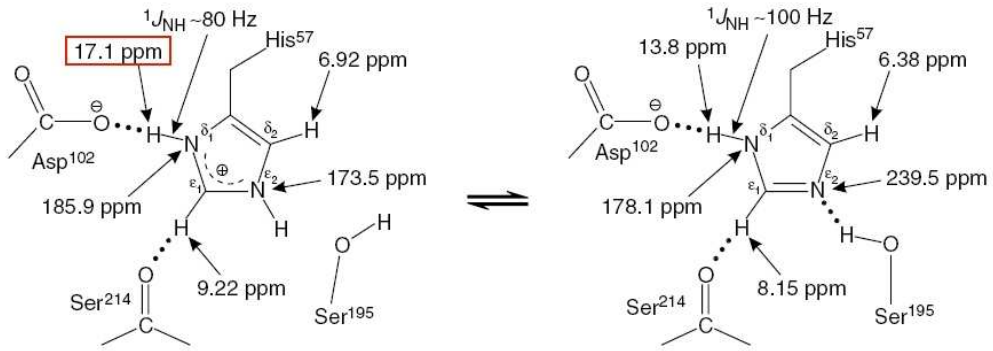


Fig. 16: Current model of the catalytic mechanism of serine proteases, showing a 'catalytic tetrad', chemical shifts and spin coupling constants. The red box indicates the NMR signal already observed by Robillard and Shulman (1972) and seems to be only present for the imidazolium form of His, not for the neutral ring. Source: (Bachovchin, 2001).

Besides this modern biochemistry textbook model, two different and opposing theories have also been proposed.

Firstly, the low-barrier hydrogen bond theory has been published in 1994 by Frey *et al.* (Frey *et al.*, 1994) and actually corresponds to a slightly modified charge-relay system. The authors have proposed that the molecular basis of proteolytic catalysis would reside in the transformation of the Asp-His H-bond to a low-barrier H-bond upon protonation of the imidazole ring in the transition state, essential for the charges to relay. LBHBs are known to be unusually strong with energies reported to range from 40-100 kJ/mol and according to Frey *et al.*, they are most easily identified

by the high chemical shifts of the H-bonded protons, with a minimum threshold of about 15 ppm. Moreover, LBHBs are characterized by an equal or nearly equal sharing of the H-bonded proton due to matching pK_a values of donor and acceptor groups and by short distances between the implicated heteroatoms ($\leq$2.65 Å).

Secondly, Ash *et al*. proposed the imidazole ring flip theory in 2000 (Ash *et al.*, 2000). The flip movements of the imidazole ring consist of a rotation about its Cβ-Cγ axis, NE2-H changing position with CE1-H and ND1-H with CD1-H. Thus, the rotation interchanges exchangeable N-H protons with non-exchangeable C-H protons, with the resulting system being either in a proton transfer position with activated Ser, or in a proton transfer blocking position. Although the authors present results that oppose the presence of a strong H-bond in the catalytic triad, they do not propose any explanation for observing the downfield proton NMR signal.

What is interesting from these two theories for our study is that both report the presence of an unusually strong hydrogen bond in the catalytic triad that gives rise to a characteristic signal in proton NMR spectra, whether or not it is an LBHB. To date, this NMR signature for a strong H-bond has been demonstrated for the catalytic triads of many enzymes other than serine proteases. These include phospholipaseC (Ryan *et al.*, 2001), aspartic proteases (Northrop, 2001), lipolytic thioesterase/protease I (Lin *et al.*, 1998), cholinesterases (Viragh *et al.*, 2000, Massiah *et al.*, 2001) and hydroxynitrile lyases (Stranzl *et al.*, 2004). It is noteworthy that all LBHBs detected thus far in enzymes involve carboxyl groups (Mildvan *et al.*, 1999).

Taking into account our "structural motif" proposal for the non-catalytic triad in α-amylases and the observations made by ^{1}H-NMR spectroscopy for the catalytic motif, we decided to analyze AHA by ^{1}H-NMR to test for the presence of a downshifted peak. A strong hydrogen bond could be the key to the elucidation of the non-catalytic triad's role and could explain how this motif can be an essential feature in the conformational stability of an entire protein.

3.2 MATERIALS AND METHODS

3.2.1 Production and purification of AHA

E. coli RR1 competent cells were transformed with the plasmid pαH12wt*, plated onto LBAA Petri dishes and incubated for 24 hours at 37 °C. The colonies were then transferred into 2 L erlenmeyer flasks containing 200 mL of TYP culture medium. After 5 days of incubation at 18 °C, the culture medium was centrifuged at 13,000 g for 50 minutes in order to eliminate the cell debris. This step and all the subsequent purification steps were carried out at 4 °C in order to protect the psychrophilic enzyme from heat denaturation.

The proteins in the supernatant were precipitated with 80 % ammonium sulfate, centrifuged at 14,500 g for 1 hour and the resulting pellet was resuspended in 50 mL of chromatography buffer (50 mM Tris, 1 mM $CaCl_2$, pH 7.5). After dialysis of the protein suspension for 24 hours against 2x3 L of chromatography buffer, three chromatographic steps were required for the purification of recombinant AHA:

(i) DEAE-agarose ion exchange chromatography (2.5x40 cm): elution of bound protein with a linear salt concentration gradient (0-0.8 M NaCl; 400x400 mL) in the chromatography buffer.

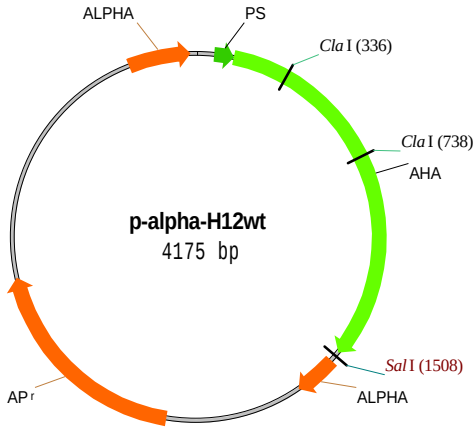
(ii) Sephadex G100 gel filtration chromatography (2.5x95 cm): actually used as a pseudo affinity chromatography as it retains the amylase longer than expected with respect to its molecular size.

(iii) Ultrogel AcA54 matrix (2.5x95 cm): a second gel filtration chromatography step.

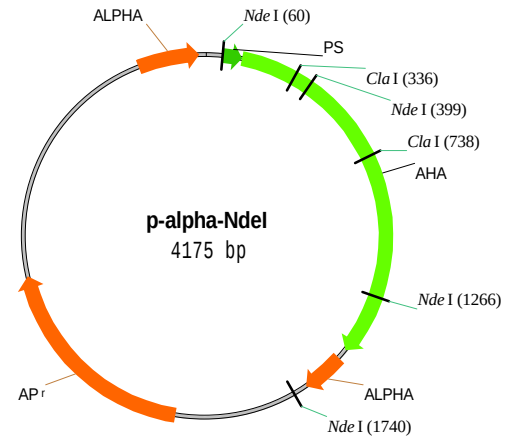
The last two columns were eluted with chromatography buffer and between the different chromatographic steps the sample was concentrated with an Amicon Ultrafiltration cell using a nitrogen pressure of 3 bars and a Millipore polyethersulfone membrane with a cutoff of 10,000 Da. The final fractions containing the α-amylase were pooled, concentrated and stored at -70 °C.

3.2.2 Construction of pET-AHA

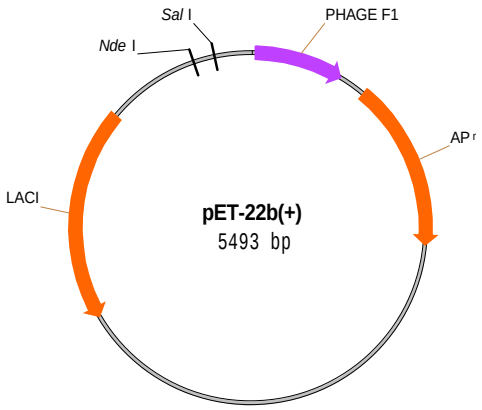
In order to produce AHA in M9 medium for ^{15}N-labeling, the AHA gene was transferred into a pET-22b(+) vector according to the following strategy:



Restriction *ClaI/SalI* :
2 fragments
ClaI(336)-*SalI*(1508) : 1172 bp
SalI(1508)-*ClaI*(336) : 3003 bp

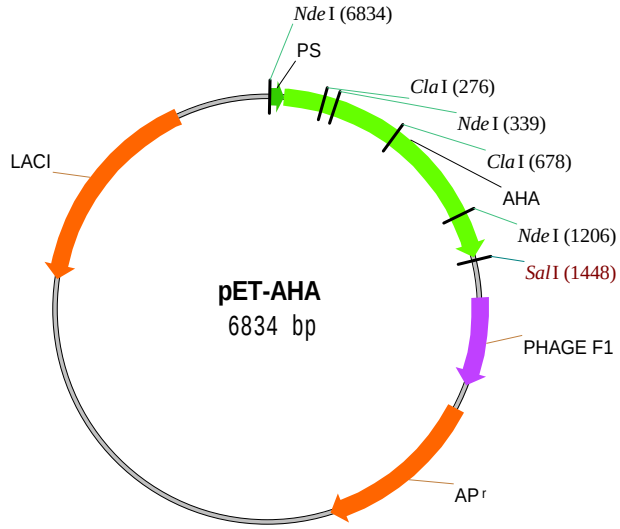


This plasmid is identical to pαH12wt*, but with an engineered *NdeI* restriction site at the ATG codon of the signal peptide.
Restriction *NdeI/ClaI* :
5 fragments
NdeI(60)-*ClaI*(336) : 276 bp
ClaI(336)-*NdeI*(399) : 63 bp
NdeI(399)- *NdeI*(1266) : 867 bp
NdeI(1266)- *NdeI*(1740) : 474 bp
NdeI(1740)- *NdeI*(60) : 2495 bp



Restriction *NdeI/SalI* :
2 fragments
NdeI(289)-*SalI*(177) : 5381 bp
SalI(177)-*NdeI*(289) : 112 bp

The fragments constituting the restriction mixtures were separated by 1 % agarose gel electrophoresis. The bands corresponding to the highlighted fragments were recovered from the gel, purified and ligated, giving rise to the production plasmid pET-AHA bearing the amylase gene and its original signal peptide:



3.2.3 Production and purification of ^{15}N-labeled AHA

^{15}N-M9 culture medium, with $^{15}NH_4Cl$ as the only nitrogen source, was used for the production of labeled AHA and was prepared as follows: $Na_2HPO_4.2H_2O$ 6 g/L; KH_2PO_4 3 g/L; NaCl 0.5 g/L (pH 7) were autoclaved and cooled to 50 °C. $^{15}NH_4Cl$ 1.1 g/L; $MgSO_4.7H_2O$ 2 mM; $CaCl_2$ 0.1 mM; glucose 0.4 % and ampicillin 100 mg/L were filtered (0.22 µm) and added to the autoclaved mixture. For the preparation of ^{15}N-labeled M9 Petri dishes, 16 g/L of agar were added before autoclaving.

E. coli BL21(DE3) competent cells were transformed with the plasmid pET-AHA, plated onto ^{15}N-M9 Petri dishes and incubated for 16 hours at 37 °C. The colonies were then transferred into 2 L erlenmeyer flasks containing 200 mL of ^{15}N-M9 culture medium. After ±10 hours of incubation at 18 °C, when the absorbance at 550 nm reached 0.5±0.1 units, the production was induced by adding 0.5 mM IPTG. The production was at its maximum after 24 hours induction and the culture medium was centrifuged at 13,000 g for 50 minutes at 4 °C. The proteins of the supernatant

were then precipitated with 95 % of ammonium sulfate, centrifuged at 14,500 g for 1 hour and the resulting pellet was resuspended in 50 mL of chromatography buffer.

Subsequent dialysis and purification were as described for unlabeled AHA.

3.2.4 Mass spectroscopy

Purified protein samples (^{15}N-labeled and unlabeled AHA) were prepared in ammonium acetate 100 mM, pH 6.6 and the proteins concentrated to 0.1 mM using a Millipore Ultrafree membrane with a cutoff of 5 kDa. 18 µL of acetonitrile/H_2O (1:1) were then added to 1 µL of each protein sample.

Analysis was performed in a quadrupole-time of flight system (Q-TOF Ultima) after ionization by electrospray. The maximal error with this system is 4 Da.

The percentage of isotopic labeling could thus be calculated with the following formula:

$$\% = \frac{\left(Mw(^{15}N - AHA) - Mw(^{14}N - AHA)\right) * 100}{number\ of\ N\ atoms}$$

3.2.5 NMR

Purified AHA was prepared in buffer MES 200 µM, $CaCl_2$ 1µM, pH 5.8, and concentrated to 0.3 mM using the same type of membrane as described above. 30 µL of D_2O were added to 430 µL of sample just before the experiment. Data was recorded at 20 °C on a Varian 800 MHz Inova NMR spectrometer at the Institut de Biologie Structurale Jean-Pierre Ebel in Grenoble, France, in collaboration with Dr. Jean-Pierre Simorre.

3.3 RESULTS

AHA was produced in *E. coli*, purified to homogeneity and concentrated to 0.3 mM. To promote His protonation, the pH was set to 5.8, a value close to the solubility limit of the protein (pH ≈ 5.4). The ^{1}H-NMR spectrum of AHA is shown in Fig. 17A.

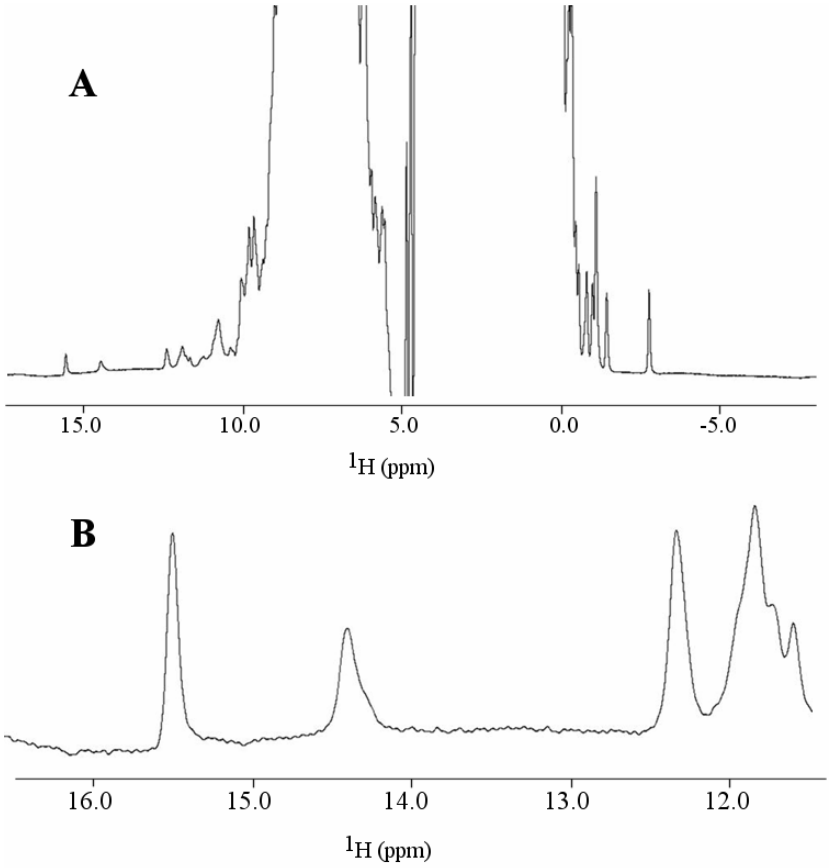


Fig. 17: 800 MHz proton NMR spectra of AHA in $^{1}H_2O$. A, overview of the proton signals for AHA. B, enlargement of the downfield spectrum in A showing the two proton signals at 14.4 and 15.5 ppm.

From the mass of overlaying signals of the protein protons, we observe a lone peak at 15.5 ppm (Fig. 17B) with a very low field resonance, similar to that attributed to the protonated His residue in catalytic triads, but very unusual for such a large structure

(50 kDa). Moreover, this peak is accompanied by a second peak at 14.4 ppm, exactly as in the ^{1}H-NMR spectrum of the thioesterase/protease I (TEP-I) of *E. coli* (Tyukhtenko *et al.*, 2002) where this peak has unambiguously been attributed to the hydroxyl hydrogen atom of the Ser of the catalytic triad. Earlier publications also mentioned the presence of the second, less downfield shifted signal in the spectra of other serine proteases (Robillard and Shulman, 1972, Steitz and Shulman, 1982).

The peak at 15.5 ppm disappears when the D_2O/H_2O ratio is raised, showing that the proton responsible for this peak is exchangeable. However, the exchange rate is much slower than that of normally H-bonded protons, indicating that either the energy of this hydrogen bond is higher as compared to a normal H-bond, or that the responsible proton displays low solvent accessibility.

Under certain conditions, the close proximity of an aromatic ring to a proton or even a proton of the aromatic ring itself could also account for the observed downfield shift: the important cyclic currents created by the delocalized electrons strongly influence the environment of the proton and thus, its NMR signal. Therefore, we first have checked this possibility using the software "shiftX"[3] (Neal *et al.*, 2003) which allows the calculation of chemical shifts based solely on the atomic coordinates of a given molecule. The protons are positioned at a standard distance from the atom they are bound to and whenever such a proton is close to an aromatic ring, its predicted chemical shift will account for it if required. After running the algorithm on AHA (PDB code: 1AQH), we checked every hydrogen atom of the structure but could not find any with a chemical shift predicted to be higher than 10 ppm. This should exclude the influence of cyclic currents on a proton as the origin of the observed signal in the α-amylase spectrum, but does not provide any proof that the proton belongs to an imidazole side chain.

Consequently, the next step was the analysis of the ^{1}H-^{15}N correlation spectrum of ^{15}N-labeled AHA. To reach this goal, the AHA gene has been inserted into the expression vector pET-22b(+), following the engineering of an *NdeI* restriction site at the start codon. The labeled enzyme was produced in M9 minimal medium containing $^{15}NH_4Cl$ as the sole nitrogen source. The production yield was approximately 3.5 mg/L.

[3] http://redpoll.pharmacy.ualberta.ca/shiftx/
Software developed by David Wishart.

About 50 mg of ^{15}N-labeled AHA have been purified to homogeneity and concentrated to 0.3 mM. Mass spectroscopy showed a mass difference of 587±4 Da between the labeled and unlabeled enzyme; thus, the isotopic labeling was calculated to be 97.7±0.7 %.

Figure 18 shows an extract of the ^{15}N-^{1}H correlation spectrum of ^{15}N-labeled AHA corresponding to part of the bulk protein. Even though the quality of the spectrum is excellent for a 50 kDa structure, the elevated number of overlaying signals precludes detailed analysis of correlations despite the high power of the NMR appliance used.

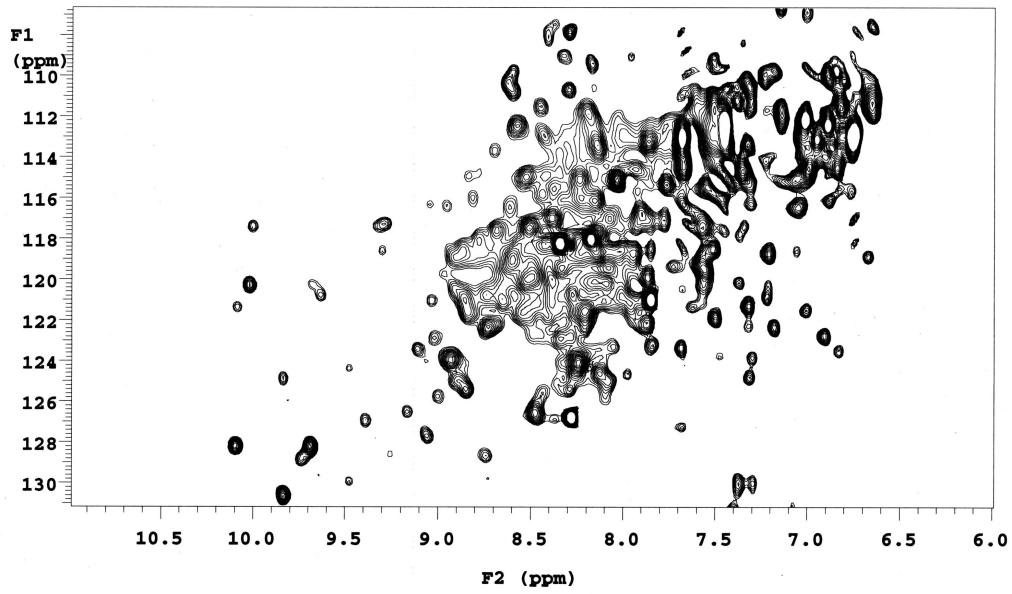


Fig. 18: 800 MHz ^{1}H-^{15}N HSQC correlation spectrum of the ^{15}N-labeled AHA. ^{1}H signal is represented on X-axis; ^{15}N signal on Y-axis. Spectrum recorded with the carrier centered on the amide proton at 120 ppm.

Nevertheless, in the low field of the ^{1}H shifts (Fig. 19), the observation of a spot at the position ^{1}H: 15.5 ppm; ^{15}N: 182 ppm demonstrates that the proton responsible for the signal at 15.5 ppm is correlated with a nitrogen atom.

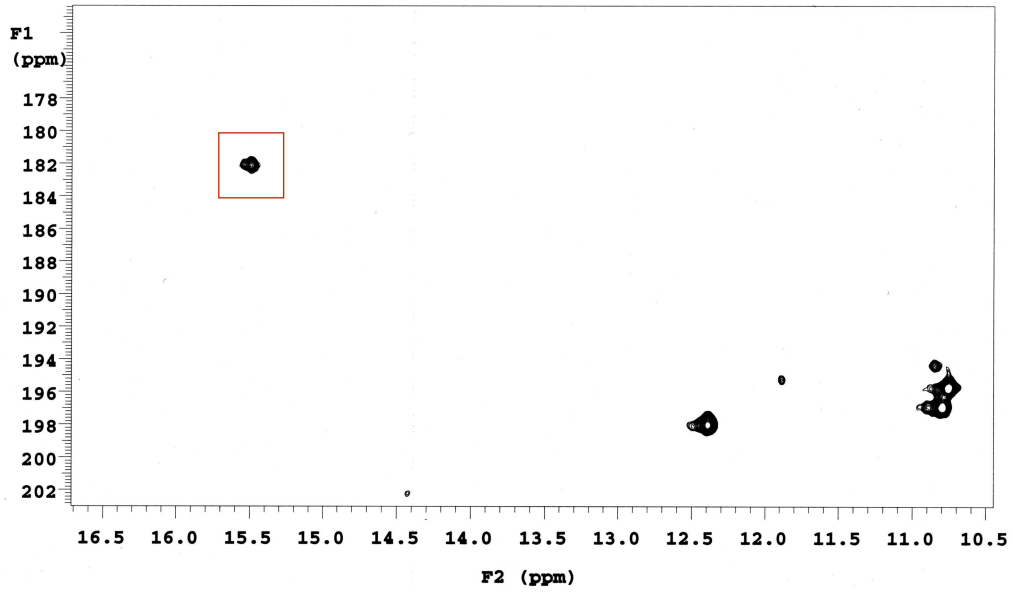


Fig. 19: 800 MHz ^{1}H-^{15}N HSQC correlation spectrum of the chemical shift range of interest. ^{1}H signal is represented on X-axis; ^{15}N signal on Y-axis. Spectrum recorded with the carrier centered at 190 ppm on the His ND1 chemical shift.

The Biological Magnetic Resonance Bank[4] (BMRB) (Seavey *et al.*, 1991) provides a database with the majority of the NMR chemical shifts observed for all atoms of all 20 amino acids that are encountered in the living world. We searched the BMRB for the chemical shifts of the hydrogen atom and of the nitrogen atom responsible for the observed spot and came to the conclusion that the only possible N-H configuration able to explain both chemical shifts is a protonated histidine imidazole ring.

In addition, Yufeng and McDermott (Yufeng and McDermott, 1999) have analyzed the ^{1}H chemical shifts as a function of the distance between the H-bonded heteroatoms implicating a protonated His side chain and carboxylic groups (Fig. 20).

[4] http://www.bmrb.wisc.edu/

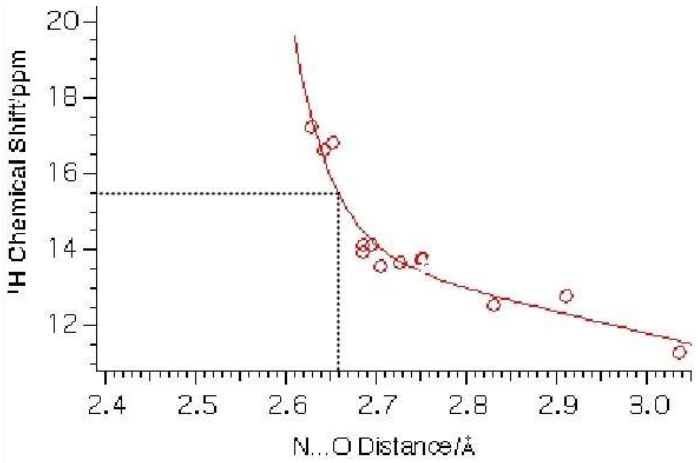


Fig. 20: ^{1}H-NMR chemical shift versus N…O distance of H-bonded heteroatoms established by the nitrogen atoms of imidazolium and the oxygen atoms of carboxylic groups. Dashed lines refer to the 15.5 ppm signal observed for AHA. Adapted from Yufeng and McDermott, 1999.

A chemical shift of 15.5 ppm corresponds to a total H-bond length of 2.66 Å, which is in perfect agreement with the 2.71±0.07 Å measured for the 5 α-amylase structures between the E19 OE2 atom and the H337 NE2 atom. As the strength of such a hydrogen bond increases with decreasing heteroatom distance, the non-catalytic triad seems to bear a remarkably strong H-bond between its Glu and His residues. According to Yufeng and McDermott (1999) and in addition to the fact that the His side chain must be in its imidazolium form and Glu must be deprotonated to account for all our observations, the difference between acceptor and donor pK_a values must not exceed 5 units. The protonated state of the histidine side chain, the short distance between the implicated heteroatoms and the weak pK_a difference are facts that have already been discussed in the charge-relay model and the LBHB theory.

Despite these interesting observations, the NMR signals could not be attributed unambiguously to the His residue composing the non-catalytic triad for several reasons: (i) AHA bears 12 His residues, (ii) none of the triad His mutants of AHA, AHA Mut5 and AmyD was stable enough to be purified, rendering impossible an analysis of these mutants by NMR, and (iii) attempts to use pulsed two-dimensional NMR techniques for the identification of the His residue failed because of the size of the protein, *i.e.* the high number of overlaying NMR signals.

3.4 CONCLUSION

We report the presence of an isolated, downfield shifted peak in the ^{1}H-NMR spectrum of AHA. The responsible proton was confirmed by ^{1}H-^{15}N NMR analysis of ^{15}N-labeled AHA to be bound to an imidazole nitrogen atom.

These observations are very similar to those reported for the spectra of serine proteases and other enzymes whose catalytic mechanism relies on a triad. In all these cases, the origin of the NMR signal was attributed to a strong hydrogen bond between the side chains of Asp/Glu and His of this catalytic motif.

The presence of such a strong H-bond in the non-catalytic triad could account for the results reported in the first section of this document and thus, would be a convincing explanation as to how the triad could dominate the conformational stability of α-amylases. Further evidence in favor of our structural hypothesis was provided by the group of McDermott and coworkers, whose observations allowed us to underline the presence of a strong hydrogen bond in α-amylases. The question whether the strong H-bond is of the LBHB type or not is difficult to answer, and finally not very important. Facts in favor of an LBHB are however the distance between the N and O heteroatoms matching with the distances reported for LBHBs, as well as the significant NMR shift. pK_a values of the triad's His and Glu residues as well as the exact position of the hydrogen between the heteroatoms are unknown.

The unambiguous attribution of the N-H signal to the His residue of the non-catalytic triad has failed for the above mentioned reasons. This probably will remain problematic, since the only technique suitable for tracking hydrogen bonds is NMR and as we used one of the most powerful NMR spectrometers currently available.

However, some additional arguments are in favor of the assignment of the NMR signal to the triad His residue. All LBHBs detected thus far on enzymes involve carboxyl groups (Mildvan *et al.*, 1999): among the 12 His residues of AHA and besides the triad His, only 3 (His117, 321 and 442) are at H-bonding distance and coplanar with an oxygen atom of a carboxylate function, but these are non-conserved His in the chloride-dependent α-amylase family and are devoid of any known function. None of these 3 His-carboxylate combinations has a Ser residue nearby. In the same context,

both strictly conserved histidines indirectly involved in the catalytic mechanism (His89 and 263) have no carboxylate at H-bonding distance, suggesting that the NMR signal at 15.5 ppm is not related to the active site.

Assuming that the His residue of the non-catalytic triad is indeed the origin of the NMR signal and thus that this triad bears a strong hydrogen bond, the sum of observations with respect to their underlying conditions guided us to the following model (Fig. 21).

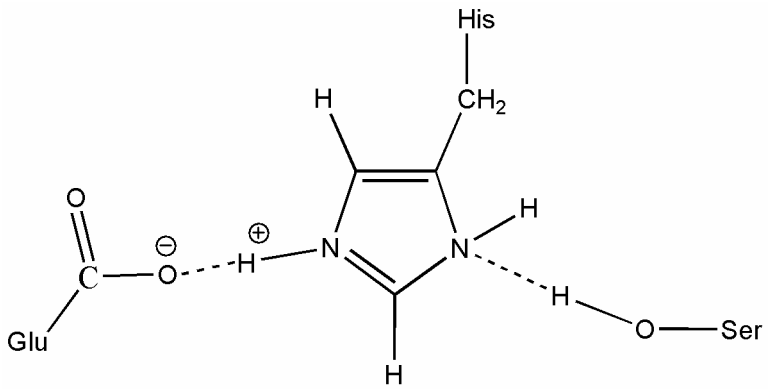


Fig. 21: Proposed model of the H-bond network for the non-catalytic triad of α-amylases.

This model accounts for the protonated state of the His imidazole ring (on the NE2 atom) and thus, it explains the NMR peak at 15.5 ppm due to an LBHB between His and Glu. On the other side of the ring, the hydrogen atom needs to be on the hydroxyl oxygen atom of Ser to account for the peak at 14.4 ppm, therefore excluding the presence of an alkoxide ion which would in any case be far too reactive to exist in this apparently inactive triad. The reason why we proposed this proton to be interacting with the His ND1 atom resides in the fact that in the correlation spectrum (Fig. 19) we observe a very weak N-H correlation signal at position ^{1}H: 14.4 ppm; ^{15}N: 202 ppm. This signal indicates that both atoms must be close and interacting.

The question as to why this non-catalytic triad is devoid of any lipolytic or proteolytic activity remains interesting. If we assume that the moderate solvent accessibility of the triad is not the (only) reason for this motif being inactive, our model could explain this lack of activity, independently of whether a catalytic triad operates via a charge-relay system or a flip movement of the His side chain. Indeed, if the proton

on the acid side is on the carboxyl group, we unavoidably encounter a conflicting situation (Fig. 22), *i.e.* a very reactive alkoxide ion would have to remain inert while being close to the $-NH_2^+$ group of the imidazole ring. Also a flip movement of the His imidazole ring would not provide any solution for this improbable and conflicting situation.

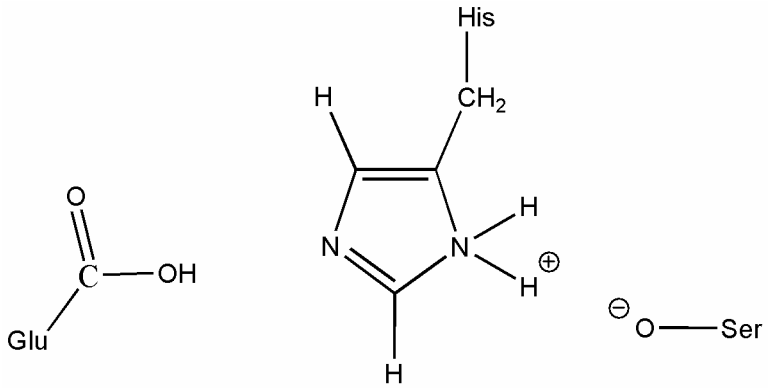


Fig. 22: Conflicting model for the non-catalytic triad of α-amylases, illustrating why no charges can be relayed.

As this scenario is very unlikely, we propose that the triad in α-amylases cannot be active as no charges can be relayed. Based on our model, it therefore appears that the precise positioning of the Ser residue is the key element which decides if a triad is active or not (Fig. 23).

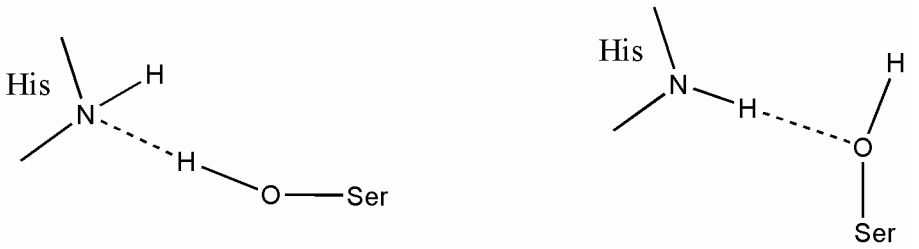


Fig. 23: Possible explanation for a triad being inactive. Left side: model as proposed for the non-catalytic triad in α-amylases; right side: model as proposed in the charge-relay system for the catalytic triad.

Nevertheless, we must not forget that the imidazole ring is flipped in the non-catalytic triad of α-amylases with respect to the catalytic triads, with the ND1 atom facing Ser OG instead of the carboxylate group. Although the chemical significance of

this strict conservation of the imidazole geometry in all the catalytic triads has not been explained to date, the inverse geometry in α-amylases is the main obvious difference that we have to accept as another possibility for the lack of Ser reactivity, although this is not easily explicable from the model as the interchange of ND1 and NE2 atoms (*i.e.* the non-catalytic protease-like orientation, see next chapter) leaves the model unaffected (Fig. 24).

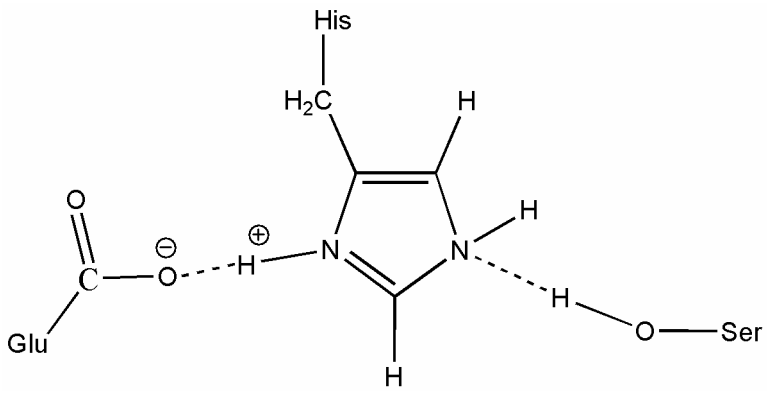


Fig. 24: Triad in the non-catalytic protease-like orientation.

4.1 INTRODUCTION

In order to identify similar triads in protein structures other than α-amylases, we scanned the entire RCSB Protein Data Bank[5] (Berman *et al.*, 2000) (almost 30,000 entries in January 2005) with the algorithm Jess[6] (Barker and Thornton, 2003) (a modified version of the TESS algorithm (Wallace *et al.*, 1997)) provided by Jonathan Barker from the European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK.

The screening of databases to analyze or identify catalytic Ser-His-Asp triads has already been reported several times. In 1993, Barth *et al.* (Barth *et al.*, 1993) have performed an extensive steric comparison based on rmsd fits of the active site residues of all the serine proteases contained in the PDB. As a result, they were able to classify these enzymes according to the (chymo)trypsin and subtilisin families, and to identify an additional serine which is highly conserved in the active site of the serine proteases and which they suggested to be part of a "catalytic tetrad". They also found several examples of Ser-His-Asp triads in non-proteolytic proteins (Barth *et al.*, 1994). Still in 1994, Artymiuk *et al.* (Artymiuk *et al.*, 1994) have used a graph-theoretic approach for the identification of the 3D patterns of amino acid side chains in protein structures. Besides the identification of several catalytic triads from their data set, they also identified an unusual triad from the pro-enzymes chymotrypsinogen and trypsinogen which does not exist in the active form. A different structural comparison of the serine proteases, using a less specific technique, has been performed by Fischer *et al.* (1994) (Fischer *et al.*, 1994). Their method, derived from geometric hashing methods used in computer vision research, treats all Cα atoms in a protein as points in space and compares proteins purely on the geometric relationships between these points. It thus can detect recurring substructural 3D motifs, and was able to identify the structural similarities of the active sites of the trypsin-like and subtilisin-like serine proteases based solely on the similarities of the Cα geometries of their constituent residues. Finally, Wallace *et al.* (1996) (Wallace *et al.*, 1996) have described how to derive a 3D

[5] http://www.rcsb.org/pdb/

[6] For information: Jess is not an acronym; thus, it has not to be written in capital letters throughout. It is named in honor of its predecessor TESS, the T replaced by the first initial of the author's name, Jonathan Barker. The author also proposes "Jonathan's Excellent Shite Shuffler" for those who absolutely want the name to be an acronym.

coordinate template to search structural databases and have illustrated their method by applying it to the catalytic triad. The algorithms used in all these studies were different from Jess (not available at the time) and the target database contained approximately 600 selected structures out of the PDB.

Furthermore, none of these studies mentions a triad in α-amylases for several reasons: (i) some of the 3D structures had not been elucidated at that time, (ii) the algorithms were not suitable or sophisticated enough, (iii) the target database was too restricted, or most importantly, (iv) the triad in its non-catalytic, α-amylase-like orientation would in any case have been missed with such scans, as all the templates used in these searches were based on the catalytic triad and none of the authors was trying to identify a non-catalytic triad. Therefore, our project to scan the entire PDB with a powerful constraint-based algorithm and a precise 3D template derived from the non-catalytic triad was pioneering work and important for underlining our "structural motif" hypothesis.

4.2 MATERIALS AND METHODS

4.2.1 Algorithm

Jess is an algorithm capable of processing a template consisting of arbitrary constraints on physical (type of atom), sequential (atom attributed to a corresponding amino acid) and geometric (atomic coordinates) properties. We can regard the template as a query and probe a database of structures to discover matching sites. As Jess uses the distances between pairs of atoms in the template to match its geometry, a tolerance factor (the delta parameter or Δ) is introduced and set by the user: in a matching structure, the distances between pairs of atoms are within Δ Å of the corresponding distance in the template.

4.2.2 Templates

Since the work of Wallace *et al.* (1996), we know that the relative positions of the two functional oxygen atoms Asp OD2 and Ser OG govern the conformation of the catalytic triad. An additional proof of the importance of the position of these two atoms

is shown in Fig. 8A, the superimposition of the catalytic and non-catalytic triads: only the positions of these atoms and of the imidazole nitrogen atoms are conserved and determine the establishment of hydrogen bonds, independent of the orientation of any other implicated residue.

Therefore, besides the fact that our templates were based on the triad of the *Pseudoalteromonas haloplanktis* α-amylase (PDB identifier: 1AQH), our first template consisted of the coordinates of ND1 and NE2 of the His337 imidazole ring and the oxygen atoms OG of Ser303 and OE2 of Glu19, the latter facing NE2. (Δ = 0.2 Å).

```
ATOM    101  OE2 GLU   19      17.704  20.729  26.257
ATOM      1  OG  SER  303      13.848  22.894  21.000
ATOM      1  ND1 HIS  337      14.748  20.544  22.412
ATOM      1  NE2 HIS  337      16.297  20.388  23.885
```

Fig. 25: Template 1.

As can be seen in the second column of template 1, a numeric code is attributed to each atom, equivalent to the mentioned sequential constraint: “101” imposes a match to any non-carbon side chain or ligand atom without residue matching (Glu could be replaced by Asp for example), whereas “1” imposes a match to any non-carbon side chain atom within the specified residue (which restricts (i) Ser OG to a strict conservation of element, name (the “G” in OG) and residue, and (ii) His nitrogen atoms to be interchangeable, allowing them to match with triads in either protease- or amylase-like orientation and to account for the random assignment of these atoms in protein structures). The fact that we also targeted protease-like triads can be considered as a control for the validity of our results.

For the second template, the coordinates of the CE1 atom of His337 were added with the numeric code “0” imposing an exact match on atom name and residue name. The oxygen atom of Ser and the imidazole nitrogen atoms were imposed to match any atom of the same element within the residue specified (numeric code “3”), *i.e.* for Ser, it could also be a main chain oxygen, and for His, a strict conservation (main chain atoms would be allowed, but in main chains they never occur in the N-C-N configuration as determined by the atomic coordinates). Thus, only the amylase-oriented triads should be retained.

```
ATOM    101  OE2 GLU   19      17.704  20.729  26.257
ATOM      3  OG  SER  303      13.848  22.894  21.000
ATOM      0  CE1 HIS  337      15.367  21.185  23.387
ATOM      3  ND1 HIS  337      14.748  20.544  22.412
ATOM      3  NE2 HIS  337      16.297  20.388  23.885
```

Fig. 26: Template 2.

4.2.3 Data treatment

Jess records matching structures, called hits, with the following attributes: (i) the rmsd value of the hit superimposed on the template, (ii) the delta value corresponding to the maximum distance difference (pair wise atom distances compared with the template), (iii) the PDB accession code, and most importantly (iv) the serial number of each atom as it appears in the PDB text file of the matching structure, which then leads to the identification of the corresponding amino acid residue.

To this list of attributes we then added the name of each protein, its enzyme classification or EC number and its form (*e.g.* free, complexed, mutant, fragment, etc.). Secondly, for each hit we consulted the Catalytic Site Atlas (CSA) (Porter *et al.*, 2004), which provides catalytic residue annotation for enzymes in the PDB[7], to determine whether the identified amino acid residues of the matching structure are catalytic or not. Whenever a structure was lacking in the CSA database, we consulted the primary literature. Finally, using SwissPdb-Viewer[8], we superimposed the complete side chains of identified triads by fitting them to the side chains of His residues to visually control the validity of the hit and the orientation of the residues (amylase or protease-like). This visual control also allows the verification of the coplanarity of both oxygen atoms with the imidazole ring, a very important feature for the correct establishment of hydrogen bonds.

[7] http://www.ebi.ac.uk/thornton-srv/databases/CSA

[8] http://www.expasy.org/spdbv
This software is developed and regularly upgraded by GlaxoSmithKline, the version used here was 3.7(SP5).

4.3 RESULTS

4.3.1 Statistics

The first template, with a Δ tolerance of 0.2 Å, led to the identification of 317 matching structures, whereas the second template (Δ = 0.5 Å) provided 272 hits. As expected, this second number is smaller due to the additional carbon atom in the template introducing supplementary constraints, and includes triads already matching with template 1. Thus, if the proteins found twice are taken into account, a total of 450 structures were identified with both templates *i.e.* 450 different PDB codes bearing a triad in either orientation. The finding of proteases, esterases, lipases and other enzymes with a similar catalytic triad, as well as of all 5 chloride-dependent α-amylases underlines the validity of our method.

However, this number does not perfectly match with the real number of triads in the PDB, as distorted triads would have escaped identification by our search even though they are present and/or conserved. A simple means of identifying such distorted motifs was to significantly increase the delta tolerance, but as a result, the number of matching structures was also much higher and the hits included triads that were not superimposable with the non-catalytic triad when the complete side chains were taken into account. We therefore retained the value of 0.2 Å for the delta parameter as being the best compromise for the search with the first template.

The results obtained with template 1 were satisfying and would have been sufficient as far as our study was concerned. As already mentioned, the introduction of the His CE1 atom into template 2 (while maintaining Δ at 0,2 Å!) was only expected to allow for the easy separation of the protease-like from the amylase-like triads; however, a few structures expected to bear a triad were lacking in our results. Therefore, we finally increased the delta tolerance to 0.5 Å for the search with the second template, which also explains why template 2 did not exclusively lead to the identification of all the amylase-like triads already found with template 1.

After grouping the hits (450) obtained with both templates, all the enzymes with a catalytic triad, all the α-amylases as well as the redundant structures, *i.e.* structures with different PDB codes but of the same enzyme (only the structure with the

best rmsd value was retained in this case) were removed from the list, thereby allowing us to focus on the two remaining categories: the non-catalytic amylase-like triads and the non-catalytic protease-like triads. The list of the first category of triads counts 41 non-redundant structures, representing 110 PDB codes out of the 450. Among these 41 structures, 20 triads are composed of Ser-His and either Glu or Asp whereas the other 21 are formed with different residues or ligands as the third member. In the category of non-catalytic protease-like triads, 14 non-redundant structures representing 32 different PDB codes have been identified. 5 of the 14 structures are formed with Glu or Asp and the rest are formed with other residues; no ligands have been identified in this category.

4.3.2 Non-catalytic amylase-like triads

To belong to this category, the structures have to bear a triad composed of a His residue in its flipped position, the Ser OG atom therefore faces the His ND1 atom and any non-carbon side chain or ligand atom may face the His NE2 atom.

For the sake of clarity, we subdivided the 41 structures of this first category into the 2 mentioned types: (i) triads composed of His, Ser and Glu/Asp (Tab. 2), and (ii) triads with other residues or ligands instead of the acid (Tab. 3).

Tab. 2: His-Ser-Glu/Asp non-catalytic amylase-like triads. For each hit, the PDB accession code, the rmsd and delta values (Å), the structure's full name, the number of redundancies (TOT.) and the identified residues are provided; numbering of the 3 latter columns as in corresponding PDB file. The double line separates triads with Glu from triads with Asp.

PDB	rmsd (Å)	Δ (Å)	ENZYME NAME	TOT.	SER	HIS	ACID
1R1S	0.08	0.05	SH2 domain	30	S97	H106	E71
1VH4	0.13	0.14	protein binding protein (stabilizer)	1	S243	H215	E187
2HMY	0.16	0.35	methyltransferase	1	S294	H292	E47
1ODO	0.16	0.18	cytochrome P450	1	S318	H291	E20
1NKG	0.18	0.27	rhamnogalacturonan lyase	1	S367	H365	E136
1VFR	0.19	0.12	NADPH oxidoreductase	1	S81	H82	E54
1N2K	0.19	0.38	sulfatase	2	S250	H227	E253
1CF2	0.22	0.18	glyceraldehyde-3-P dehydrogenase	1	S302	H299	E301
1JY1	0.24	0.2	tyrosyl-DNA phosphodiesterase	2	S419	H372	E415
1IZN	0.31	0.2	actin capping protein	1	S171	H152	E221
1QHM	0.13	0.17	pyruvate formate lyase	6	S1083	H1084	D1074
1BGL	0.15	0.28	β-galactosidase	2	S971	H972	D591
1KW0	0.17	0.31	human Phe-hydroxylase	2	S273	H271	D151
1H46	0.18	0.28	exoglucanase	1	S43	H42	D74
1UKP	0.18	0.1	β-amylase	3	S464	H335	D275
1ASG	0.18	0.14	aspartate aminotransferase	7	S141	H145	D223
1DE6	0.2	0.2	rhamnose isomerase	2	S177	H229	D231
1SI0	0.21	0.15	iron binding protein	2	S278	H267	D205
1MRJ	0.23	0.18	ribosome inactivating protein	2	S61	H51	D1
1DEL	0.28	0.17	deoxynucleoside-P kinase	1	S7	H194	D218

These triads have been analyzed in detail to check if their existence is known and if they could serve as a structural motif, according to our working hypothesis.

Structures bearing an amylase-like Ser-His-*Glu* triad

The primary literature mentions neither a triad nor any of the identified residues for the structures **2HMY** (O'Gara *et al.*, 1999), **1ODO** (Podust *et al.*, 2004), **1NGK** (McDonough *et al.*, 2004) and **1VFR** (Koike *et al.*, 1996, Koike *et al.*, 1998). No literature is available for structure **1VH4**.

- **1R1S** (Cho *et al.*, 2004) **SH2 domains**

This structure is the best representative of 30 (!) enzymes bearing a Sarcoma virus (Src) homology 2 domain, or SH2 domain. It is important to note that the 30 identified structures correspond to 9 different and independent enzymes. The perfect matching with the template, as shown by the low rmsd and Δ values, is worth mentioning. SH2 domains are non-catalytic regions of proteins shown to play an important role in cellular signaling pathways. They have been found in many cytoplasmic signaling proteins and have been shown to bind to phosphorylated growth factor receptors and other proteins phosphorylated in response to extracellular signals. These domains are approximately 100 amino acid residues in length and are often found in conjunction with a 50 amino acid domain known as a Src homology 3 (SH3) domain.

The sequence position of His106 is variable but not its position in the 3D structure: within SH2 domains, this highly conserved residue is always located on the βD4-strand and is implicated in binding of phosphotyrosine. It is known to form hydrogen bonds with E71 and other residues. However, the primary literature for all 30 identified structures mentions neither the Ser residue, nor the word “triad”.

- **1N2K** (Chruszcz *et al.*, 2003) **Sulfatase**

H227 is close to the catalytic residue H229. However, the corresponding literature mentions neither a triad nor any of the implicated residues.

- **1CF2** (Charron *et al.*, 2000) **GAPDH**

This D-glyceraldehyde-3-phosphate dehydrogenase (GAPDH) originates from the hyperthermophilic Archaeon *Methanothermus fervidus*. The residue E301 is strictly conserved within the whole family of GAPDHs and contributes to the protein stability through hydrogen bonding. The reason why Jess identified only 1CF2 and not the entire family is that residues H299 and S302 are conserved only among hyperthermophilic archaeal GAPDHs and 1CF2 is the only such 3D structure available. This conservation perfectly fits with our working hypothesis whereby the triad could serve as a stabilizing motif in these hyperstable proteins.

- **1JY1** (Davies *et al.*, 2002) **Tyrosyl-DNA phosphodiesterase**

Esterases are known to bear a catalytic triad and this structure makes no exception to the rule; however, the residues we identified for this human tyrosyl-DNA phosphodiesterase are different to the catalytic triad and not mentioned in the primary literature.

- **1IZN** (Yamashita *et al.*, 2003) **Actin capping protein**

The available literature on this heterodimeric actin capping protein only mentions residue E221 as being crucial for (i) maintaining the overall architecture by substantially contributing to the stabilization of the individual subunit structure, and (ii) stabilizing the dimer, as E221 also interacts with residues of the second subunit *via* hydrogen bonding.

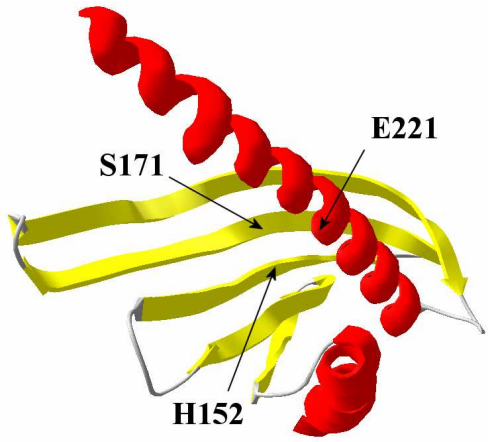


Fig. 27: Partial representation of the C-terminal domain of the β subunit, illustrating the position of the triad.

The triad residues are all located in the C-terminal domain of the β subunit, but on different secondary structures (Fig. 27). E221 is located in the middle of a long helix which stretches over an anti-parallel β-sheet formed by five consecutive β-strands, whereas H152 and S171 are found on two adjacent β-strands. As neither His nor Ser are mentioned in the literature, the unexpected presence of a motif such as a "stabilizing" triad could provide further information on how the overall fold is maintained.

Structures bearing an amylase-like Ser-His-*Asp* triad

Concerning the structures **1QHM** (Leppanen *et al.*, 1999), **1BGL** (Jacobson *et al.*, 1994), **1KW0** (Andersen *et al.*, 2002), **1UKP** (Kang *et al.*, 2003), **1DE6** (Korndorfer *et al.*, 2000), **1SI0** (Shouldice *et al.*, 2004), **1MRJ** (Huang *et al.*, 1995) and **1DEL** (Teplyakov *et al.*, 1996), neither a triad nor any of the residues composing this triad are mentioned in the relevant literature.

The only 2 structures requiring additional information are:

- **1H46** (Munoz *et al.*, 2003) **Exoglucanase**

The particularity of this hit is that it is not the Ser OG atom which is implicated in the triad but the main chain carbonyl oxygen atom of Ser. Furthermore, the identified triad of this exoglucanase or exocellulase is distinct from the active site.

- **1ASG** (Danishefsky *et al.*, 1991) **Aspartate aminotransferase**

Residues D223 and H145 of this aspartate aminotransferase are part of the 25 known catalytic residues, of which D223, K258 and W142 are the most important. D223 is implicated directly in catalysis whereas H145 is also strictly conserved but only seems to be involved in formation of the active site pocket. The reason why we kept this convincingly matching structure in the list is because S141 is not mentioned in the literature.

Structures bearing an amylase-like Ser-His-*X* triad

As mentioned, this is the second type of non-catalytic amylase-like triad in which Asp or Glu are substituted by other residues or even ligands. In many cases, this substituted third member of the triad is a Ser residue. However, cross-checking of these structures showed that their complete side chains are not superimposable with each other, prompting us to consider them independently.

Tab. 3: His-Ser-X non-catalytic amylase-like triads.

PDB	rmsd (Å)	Δ (Å)	ENZYME NAME	TOT.	SER	HIS	3rd RES or LIGAND
1HXH	0.05	0.07	hydroxysteroid dehydrogenase	1	S141	H182	S180
1GU6	0.12	0.14	cytochromeC nitrite reductase	1	S255	H351	Q378
1LI1	0.13	0.17	collagen IV	4	S163	H121	S142
1OH9	0.14	0.15	acetylglutamate kinase	3	S33	H34	N248
1M4N	0.15	0.2	1-aminocyclopropane-1-carboxylate synthase	1	S287	H266	H226
1MGN	0.16	0.31	metmyoglobin	5	S92	H93	heme-Fe154
1O7G	0.16	0.38	naphtalene 1,2-dioxygenase	7	S347	H261	S310
1VK9	0.17	0.15	hypothetical protein	1	S96	H74	S94
1E5Q	0.17	0.15	saccharopine oxydoreductase	1	S74	H81	S99
1TS3	0.17	0.17	toxin of toxin shock syndrom	1	S111	H141	Q136
1A2K	0.17	0.43	nuclear transport factor NTF2	3	S83	H73	Q71
1K0W	0.18	0.14	ribulose-P-epimerase	3	S30	H80	N28
1A8R	0.18	0.18	GTP cyclohydrolase	1	S112	H113	GTP
1JYW	0.18	0.2	Lac Z	2	S998	H540	O6 of substrate
1L5V	0.19	0.31	maltodextrin phosphorylase	1	S132	H60	T134
1TME	0.19	0.32	virus shell protein	1	S119	H122	S240
1QO7	0.2	0.31	epoxide hydrolase	1	S148	H97	Q86
1RKX	0.26	0.18	glucose dehydratase	1	S40	H17	S61
1W15	0.33	0.16	calcium binding protein (synaptotagmin)	1	S400	H420	H404
1B8Y	0.34	0.16	metalloproteinase (cat. dom. of stromelysin)	1	S133	H96	T131
1CIY	0.36	0.19	*Bacillus* insecticidal toxin	1	S428	H429	R265

For structures **1GU6** (Bamford *et al.*, 2002), **1LI1** (Than *et al.*, 2002), **1OH9** (Gil-Ortiz *et al.*, 2003), **1M4N** (Capitani *et al.*, 2003b), **1A2K** (Stewart *et al.*, 1998), **1A8R** (Nar *et al.*, 1995, Auerbach *et al.*, 2000), **1L5V** (Geremia *et al.*, 2002), **1RKX** (Vogan *et al.*, 2004), **1W15** (Dai *et al.*, 2004), **1B8Y** (Parker *et al.*, 1999, Pavlovsky *et al.*, 1999), **1TME** (Grant *et al.*, 1992), **1QO7** (Zou *et al.*, 2000) and **1CIY** (Grochulski *et al.*, 1995, Li *et al.*, 1991), neither a triad nor any of the identified residues are mentioned.

- **1HXH** (Benach *et al.*, 2002) **Hydroxysteroid dehydrogenase**

An almost perfect superimposition with the non-catalytic triad of AHA is demonstrated by this triad, with surprisingly low rmsd and Δ values of, respectively, 0.05 and 0.07 Å. This bacterial hydroxysteroid dehydrogenase from *Comamonas testosteroni* is in its free form in the PDB file. Mutagenesis studies have demonstrated that His182 is important for cofactor NAD(H) binding, although not directly involved. His182 is known to establish hydrogen bonds to the side chains of Ser141 and Ser180 as well as the main chain atoms of Ser138, and appears to form a salt bridge with the side chain of Glu241. This network of interactions seems responsible for linking the C-terminal substrate binding region to the active site and thus illustrates that "*residues not directly located at the active site contribute to catalysis in a complex manner through support of correct conformations of the active-site residues*" (Benach *et al.*, 2002). This

statement is absolutely in accordance with our hypothesis and the authors do not mention any triad configuration.

- **1MGN** (Hargrove *et al.*, 1994, Quillin *et al.*, 1993) **Myoglobin**

Interestingly, these triads (5 hits) are composed of S92, H93 and the Fe^{3+} ion of the heme in the myoglobin or metmyoglobin structures. Such motifs are called His-coordinated hemes. However, this does not mean that all hemes in all structures are His-coordinated, or that all His-coordinated hemes are superimposable with the amylase template. This might be the reason why Jess did not identify any matching hemoglobin structure, although hemoglobin is known to coordinate hemes with His residues. However, myoglobin and related structures were found to bear such a superimposable Ser-His-Fe^{3+} triad.

In relation to our hypothesis, several very interesting observations have been reported for myoglobin structures:

(i) the energy of the Fe^{3+}-H93 bond equals that of a covalent bond,

(ii) S92 is known to establish stabilizing hydrogen bonds with H93,

(iii) the Fe^{3+}-H93 bond length (2-2.6 Å, depending on the considered structure) as well as the protonation state of H93 are critical for efficient heme binding (His needs to be deprotonated; pK_a estimated to be 5.8),

(iv) the ^{1}H-NMR spectrum shows a highly downshifted peak, absent in the H93Y mutant spectrum and which has been attributed to the side chain amine proton of H93 (Lloyd *et al.*, 1996).

(v) Finally, mutational studies have demonstrated that the H93Y mutation leads to a significant increase in the rate of heme loss. However, although the tyrosyl side chain is larger than that of the histidyl residue, the heme binding site can adjust to accommodate the new residue without concomitant decrease in overall stability, thus possibly explaining why the mutant remains stable (Hildebrand *et al.*, 1995). Furthermore, the S92D mutation results in almost no change in stability and this can be explained either by the introduction of new hydrogen bonds by the substituting residue, or by the hydrogen bonding of the L89 carbonyl group to the H93 imidazole ring, sufficient to preserve the wild-type configuration. These results differ from our observations as none of the AHA mutants prepared remained stable.

Although the coordination of the heme by a His residue is well described, none of the authors noticed the similarity between the triad motif and the structure of the coordinated heme. His and Ser residues are often considered as being catalytic, but from our point of view, they are only statically implicated in binding, in contrast to a dynamic catalytic action such as a charge-relay system or any movement of the residues. These are the reasons why we kept the myoglobin structures among the valid hits.

- **1O7G** (Karlsson *et al.*, 2003) **Naphtalene 1,2-dioxygenase**

This as yet unknown and non-catalytic Ser-His-Ser triad is in the amylase-like configuration with S347 being the conserved Ser residue providing the OG atom for putative hydrogen bonding. The identified oxygen atom of S310 is the main chain (amide) oxygen atom.

- **1VK9 Hypothetical protein**

There is no literature available for this hypothetical protein with unknown function from *Thermotoga maritima*. Concerning the triad and in contrast to the preceding structure, the identified Ser oxygen atoms are both OG atoms.

- **1E5Q** (Johansson *et al.*, 2000) **Saccharopine oxydoreductase**

Of the 3 identified residues, only S99 is known to establish a hydrogen bond with saccharopine; S74 and H81 are not mentioned. Indeed, S99 is located between the triad and the saccharopine molecule and thus the triad could serve in the positioning of S99.

- **1TS3** (Earhart *et al.*, 1998) **Toxin of the toxin shock syndrome**

Here, the Q136 OE1 atom forms hydrogen bonds with the H141 NE2 and the L113 N atoms and substitution of Q136 results in reorganization of a loop (β7-β9). This residue is therefore known to be structurally important. Although the importance of many neighboring residues of the triad is discussed in the literature (*e.g.* residues 140, 135, 139, 113, 114), the function of the S111 residue is not revealed.

- **1K0W** (Luo *et al.*, 2001) **Ribulose-P-epimerase**

Five residues in these epimerases, among which is N28, are highly conserved and known to establish hydrogen bonding with the phosphate group of the substrate.

The N28A mutation leads to a much higher K_M value, underlining the structural importance of this residue. Still, neither a triad nor S30 and H80 are mentioned.

- **1JYW** (Juers *et al.*, 2001) **Lac Z**

This PDB file of the β-galactosidase lacZ is comparable to 1BGL (Tab. 2); thus, both structures bear the as yet unknown S971-H972-D591 triad as identified for 1BGL, with identical rmsd and delta values. In addition to this motif, a second triad formed by residues S998, H540 and the substrate has been identified, but only for 1JYW as no substrate is bound in the case of 1BGL.

The implicated ligand is the heteroatom O6 of the sugar residue of the synthetic substrate 1-O-(p-nitrophenyl)-β-D-galactopyranose. The role of H540 together with other residues in substrate binding is described but residue S998 has not been analyzed. As for 1MGN (myoglobin) and in contrast to the authors opinion, we consider H540 in this study to be non-catalytic, as it only helps in holding the substrate in a suitable position without actually participating in catalysis.

4.3.3 Non-catalytic protease-like triads

In contrast to the non-catalytic amylase-like triads, structures of this second category have to bear a triad composed of a His residue in its protease-like catalytic position, the Ser OG atom now facing the His NE2 atom and any non-carbon side chain or ligand atom facing the His ND1 atom.

Tab. 4: Non-catalytic protease-like triads. The double line separates triads with Glu/Asp from triads with other residues.

PDB	rmsd (Å)	Δ (Å)	ENZYME NAME	TOT.	SER	HIS	3rd RES
1IHU	0.07	0.13	arsenite-translocating ATPase	3	S420	H148	E416
2BSP	0.11	0.1	pectate lyase	2	S18	H193	D150
1QNQ	0.12	0.16	β-mannanase	3	S158	H198	D195
1U43	0.19	0.33	2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase	1	S35	H42	D8
1R8W	0.26	0.18	glycerol dehydratase	2	S7	H368	D312
1OZF	0.07	0.14	acetolactate synthase (lyase)	3	S445	H469	S443
1MLA	0.1	0.08	acyltransferase	1	S295	H271	T293
1QWK	0.14	0.19	aldo-keto reductase (hypothetical prot)	1	S174	H187	N157
1IOW	0.17	0.18	D-Ala ligase	1	S274	H63	S22
1CL2	0.19	0.28	cystathionine β-lyase	1	S130	H103	T148
1A4M	0.2	0.18	adenosine deaminase	3	S291	H258	N289
1PMM	0.2	0.18	glutamate decarboxylase	1	S271	H241	S269
5RLA	0.22	0.16	arginase	9	S271	H111	H228
1G1Y	0.26	0.18	α-amylase II (neopullulanase)	1	S548	H527	T522

Only one PDB entry bearing a protease-like triad with a *Glu* residue and representing 3 similar structures with very low rmsd values, has been retained in the list of hits: **1IHU** (Zhou *et al.*, 2001). This type of ATPase is composed of 2 homologous domains, A1 and A2. The triad is located at the interface of these domains but at the opposite end of the molecule with respect to the nucleotide binding site. In this region, 3 arsenite (AsIII) or antimony (SbIII) ions are bound in the coordination of which H148 and S420 are implicated. Binding of these metalloids tightens the interaction between A1 and A2 and is important for ATP hydrolysis; H148 and S420 are thus structurally important residues, in agreement with our hypothesis.

Four triads with *Asp* have been retained; for structures **2BSP** (Pickersgill *et al.*, 1994) and **1R8W** (O'Brien *et al.*, 2004), representing 2 similar enzymes each, neither the triad nor any of its composing residues have been mentioned. This is also the case for the two remaining structures, but some additional information is given below.

In the 3 hits represented by **1QNQ** (Sabini *et al.*, 2000), the matching atom of S158 is the main chain oxygen. These β-mannanases, or β-mannosidases (EC 3.2.1.78), belong to glycoside hydrolase family 5 in Henrissat's classification and display a $(\beta/\alpha)_8$ architecture, this being different to the typical α-amylase $(\beta/\alpha)_8$ fold. Thus, β-mannanases belong to clan GH-A, in contrast to clan GH-H for α-amylases.

For **1U43** (Steinbacher *et al.*, 2002), all three residues of the triad are known to be conserved: H42 and D8 coordinate a zinc ion required for catalysis and S35 binds to a phosphate group of the substrate.

Finally, 9 structures representing a total of 21 redundant PDB entries bear a *Ser-His-X* triad. The primary publications mention neither a triad nor any of the identified residues for **1OZF** (Pang *et al.*, 2004), **1IOW** (Fan *et al.*, 1997), **1CL2** (Clausen *et al.*, 1997), **1A4M** (Wang and Quiocho, 1998), **5RLA** (Scolnick *et al.*, 1997) and **1G1Y** (Kondo *et al.*, 2001). For **1QWK**, no literature is available.

The structure **1MLA** (Serre *et al.*, 1995), with a typical $(\alpha/\beta)_8$ hydrolase architecture but with a unique fold, bears a catalytic triad composed of the residues S92, H201 and N250. This triad is distinct from that which we identified.

Surprisingly, the complete triad of the glutamate decarboxylase **1PMM** (Capitani *et al.*, 2003a) is almost perfectly superimposable with the triad of structure **1OZF**, although these have nothing else in common.

4.3.4 Additional remarks

As mentioned earlier in this document, a number of authors have already identified structures other than proteases and lipases that bear triads somehow similar to the catalytic motif. We thus wanted to compare our search results with those of other studies.

Wallace *et al*. (1996) have described such a motif in immunoglobulin 2igG and in cyclosporine. The first structure has not been identified by our method and these authors remain the only team that has discussed the occurrence of a triad in this immunoglobulin. In contrast, the cyclosporine structure was identified in our study, but deleted from the list as we now know this triad to be catalytic. This fact had already been indicated by Wallace *et al*.

David Blow (1990) mentioned the finding of an unknown triad in the chloride-*in*dependent TAKA-amylase. This is intriguing and as far as our amylase-based study is concerned, it is even an uncomfortable observation since we have not identified a convincing triad in this enzyme with Jess. We therefore analyzed this case in detail.

Indeed, using SwissPdb-Viewer and the respective PDB files, we could identify a strongly distorted non-catalytic triad; too distorted (distances and coplanarity) to be identified by Jess and superimposed with our templates. Moreover, the position of the triad in the structures of chloride-dependent α-amylases is highly conserved but completely differs from the position in the triad of the chloride-independent TAKA-amylase (Fig. 28), what is however not important for Jess analysis.

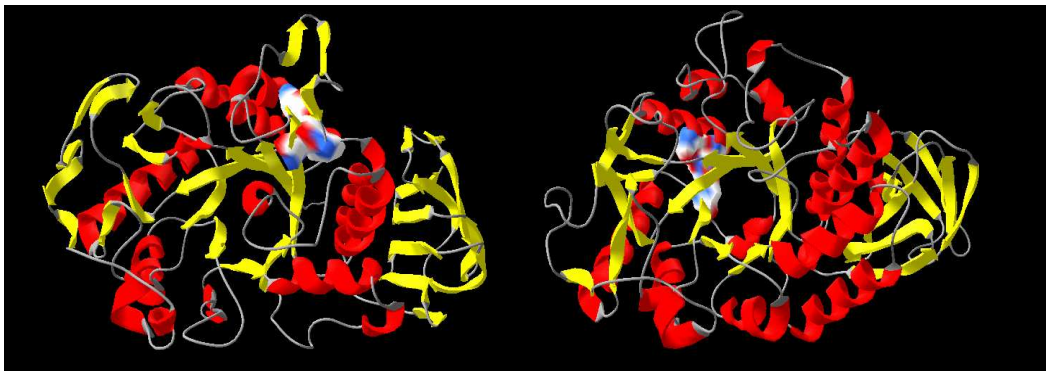


Fig. 28: Position of the triad in AHA (left structure) and in TAKA-amylase (*Aspergillus oryzae*) (right structure). Both molecules are oriented identically with domain C on the right side.

We also investigated other chloride-independent α-amylases such as from *Bacillus* sp. Again, a (too) highly distorted triad was found and the position of the motif is found to be different from both positions shown in Fig. 28.

This explains why Jess could not identify any triad in chloride-independent α-amylases and why we still may consider these enzymes to be devoid of a triad as described in this work.

4.4 CONCLUSION

After removal of enzymes bearing a catalytic triad, all the α-amylases, as well as the redundant structures from the list of 450 identified structures, 142 PDB entries were retained, showing that the occurrence of a non-catalytic triad in either orientation is not restricted to Cl^--dependent α-amylases.

The primary literature shows that these triads have not been reported so far. With no exception, these roles are described to be of a structural nature: coordination of ions and hemes, cofactor binding, binding and positioning of substrates either directly or indirectly, stabilization through hydrogen bonding, intra- and intersubunit stabilization are the keywords found in the corresponding publications.

It is perhaps noteworthy that out of the 25 non-redundant triads in both orientations and implicating Glu or Asp, 11 (44 %) are enzymes acting on carbohydrates. However, none of these structures belong to glycoside hydrolase family 13 or to the GH-H clan, as do the α-amylases.

The 142 PDB codes were split into two categories: those with an amylase-like and those with a protease-like non-catalytic triad. Within these categories, the entries were further ordered as a function of the third member of the triad, besides His and Ser which are the conserved residues in amylase and protease-like triads.

The identified triads that are in the amylase-like orientation and composed of Ser-His and Glu/Asp are the most important for us, as they are found in exact this configuration in α-amylases, our starting point. Glu and Asp were treated equivalently, since the chemical function of their side chain is identical. Out of the retained 142 entries, these structures alone account for 69 entries with some very interesting matches such as the SH2 domains, GAPDH from a hyperthermophilic Archaeon and the actin capping protein. For the 69 structures, the mean distances between the respective heteroatoms are 2.82±0.14 Å for Ser-His and 2.77±0.12 Å for His-Glu/Asp. The 11 hits in the protease-like orientation and implicating Glu/Asp display a greater mean distance between the Ser and His heteroatoms of interest (2.89±0.14 Å), but an even shorter

mean distance between His and Glu/Asp (2.71±0.1 Å). It is interesting to note that the distance between His and the acid is in accordance with the establishment of a strong hydrogen bond.

The definition of an LBHB, a strong or even of a regular hydrogen bond does not set any restrictions on the nature of the amino acid or the heteroatoms establishing the bond, therefore, nothing allowed us to *a priori* exclude the Ser-His-X triads whenever they fulfilled all the other conditions, such as distances and coplanarity. However, it is worth remembering that all LBHBs detected thus far on enzymes involve carboxyl groups (Mildvan *et al.*, 1999). The identified Ser-His-X triads can consequently be fortuitous hits, but the experimental data, especially on myoglobin, are appealing and suggest that these triads could be a not as yet described conformation favoring strong interactions or binding between the partners. Moreover, the distances are compatible with the establishment of a H-bond between His and the third residue for most of the retained Ser-His-X structures.

Indeed, the question as to whether a hit is fortuitous or not, can be addressed for all the identified triads. The probability of finding an arrangement of 4 or 5 atoms, such as in our templates, in a target database of ±30,000 structures that are composed of a few residues up to hundreds or even thousands of amino acids, is certainly not nil and is indeed very difficult to evaluate. Some arguments underline the validity of our screening method and discriminating between the different hits and thus indicate that the identified triads can not all be fortuitous and without any function. Firstly, in many cases, Jess identified more than one protein in the same family, which reduces the probability for each single hit to be fortuitous. Secondly, the primary literature, when describing the role of one or some of the implicated residues, always describes the roles to be of a structural nature, in agreement with our hypothesis. The third and most important argument underlining the validity of our hits is as follows. In our templates, we allowed the oxygen atom of the acid to be replaced by any non-carbon side chain (or ligand) atom without residue matching. If we only take into account the amino acid residues and not the ligands, the possibilities for replacing the Glu/Asp oxygen atoms in the amylase orientation, with respect to the imposed restrictions, are: Met SD, Thr OG, Tyr OH, Trp NE, Asn OD1 and ND2, Gln OE1 and NE2, Cys SG, Lys NZ, Arg NE, NH1 and NH2 and His ND1 and NE2 atoms. Thus, besides Glu/Asp, 15 possibilities exist that are all statistically equal for Jess. However, the list of matching amylase-like

structures displays 69 hits with Glu/Asp, against 41 hits for all 15 possibilities mentioned above. If we add the chloride-dependent α-amylases to our estimation, we find a total of 69 structures + 36 α-amylase structures, *i.e.* 105 hits with Ser-His-Glu/Asp triads versus 41 hits with a Ser-His-X triad. In other words, if all the hits were fortuitous, we should observe a more even distribution of all the replacement possibilities, this being clearly not the case.

Even though these arguments do not provide any guarantee for selection of an arrangement without any role and which we mistook for a triad, they strongly support the assumption that the presence of a non-catalytic triad cannot be fortuitous in all 142 cases.

These aspects indicate that the structures identified with our methods should be valid hits and that the presence of a non-catalytic triad is neither restricted to Cl^--dependent α-amylases, nor is it fortuitous. Even more important is that in the light of the experimental data of some of the identified proteins, the non-catalytic triad could indeed fulfill an important structural role, in agreement with our working hypothesis and our experimental data.

5. GENERAL CONCLUSION AND PERSPECTIVES

The strict conservation, from bacteria to human, of the 3 residues constituting the non-catalytic triad in chloride-dependent α-amylases indicates that the presence of this motif cannot be fortuitous. The results of our mutagenesis studies indicate that the non-catalytic triad fulfills undeniably an important structural role. This is the first time that a triad, previously only known as a catalytic motif, is reported to be of structural importance with only minor differences if compared to the catalytic triad. We therefore propose that the non-catalytic triad is a novel structural motif in the conformational stability of proteins.

In an attempt to explain how the two putative hydrogen bonds of the non-catalytic triad may influence the stability of an entire structure, we investigated this motif for the presence of an unusually strong hydrogen bond, which are most readily studied by NMR spectroscopy. The occurrence of a strong hydrogen bond in the non-catalytic triad could account for the instability of mutants of the participating residues: as calculated for AHA, the disruption of an LBHB can result in a ±10% decrease of the enthalpy of unfolding of the entire enzyme.

NMR spectra of AHA indeed revealed features that are identical with the NMR signature of catalytic triads. Among these observations, the most important is the downfield resonance signal indicative of a low-barrier hydrogen bond. In our case, this signal was attributed to a His imidazolium ring, but the unambiguous attribution to the His residue of the non-catalytic triad has failed for the mentioned reasons. However, many additional aspects suggest that the origin of the signals for AHA is very likely to be the non-catalytic triad, bearing a strong or low-barrier hydrogen bond between a His imidazolium nitrogen atom and a Glu/Asp carboxyl oxygen atom.

Assuming that the His residue of the non-catalytic triad is indeed the origin of our observations, we proposed a model accounting for the sum of NMR signals with respect to their underlying conditions. The model could also explain why the triad in its non-catalytic form cannot have an auxiliary hydrolytic activity as no charges can be relayed. This leads us to the discussion of the possible role of the Ser residue in the triad. In contrast with the catalytic triad, we excluded the formation of an alkoxide ion

in the non-catalytic triad, but proposed that this residue is essential for positioning and thus, for avoiding the charges to be relayed.

Finally, we scanned the PDB with a powerful constraint-based algorithm and found that the presence of such a non-catalytic triad is not restricted to chloride-dependent α-amylases. The primary literature of the identified structures bearing a triad neither mentions the word triad nor describes any resemblance with this motif. Moreover, in the cases where the role of some of the residues is discussed, it is without exception described to fulfill a structural role. The sum of these observations underlines once more that the non-catalytic triad is a novel *and* structural motif in the conformational stability of proteins.

Concerning the perspectives, the idea of a novel structural motif should immediately lead in the protein engineer's mind to the question if it is possible to introduce this motif in structures initially devoid of it. Although many enzymes display a $(\beta/\alpha)_8$ barrel fold, this fold still varies enough to render it very difficult to find suitable sites for the introduction of the triad residues. Extreme care should be taken to respect the bond distances and coplanarity of the imidazole ring and the oxygen atoms, essential for the correct establishment of hydrogen bonds of any type. However, a similar project has successfully been realized with the catalytic triad, demonstrating that the transfer of this motif with establishment of a strong hydrogen bond is possible (Tanaka and Yada, 2004). In this study, the catalytic residues of an aspartic protease were substituted with the catalytic triad characteristic of a serine protease using trypsin as model. As a result, some mutants showed correct peptide hydrolysis and were significantly inhibited by serine protease inhibitors.

BIBLIOGRAPHY

Aghajari, N., Feller, G., Gerday, C. and Haser, R. (1996): Crystallization and preliminary X-ray diffraction studies of α-amylase from the antarctic psychrophile *Alteromonas haloplanctis* A23. *Protein Sci.,* **5,** 2128-2129.

Aghajari, N., Feller, G., Gerday, C. and Haser, R. (1998a): Crystal structures of the psychrophilic α-amylase from *Alteromonas haloplanctis* in its native form and complexed with an inhibitor. *Protein Sci,* **7,** 564-572.

Aghajari, N., Feller, G., Gerday, C. and Haser, R. (1998b): Structures of the psychrophilic *Alteromonas haloplanctis* α-amylase give insights into cold adaptation at a molecular level. *Structure,* **6,** 1503-1516.

Aghajari, N., Feller, G., Gerday, C. and Haser, R. (2002): Structural basis of alpha-amylase activation by chloride. *Protein Sci,* **11,** 1435-41.

Andersen, O. A., Flatmark, T. and Hough, E. (2002): Crystal structure of the ternary complex of the catalytic domain of human phenylalanine hydroxylase with tetrahydrobiopterin and 3-(2-thienyl)-L-alanine, and its implications for the mechanism of catalysis and substrate activation. *J. Mol. Biol.,* **320,** 1095-108.

Artymiuk, P. J., Poirrette, A. R., Grindley, H. M., Rice, D. W. and Willett, P. (1994): A graph-theoretic approach to the identification of three-dimensional patterns of amino acid side-chains in protein structures. *J. Mol. Biol.,* **243,** 327-44.

Ash, E. L., Sudmeier, J. L., Day, R. M., Vincent, M., Torchilin, E. V., Haddad, K. C., Bradshaw, E. M., Sanford, D. G. and Bachovchin, W. W. (2000): Unusual ^{1}H NMR chemical shifts support (His) C(epsilon) 1...O=C H-bond: proposal for reaction-driven ring flip mechanism in serine protease catalysis. *Proc. Natl. Acad. Sci. U S A,* **97,** 10371-6.

Auerbach, G., Herrmann, A., Bracher, A., Bader, G., Gutlich, M., Fischer, M., Neukamm, M., Garrido-Franco, M., Richardson, J., Nar, H., Huber, R. and Bacher, A. (2000): Zinc plays a key role in human and bacterial GTP cyclohydrolase I. *Proc. Natl. Acad. Sci. U S A,* **97,** 13567-72.

Bachovchin, W. W. (2001): Contributions of NMR spectroscopy to the study of hydrogen bonds in serine protease active sites. *Magn. Reson. Chem.,* **39,** S199-S213.

Bamford, V. A., Angove, H. C., Seward, H. E., Thomson, A. J., Cole, J. A., Butt, J. N., Hemmings, A. M. and Richardson, D. J. (2002): Structure and spectroscopy of the periplasmic cytochrome c nitrite reductase from *Escherichia coli*. *Biochemistry,* **41,** 2921-31.

Barker, J. A. and Thornton, J. M. (2003): An algorithm for constraint-based structural template matching: application to 3D templates with statistical analysis. *Bioinformatics,* **19,** 1644-9.

Barth, A., Frost, K., Wahab, M., Brandt, W., Schadler, H. D. and Franke, R. (1994): Classification of serine proteases derived from steric comparisons of their active sites, part II: "Ser, His, Asp arrangements in proteolytic and nonproteolytic proteins". *Drug Des. Discov.,* **12,** 89-111.

Barth, A., Wahab, M., Brandt, W. and Frost, K. (1993): Classification of serine proteases derived from steric comparisons of their active sites. *Drug Des. Discov.,* **10,** 297-317.

Baskakov, I. and Bolen, D. W. (1998): Forcing thermodynamically unfolded proteins to fold. *J. Biol. Chem.,* **273,** 4831-4.

Benach, J., Filling, C., Oppermann, U. C., Roversi, P., Bricogne, G., Berndt, K. D., Jornvall, H. and Ladenstein, R. (2002): Structure of bacterial 3beta/17beta-hydroxysteroid dehydrogenase at 1.2 Å resolution: a model for multiple steroid recognition. *Biochemistry,* **41,** 14659-68.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. and Bourne, P. E. (2000): The Protein Data Bank. *Nucleic Acids Res.,* **28,** 235-42.

Blow, D. (1990): Enzymology. More of the catalytic triad. *Nature,* **343,** 694-5.

Blow, D. M., Birktoft, J. J. and Hartley, B. S. (1969): Role of a buried acid group in the mechanism of action of chymotrypsin. *Nature,* **221,** 337-40.

Boel, E., Brady, L., Brzozowski, A. M., Derewenda, Z., Dodson, G. G., Jensen, V. J., Petersen, S. B., Swift, H., Thim, L. and Woldike, H. F. (1990): Calcium binding in alpha-amylases: an X-ray diffraction study at 2.1 Å resolution of two enzymes from *Aspergillus*. *Biochemistry,* **29,** 6244-9.

Brayer, G. D., Luo, Y. and Withers, S. G. (1995): The structure of human pancreatic alpha-amylase at 1.8 Å resolution and comparisons with related enzymes. *Protein Sci.,* **4,** 1730-42.

Buisson, G., Duee, E., Haser, R. and Payan, F. (1987): Three dimensional structure of porcine pancreatic alpha-amylase at 2.9 Å resolution. Role of calcium in structure and activity. *Embo J.,* **6,** 3909-16.

Capitani, G., De Biase, D., Aurizi, C., Gut, H., Bossa, F. and Grutter, M. G. (2003a): Crystal structure and functional analysis of *Escherichia coli* glutamate. decarboxylase *Embo J.,* **22,** 4027-37.

Capitani, G., Eliot, A. C., Gut, H., Khomutov, R. M., Kirsch, J. F. and Grutter, M. G. (2003b): Structure of 1-aminocyclopropane-1-carboxylate synthase in complex with an amino-oxy analogue of the substrate: implications for substrate binding. *Biochim. Biophys. Acta,* **1647,** 55-60.

Charron, C., Talfournier, F., Isupov, M. N., Littlechild, J. A., Branlant, G., Vitoux, B. and Aubry, A. (2000): The crystal structure of D-glyceraldehyde-3-phosphate dehydrogenase from the hyperthermophilic archaeon *Methanothermus fervidus* in the presence of NADP(+) at 2.1 Å resolution. *J. Mol. Biol.,* **297,** 481-500.

Cho, S., Velikovsky, C. A., Swaminathan, C. P., Houtman, J. C., Samelson, L. E. and Mariuzza, R. A. (2004): Structural basis for differential recognition of tyrosine-phosphorylated sites in the linker for activation of T cells (LAT) by the adaptor Gads. *Embo J,* **23,** 1441-51.

Chruszcz, M., Laidler, P., Monkiewicz, M., Ortlund, E., Lebioda, L. and Lewinski, K. (2003): Crystal structure of a covalent intermediate of endogenous human arylsulfatase A. *J Inorg Biochem,* **96,** 386-92.

Clausen, T., Huber, R., Messerschmidt, A., Pohlenz, H. D. and Laber, B. (1997): Slow-binding inhibition of *Escherichia coli* cystathionine beta-lyase by L-

aminoethoxyvinylglycine: a kinetic and X-ray study. *Biochemistry,* **36,** 12633-43.

Da Lage, J. L., Feller, G. and Janecek, S. (2004): Horizontal gene transfer from Eukarya to bacteria and domain shuffling: the alpha-amylase model. *Cell. Mol. Life Sci.,* **61,** 97-109.

Dai, H., Shin, O. H., Machius, M., Tomchick, D. R., Sudhof, T. C. and Rizo, J. (2004): Structural basis for the evolutionary inactivation of Ca^{2+} binding to synaptotagmin 4. *Nature Struct. Mol. Biol.,* **11,** 844-9.

D'Amico, S., Gerday, C. and Feller, G. (2000): Structural similarities and evolutionary relationships in chloride-dependent alpha-amylases. *Gene,* **253,** 95-105.

D'Amico, S., Gerday, C. and Feller, G. (2003): Temperature adaptation of proteins: engineering mesophilic-like activity and stability in a cold-adapted alpha-amylase. *J. Mol. Biol.,* **332,** 981-8.

Danishefsky, A. T., Onnufer, J. J., Petsko, G. A. and Ringe, D. (1991): Activity and structure of the active-site mutants R386Y and R386F of *Escherichia coli* aspartate aminotransferase. *Biochemistry,* **30,** 1980-5.

Davies, D. R., Interthal, H., Champoux, J. J. and Hol, W. G. (2002): The crystal structure of human tyrosyl-DNA phosphodiesterase, Tdp1. *Structure,* **10,** 237-48.

Earhart, C. A., Mitchell, D. T., Murray, D. L., Pinheiro, D. M., Matsumura, M., Schlievert, P. M. and Ohlendorf, D. H. (1998): Structures of five mutants of toxic shock syndrome toxin-1 with reduced biological activity. *Biochemistry,* **37,** 7194-202.

Fan, C., Park, I. S., Walsh, C. T. and Knox, J. R. (1997): D-alanine:D-alanine ligase: phosphonate and phosphinate intermediates with wild type and the Y216F mutant. *Biochemistry,* **36,** 2531-8.

Feller, G., d'Amico, D. and Gerday, C. (1999): Thermodynamic stability of a cold-active alpha-amylase from the Antarctic bacterium *Alteromonas haloplanctis. Biochemistry,* **38,** 4613-9.

Feller, G., le Bussy, O., Houssier, C. and Gerday, C. (1996): Structural and functional aspects of chloride binding to *Alteromonas haloplanctis* alpha-amylase. *J. Biol. Chem.,* **271,** 23836-23841.

Feller, G., Lonhienne, T., Deroanne, C., Libioulle, C., Van Beeumen, J. and Gerday, C. (1992): Purification, characterization, and nucleotide sequence of the thermolabile α-amylase from the antarctic psychrotroph *Alteromonas haloplanctis* A23. *J. Biol. Chem.,* **267,** 5217-5221.

Feller, G., Payan, F., Theys, F., Qian, M., Haser, R. and Gerday, C. (1994): Stability and structural analysis of α-amylase from the antarctic psychrophile *Alteromonas haloplanctis* A23. *Eur. J. Biochem.,* **222,** 441-447.

Fischer, D., Wolfson, H., Lin, S. L. and Nussinov, R. (1994): Three-dimensional, sequence order-independent structural comparison of a serine protease against the crystallographic database reveals active site similarities: potential implications to evolution and to protein folding. *Protein Sci.,* **3,** 769-78.

Frey, P. A., Whitt, S. A. and Tobin, J. B. (1994): A low-barrier hydrogen bond in the catalytic triad of serine proteases. *Science,* **264,** 1927-30.

Geremia, S., Campagnolo, M., Schinzel, R. and Johnson, L. N. (2002): Enzymatic catalysis in crystals of *Escherichia coli* maltodextrin phosphorylase. *J. Mol. Biol.,* **322,** 413-23.

Gil-Ortiz, F., Ramon-Maiques, S., Fita, I. and Rubio, V. (2003): The course of phosphorus in the reaction of N-acetyl-L-glutamate kinase, determined from the structures of crystalline complexes, including a complex with an AlF(4)(-) transition state mimic. *J. Mol. Biol.,* **331,** 231-44.

Grant, R. A., Filman, D. J., Fujinami, R. S., Icenogle, J. P. and Hogle, J. M. (1992): Three-dimensional structure of Theiler virus. *Proc. Natl. Acad. Sci. U S A,* **89,** 2061-5.

Grochulski, P., Masson, L., Borisova, S., Pusztai-Carey, M., Schwartz, J. L., Brousseau, R. and Cygler, M. (1995): *Bacillus thuringiensis* CryIA(a) insecticidal toxin: crystal structure and channel formation. *J. Mol. Biol.,* **254,** 447-64.

Hargrove, M. S., Singleton, E. W., Quillin, M. L., Ortiz, L. A., Phillips, G. N., Jr., Olson, J. S. and Mathews, A. J. (1994): His64(E7)-->Tyr apomyoglobin as a reagent for measuring rates of hemin dissociation. *J. Biol.Chem.,* **269,** 4207-14.

Henrissat, B. (1991): A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J,* **280 (Pt 2),** 309-16.

Hibbert, F. and Emsley, J. (1990): Hydrogen Bonding and Chemical Reactivity. *Adv. Phys. Organ. Chem.,* **26,** 255-379.

Hildebrand, D. P., Burk, D. L., Maurus, R., Ferrer, J. C., Brayer, G. D. and Mauk, A. G. (1995): The proximal ligand variant His93Tyr of horse heart myoglobin. *Biochemistry,* **34,** 1997-2005.

Huang, Q., Liu, S., Tang, Y., Jin, S. and Wang, Y. (1995): Studies on crystal structures, active-centre geometry and depurinating mechanism of two ribosome-inactivating proteins. *Biochem. J,* **309 (Pt 1),** 285-98.

Jacobson, R. H., Zhang, X. J., DuBose, R. F. and Matthews, B. W. (1994): Three-dimensional structure of beta-galactosidase from *E. coli*. *Nature,* **369,** 761-6.

Janecek, S. (1994): Sequence similarities and evolutionary relationships of microbial, plant and animal alpha-amylases. *Eur. J. Biochem.,* **224,** 519-24.

Janecek, S. (1997): alpha-Amylase family: molecular biology and evolution. *Prog. Biophys. Mol. Biol.,* **67,** 67-97.

Johansson, E., Steffens, J. J., Lindqvist, Y. and Schneider, G. (2000): Crystal structure of saccharopine reductase from *Magnaporthe grisea*, an enzyme of the alpha-aminoadipate pathway of lysine biosynthesis *Structure,* **8,** 1037-47.

Juers, D. H., Heightman, T. D., Vasella, A., McCarter, J. D., Mackenzie, L., Withers, S. G. and Matthews, B. W. (2001): A structural view of the action of *Escherichia coli* (lacZ) beta-galactosidase. *Biochemistry,* **40,** 14781-94.

Kang, Y. N., Adachi, M., Mikami, B. and Utsumi, S. (2003): Change in the crystal packing of soybean beta-amylase mutants substituted at a few surface amino acid residues. *Protein Eng,* **16,** 809-17.

Karlsson, A., Parales, J. V., Parales, R. E., Gibson, D. T., Eklund, H. and Ramaswamy, S. (2003): Crystal structure of naphthalene dioxygenase: side-on binding of dioxygen to iron. *Science,* **299,** 1039-42.

Koike, H., Sasaki, H., Kobori, T., Zenno, S., Saigo, K., Murphy, M. E., Adman, E. T. and Tanokura, M. (1998): 1.8 Å crystal structure of the major NAD(P)H:FMN

oxidoreductase of a bioluminescent bacterium, *Vibrio fischeri*: overall structure, cofactor and substrate-analog binding, and comparison with related flavoproteins. *J. Mol. Biol.*, **280,** 259-73.

Koike, H., Sasaki, H., Tanokura, M., Zenno, S. and Saigo, K. (1996): Crystallization and preliminary crystallographic analysis of the major NAD(P)H: FMN oxidoreductase of *Vibrio fischeri* ATCC 7744. *J. Struct. Biol.*, **117,** 70-2.

Kondo, S., Ohtaki, A., Tonozuka, T., Sakano, Y. and Kamitori, S. (2001): Studies on the hydrolyzing mechanism for cyclodextrins of *Thermoactinomyces vulgaris* R-47 alpha-amylase 2 (TVAII). X-ray structure of the mutant E354A complexed with beta-cyclodextrin, and kinetic analyses on cyclodextrins. *J. Biochem. (Tokyo)*, **129,** 423-8.

Korndorfer, I. P., Fessner, W. D. and Matthews, B. W. (2000): The structure of rhamnose isomerase from *Escherichia coli* and its relation with xylose isomerase illustrates a change between inter and intra-subunit complementation during evolution. *J. Mol. Biol.*, **300,** 917-33.

Leppanen, V. M., Merckel, M. C., Ollis, D. L., Wong, K. K., Kozarich, J. W. and Goldman, A. (1999): Pyruvate formate lyase is structurally homologous to type I ribonucleotide reductase. *Structure*, **7,** 733-44.

Levitzki, A. and Steer, M. L. (1974): The allosteric activation of mammalian alpha-amylase by chloride. *Eur. J. Biochem.*, **41,** 171-80.

Li, J. D., Carroll, J. and Ellar, D. J. (1991): Crystal structure of insecticidal delta-endotoxin from *Bacillus thuringiensis* at 2.5 Å resolution. *Nature*, **353,** 815-21.

Lin, T. H., Chen, C., Huang, R. F., Lee, Y. L., Shaw, J. F. and Huang, T. H. (1998): Multinuclear NMR resonance assignments and the secondary structure of *Escherichia coli* thioesterase/protease I: a member of a new subclass of lipolytic enzymes. *J. Biomol. NMR*, **11,** 363-80.

Lloyd, E., Burk, D. L., Ferrer, J. C., Maurus, R., Doran, J., Carey, P. R., Brayer, G. D. and Mauk, A. G. (1996): Electrostatic modification of the active site of myoglobin: characterization of the proximal Ser92Asp variant. *Biochemistry*, **35,** 11901-12.

Lonhienne, T., Gerday, C. and Feller, G. (2000): Psychrophilic enzymes: revisiting the thermodynamic parameters of activation may explain local flexibility. *Biochim. Biophys. Acta*, **1543,** 1-10.

Luo, Y., Samuel, J., Mosimann, S. C., Lee, J. E., Tanner, M. E. and Strynadka, N. C. (2001): The structure of L-ribulose-5-phosphate 4-epimerase: an aldolase-like platform for epimerization. *Biochemistry*, **40,** 14763-71.

Massiah, M. A., Viragh, C., Reddy, P. M., Kovach, I. M., Johnson, J., Rosenberry, T. L. and Mildvan, A. S. (2001): Short, strong hydrogen bonds at the active site of human acetylcholinesterase: proton NMR studies. *Biochemistry*, **40,** 5682-90.

McCarter, J. D. and Withers, S. G. (1994): Mechanisms of enzymatic glycoside hydrolysis. *Curr. Opin. Struct. Biol.*, **4,** 885-92.

McDonough, M. A., Kadirvelraj, R., Harris, P., Poulsen, J. C. and Larsen, S. (2004): Rhamnogalacturonan lyase reveals a unique three-domain modular structure for polysaccharide lyase family 4. *FEBS Lett.*, **565,** 188-94.

Mildvan, A. S., Harris, T. K. and Abeygunawardana, C. (1999): Nuclear Magnetic Resonance Methods for the Detection and Study of Low-Barrier Hydrogen Bonds on Enzymes. *Methods Enzymol.*, **308,** 219-245.

Munoz, I. G., Mowbray, S. L. and Stahlberg, J. (2003): The catalytic module of Cel7D from *Phanerochaete chrysosporium* as a chiral selector: structural studies of its complex with the beta blocker (R)-propranolol. *Acta Crystallogr. D Biol. Crystallogr.*, **59,** 637-43.

Nar, H., Huber, R., Auerbach, G., Fischer, M., Hosl, C., Ritz, H., Bracher, A., Meining, W., Eberhardt, S. and Bacher, A. (1995): Active site topology and reaction mechanism of GTP cyclohydrolase I. *Proc. Natl. Acad. Sci. U S A,* **92,** 12120-5.

Neal, S., Nip, A. M., Zhang, H. and Wishart, D. S. (2003): Rapid and accurate calculation of protein ^{1}H, ^{13}C and ^{15}N chemical shifts *J. Biomol. NMR,* **26,** 215-40.

Northrop, D. B. (2001): Follow the protons: a low-barrier hydrogen bond unifies the mechanisms of the aspartic proteases. *Acc. Chem. Res.*, **34,** 790-7.

Numao, S., Maurus, R., Sidhu, G., Wang, Y., Overall, C. M., Brayer, G. D. and Withers, S. G. (2002): Probing the role of the chloride ion in the mechanism of human pancreatic alpha-amylase. *Biochemistry,* **41,** 215-25.

O'Brien, J. R., Raynaud, C., Croux, C., Girbal, L., Soucaille, P. and Lanzilotta, W. N. (2004): Insight into the mechanism of the B12-independent glycerol dehydratase from *Clostridium butyricum*: preliminary biochemical and structural characterization. *Biochemistry,* **43,** 4635-45.

O'Gara, M., Zhang, X., Roberts, R. J. and Cheng, X. (1999): Structure of a binary complex of HhaI methyltransferase with S-adenosyl-L-methionine formed in the presence of a short non-specific DNA oligonucleotide. *J. Mol. Biol.,* **287,** 201-9.

Pang, S. S., Duggleby, R. G., Schowen, R. L. and Guddat, L. W. (2004): The crystal structures of *Klebsiella pneumoniae* acetolactate synthase with enzyme-bound cofactor and with an unusual intermediate. *J. Biol. Chem.,* **279,** 2242-53.

Parker, M. H., Lunney, E. A., Ortwine, D. F., Pavlovsky, A. G., Humblet, C. and Brouillette, C. G. (1999): Analysis of the binding of hydroxamic acid and carboxylic acid inhibitors to the stromelysin-1 (matrix metalloproteinase-3) catalytic domain by isothermal titration calorimetry. *Biochemistry,* **38,** 13592-601.

Pavlovsky, A. G., Williams, M. G., Ye, Q. Z., Ortwine, D. F., Purchase, C. F., 2nd, White, A. D., Dhanaraj, V., Roth, B. D., Johnson, L. L., Hupe, D., Humblet, C. and Blundell, T. L. (1999): X-ray structure of human stromelysin catalytic domain complexed with nonpeptide inhibitors: implications for inhibitor selectivity. *Protein Sci.,* **8,** 1455-62.

Pickersgill, R., Jenkins, J., Harris, G., Nasser, W. and Robert-Baudouy, J. (1994): The structure of *Bacillus subtilis* pectate lyase in complex with calcium. *Nat. Struct. Biol.,* **1,** 717-23.

Podust, L. M., Bach, H., Kim, Y., Lamb, D. C., Arase, M., Sherman, D. H., Kelly, S. L. and Waterman, M. R. (2004): Comparison of the 1.85 Å structure of CYP154A1 from *Streptomyces coelicolor* A3(2) with the closely related

CYP154C1 and CYPs from antibiotic biosynthetic pathways. *Protein Sci.,* **13,** 255-68.

Porter, C. T., Bartlett, G. J. and Thornton, J. M. (2004): The Catalytic Site Atlas: a resource of catalytic sites and residues identified in enzymes using structural data. *Nucleic Acids Res,* **32,** D129-33.

Qian, M., Ajandouz el, H., Payan, F. and Nahoum, V. (2005): Molecular basis of the effects of chloride ion on the acid-base catalyst in the mechanism of pancreatic alpha-amylase. *Biochemistry,* **44,** 3194-201.

Qian, M., Haser, R. and Payan, F. (1993): Structure and molecular model refinement of pig pancreatic alpha- amylase at 2.1 Å resolution. *J. Mol. Biol.,* **231,** 785-99.

Quillin, M. L., Arduini, R. M., Olson, J. S. and Phillips, G. N., Jr. (1993): High-resolution crystal structures of distal histidine mutants of sperm whale myoglobin. *J. Mol. Biol.,* **234,** 140-55.

Ramasubbu, N., Paloth, V., Luo, Y., Brayer, G. D. and Levine, M. J. (1996): Structure of Human Salivary α-Amylase at 1.6 Å Resolution: Implications for Its Role in the Oral Cavity. *Acta Crystallogr.,* **D52,** 435-446.

Robillard, G. and Shulman, R. G. (1972): High resolution nuclear magnetic resonance study of the histidine--aspartate hydrogen bond in chymotrypsin and chymotrypsinogen. *J. Mol. Biol.,* **71,** 507-11.

Ryan, M., Liu, T., Dahlquist, F. W. and Griffith, O. H. (2001): A catalytic diad involved in substrate-assisted catalysis: NMR study of hydrogen bonding and dynamics at the active site of phosphatidylinositol-specific phospholipase C. *Biochemistry,* **40,** 9743-50.

Sabini, E., Schubert, H., Murshudov, G., Wilson, K. S., Siika-Aho, M. and Penttila, M. (2000): The three-dimensional structure of a *Trichoderma reesei* beta-mannanase from glycoside hydrolase family 5. *Acta Crystallogr. D Biol. Crystallogr.,* **56,** 3-13.

Schramm, M. and Loyter, A. (1966): Purification of alpha-Amylases by Precipitation of Amylase-Glycogen Complexes. *Methods Enzymol.,* **8,** 533-537.

Scolnick, L. R., Kanyo, Z. F., Cavalli, R. C., Ash, D. E. and Christianson, D. W. (1997): Altering the binuclear manganese cluster of arginase diminishes thermostability and catalytic function. *Biochemistry,* **36,** 10558-65.

Seavey, B. R., Farr, E. A., Westler, W. M. and Markley, J. L. (1991): A relational database for sequence-specific protein NMR data. *J. Biomol. NMR,* **1,** 217-36.

Serre, L., Verbree, E. C., Dauter, Z., Stuitje, A. R. and Derewenda, Z. S. (1995): The *Escherichia coli* malonyl-CoA:acyl carrier protein transacylase at 1.5 Å resolution. Crystal structure of a fatty acid synthase component. *J. Biol. Chem.,* **270,** 12961-4.

Shouldice, S. R., Skene, R. J., Dougan, D. R., Snell, G., McRee, D. E., Schryvers, A. B. and Tari, L. W. (2004): Structural basis for iron binding and release by a novel class of periplasmic iron-binding proteins found in gram-negative pathogens. *J. Bacteriol.,* **186,** 3903-10.

Steinbacher, S., Kaiser, J., Wungsintaweekul, J., Hecht, S., Eisenreich, W., Gerhardt, S., Bacher, A. and Rohdich, F. (2002): Structure of 2C-methyl-d-erythritol-2,4-cyclodiphosphate synthase involved in mevalonate-independent biosynthesis of isoprenoids. *J. Mol. Biol.,* **316,** 79-88.

Steitz, T. A. and Shulman, R. G. (1982): Crystallographic and NMR studies of the serine proteases. *Annu Rev. Biophys. Bioeng.*, **11,** 419-44.

Stewart, M., Kent, H. M. and McCoy, A. J. (1998): Structural basis for molecular recognition between nuclear transport factor 2 (NTF2) and the GDP-bound form of the Ras-family GTPase Ran. *J. Mol. Biol.*, **277,** 635-46.

Stranzl, G. R., Gruber, K., Steinkellner, G., Zangger, K., Schwab, H. and Kratky, C. (2004): Observation of a short, strong hydrogen bond in the active site of hydroxynitrile lyase from *Hevea brasiliensis* explains a large pK_a shift of the catalytic base induced by the reaction intermediate. *J. Biol. Chem.*, **279,** 3699-707.

Strobl, S., Maskos, K., Betz, M., Wiegand, G., Huber, R., Gomis-Ruth, F. X. and Glockshuber, R. (1998): Crystal structure of yellow meal worm alpha-amylase at 1.64 A resolution. *J. Mol. Biol.*, **278,** 617-28.

Tanaka, T. and Yada, R. Y. (2004): Redesign of catalytic center of an enzyme: aspartic to serine proteinase. *Biochem. Biophys. Res. Commun*, **323,** 947-53.

Teplyakov, A., Sebastiao, P., Obmolova, G., Perrakis, A., Brush, G. S., Bessman, M. J. and Wilson, K. S. (1996): Crystal structure of bacteriophage T4 deoxynucleotide kinase with its substrates dGMP and ATP. *Embo J.*, **15,** 3487-97.

Than, M. E., Henrich, S., Huber, R., Ries, A., Mann, K., Kuhn, K., Timpl, R., Bourenkov, G. P., Bartunik, H. D. and Bode, W. (2002): The 1.9- Å crystal structure of the noncollagenous (NC1) domain of human placenta collagen IV shows stabilization via a novel type of covalent Met-Lys cross-link. *Proc. Natl. Acad. Sci. U S A*, **99,** 6607-12.

Tyukhtenko, S. I., Litvinchuk, A. V., Chang, C. F., Leu, R. J., Shaw, J. F. and Huang, T. H. (2002): NMR studies of the hydrogen bonds involving the catalytic triad of *Escherichia coli* thioesterase/protease I. *FEBS Lett.*, **528,** 203-6.

Uitdehaag, J. C., Mosi, R., Kalk, K. H., van der Veen, B. A., Dijkhuizen, L., Withers, S. G. and Dijkstra, B. W. (1999): X-ray structures along the reaction pathway of cyclodextrin glycosyltransferase elucidate catalysis in the alpha-amylase family. *Nat. Struct. Biol.*, **6,** 432-6.

Viragh, C., Harris, T. K., Reddy, P. M., Massiah, M. A., Mildvan, A. S. and Kovach, I. M. (2000): NMR evidence for a short, strong hydrogen bond at the active site of a cholinesterase. *Biochemistry*, **39,** 16200-5.

Vogan, E. M., Bellamacina, C., He, X., Liu, H. W., Ringe, D. and Petsko, G. A. (2004): Crystal structure at 1.8 Å resolution of CDP-D-glucose 4,6-dehydratase from *Yersinia pseudotuberculosis*. *Biochemistry*, **43,** 3057-67.

Wallace, A. C., Borkakoti, N. and Thornton, J. M. (1997): TESS: a geometric hashing algorithm for deriving 3D coordinate templates for searching structural databases. Application to enzyme active sites. *Protein Sci.*, **6,** 2308-23.

Wallace, A. C., Laskowski, R. A. and Thornton, J. M. (1996): Derivation of 3D coordinate templates for searching structural databases: application to Ser-His-Asp catalytic triads in the serine proteinases and lipases. *Protein Sci.*, **5,** 1001-13.

Wang, Z. and Quiocho, F. A. (1998): Complexes of adenosine deaminase with two potent inhibitors: X-ray structures in four independent molecules at pH of maximum activity. *Biochemistry,* **37,** 8314-24.

Yamashita, A., Maeda, K. and Maeda, Y. (2003): Crystal structure of CapZ: structural basis for actin filament barbed end capping. *Embo J.,* **22,** 1529-38.

Yufeng, W. and McDermott, A. E. (1999) In: *Modeling NMR Chemicals Shifts: Gaining Insights into Structure and Environment* (Eds, Facelli, J. C. and De Dios, A. C.) ACS Books, pp. 177-193.

Zhou, T., Radaev, S., Rosen, B. P. and Gatti, D. L. (2001): Conformational changes in four regions of the *Escherichia coli* ArsA ATPase link ATP hydrolysis to ion translocation. *J. Biol. Chem.,* **276,** 30414-22.

Zou, J., Hallberg, B. M., Bergfors, T., Oesch, F., Arand, M., Mowbray, S. L. and Jones, T. A. (2000): Structure of *Aspergillus niger* epoxide hydrolase at 1.8 Å resolution: implications for the structure and function of the mammalian microsomal class of epoxide hydrolases. *Structure,* **8,** 111-22.

ADDENDUM 1: Purification of AmyD

The α-amylase from *Drosophila melanogaster* has been produced and purified for crystallization in order to elucidate its 3D-structure by X-ray diffraction. This supplementary chapter describes the purification protocol used for AmyD and carried out subsequent to the production of this enzyme as described in chapter 2, § 2.2.10.

If not stated otherwise, all the purification steps were realized at a temperature of 4 °C.

Following centrifugation to separate the culture supernatant from the cells, the proteins of the supernatant were concentrated by **ammonium sulfate precipitation** at a concentration of 80 % saturation. After 1 hour of precipitation, the suspension was centrifuged at 13,000 g for 50 minutes and the pellet resuspended in 70 mL of buffer A (50 mM Tris, pH 7.5 containing 1 mM $CaCl_2$) and then **dialyzed** against 2x2 L of buffer B (20 mM HEPES, pH 7.5 containing 20 mM NaCl) for 24 hours.

At this point, the volume of the sample and the amylolytic activity (using the EPS assay as described in chapter 2, § 2.2.4) were measured so as to enable precipitation of the α-amylase by glycogen in ethanolic conditions (Schramm and Loyter, 1966). The calculation of the required amount of glycogen (glycogen to enzyme ratio w/w 4:1) was based on the measured amylolytic activity, with 50,000 EPS units corresponding to approximately 0.4 mg/mL of enzyme.

Cooled, pure **ethanol** was firstly added dropwise to this solution to a final concentration of 40 % and stored on ice with constant stirring for 40 minutes. The suspension was then centrifuged at 28,000 g for 10 minutes to eliminate the precipitate which mainly consisted of nucleic acids.

The calculated amount of **glycogen** dissolved in 10 mL of high quality water was added dropwise to the supernatant and the solution was stored on ice for 40 minutes with constant stirring. This suspension was then centrifuged at 20,000 g for 10 minutes. The residual amylolytic activity in the supernatant was determined and the addition of **glycogen** (glycogen to enzyme ratio w/w 4:1) was repeated.

Finally, the pooled pellets were resuspended in 100 mL of buffer A and **dialyzed** against 2x2 L of the same buffer for 24 hours.

The chromatography steps used were:

(i) **Q-Sepharose**

The solution was loaded onto this anion exchanger and eluted with a linear NaCl gradient (0-0.4 M; 500/500 mL) in the same buffer.

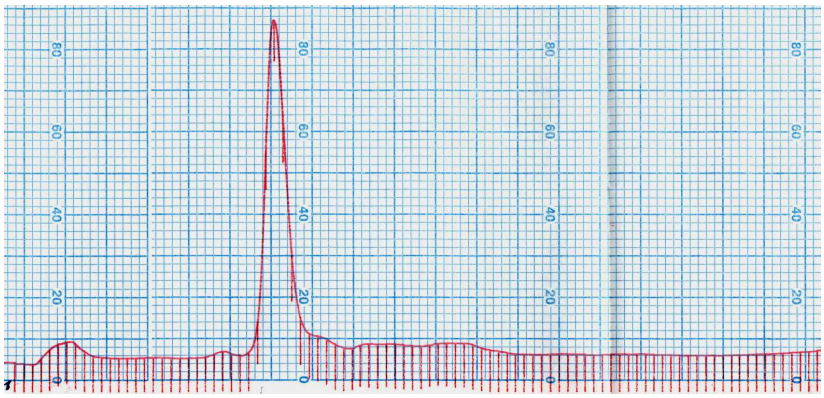


Fig. 29: Q-Sepharose chromatogram (A_{280}). The column (1.25x35 cm) was eluted with a linear NaCl gradient (0-0.4 M; 250/250 mL) in Tris buffer (as described above). Elution rate was 50 mL/h and the fraction volume was 7 mL.

The fractions displaying amylolytic activity were concentrated to 10 mL by ultrafiltration on a Millipore polyethersulfone membrane with a cutoff of 10,000 Da in an Amicon Ultrafiltration cell under a nitrogen pressure of 3 bars.

(ii) **Sephadex G100**

The concentrated sample was loaded onto this gel filtration matrix and eluted with 500 mL of buffer A. The pseudo-affinity of the α-amylase for this matrix was used to remove protein contaminants of a similar size.

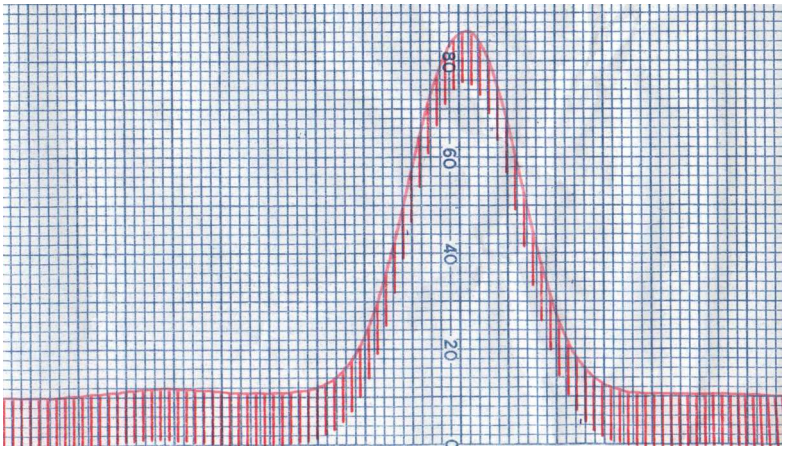


Fig. 30: Sephadex G100 chromatogram (A_{280}). The column (2.5x95 cm) was eluted with buffer A. Elution rate was 30 mL/h and the fraction volume was 3.8 mL.

Again, the fractions of interest were concentrated to 10 mL by ultrafiltration as described above.

(iii) **Ultrogel AcA54**

This final chromatographic step eliminates non-proteinacious contaminants such as salts and nucleic acids. This polishing step also allows the changing from Tris buffer to MOPS buffer (10 mM MOPS, 10 mM NaCl, 0.5 mM $CaCl_2$, pH 7.5), suitable for crystallization.

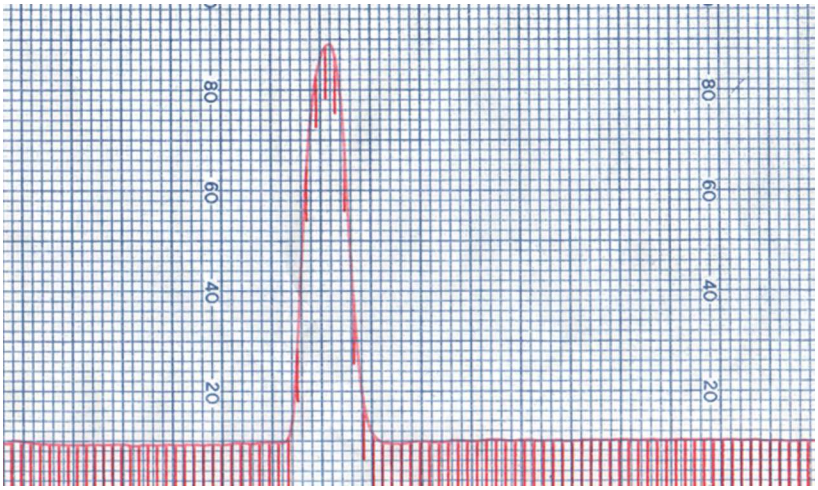


Fig. 31: Ultrogel AcA54 chromatogram (A_{280}). The column (2.5x95 cm) was eluted with MOPS buffer (as described above). Elution rate was 35 mL/h and the fraction volume was 4 mL.

The fractions of interest were pooled and concentrated to 20 mg/mL for crystallization.

This purification protocol yields approximately 50-60 mg of enzyme per liter of culture.

ADDENDUM 2: Abbreviation list (*ad* Fig.6)

This list of chloride-dependent α-amylases is taken from (D'Amico *et al.*, 2000) and provides the details about the abbreviations used in Fig. 6, as well as the corresponding accession codes in databanks.

	Species	Abbreviation	Accession No.
Mammals	Pig pancreatic	PPA	AF064742
	Mouse pancreatic	muspa	P00688
	Rat pancreatic	ratpa	P00689
	Mouse salivary	mussa	P00687
	Human carcinoma(2B)	hum2b	P19961
	Human pancreatic	HPA	P04746
	Human salivary	HSA	P04745
Bird	*Gallus gallus*	galga	Q98942
Acari	*Euroglyphus maynei*	eurma	AF144061
	Dermatophagoides pteronyssinus	derpt	AF144060
Insects	*Drosophila ananassae*	droan	G2253593
	Drosophila melanogaster	drome	G2252770
	Drosophila subobscura	drosu	G2252794
	Aedes aegypti	aedae	O02652
	Aedes atropalpus	aedat	Q16924
	Culex tarsalis	culta	Q23767
	Drosophila mauritiana	droma	P54215
	Drosophila simulans	drosi	Q24643
	Drosophila sechellia	drose	Q24642
	Drosophila teissieri	drote	Q24676
	Drosophila yakuba	droya	P51548
	Drosophila erecta	droer	Q23932
	Drosophila orena	droor	Q24610
	Drosophila pseudoobscura	drops	Q24609
	Drosophila virilis	drovi	Q24737
	Tenebrio molitor	TMA	S75702
	Tribolium castaneum	trica	P09107
	Anopheles merus	anome	U01210
	Anopheles gambiae	anoga	L04753
	Ostrinia nubilalis	ostnu	U04223
	Phaedon cochleariae	phaco	Y17902
	Lutzomyia longipalpis	lutlo	AF132512
Molluscs	*Crassostea gigas*	cragi	O02622
	Pecten maximus	pecma	P91778
Crustacean	*Penaeus vannamei*	penva	Q26193
Nematode	*Caenorhabditis elegans*	caeel	P91982
Bacteria	*Pseudomonas* sp. KFCC10818	pseudo	U40056
	Alteromonas haloplanctis	AHA	P29957
	Thermomonospora curvata	thecu	P29750

ADDENDUM 3: Jess results

The following table provides the complete list of identified triads with the algorithm Jess. The columns (from 1 to 6) correspond to the identified PDB code, the calculated rmsd and Δ values given in Å and the triad residues.

PDB#	rmsd (Å)	Δ (Å)	X	Ser	His
1A07	0,26	0,18	E162	S190	H204
1A08	0,41	0,15	E162	S190	H204
1A0L	0,09	0,13	D102	S195	H57
1A1B	0,39	0,18	E162	S190	H204
1A1C	0,39	0,16	E162	S190	H204
1A1E	0,21	0,15	E162	S190	H204
1A2K	0,17	0,43	Q71	S83	H73
1A4G	0,16	0,17	Q137	S439	H100
1A4L	0,26	0,17	N789	S791	H758
1A4M	0,2	0,18	N1789	S1791	H1758
1A4Q	0,15	0,13	Q137	S439	H100
1A5G	0,18	0,35	D102	S195	H57
1A7U	0,17	0,2	D228	S98	H257
1A81	0,3	0,18	E26	S54	H63
1A88	0,14	0,07	D226	S96	H255
1A8Q	0,19	0,22	D223	S94	H252
1A8R	0,18	0,18	GTP412	S112	H113
1AD8	0,2	0,3	D102	S195	H57
1AH2	0,24	0,17	H62	S215	H62
1AIX	0,19	0,36	D102	S195	H57
1AK4	0,11	0,13	D123	S99	H92
1AK9	0,08	0,09	D32	S221	H64
1AMR	0,24	0,17	D222	S139	H143
1AOT	0,42	0,188	E18	S46	H60
1AOU	0,17	0,17	E18	S46	H60
1AQH	***0***	***0***	***E19***	***S303***	***H337***
1AQM	0,05	0,08	E19	S303	H337
1AQN	0,14	0,19	D32	S221	H64
1ARB	0,21	0,19	D113	S194	H57
1ARG	0,26	0,15	D222	S139	H143
1ASG	0,18	0,14	D222	S139	H143
1ASL	0,29	0,19	D222	S139	H143
1ASM	0,26	0,13	D222	S139	H143
1AU9	0,19	0,28	D32	S221	H64
1B0I	0,09	0,19	E19	S303	H337
1B2Y	0,09	0,1	E27	S340	H386
1B8Y	0,34	0,16	T131	S133	H96
1B9S	0,25	0,16	E275	S295	H273
1BE8	0,13	0,18	D32	S221	H64
1BFN	0,2	0,19	D275	S464	H335
1BGL	0,15	0,28	D591	S971	H972
1BH6	0,19	0,41	D32	S221	H64
1BJE	0,24	0,13	Fe HEM 154	S92	H93
1BMA	0,19	0,14	D108	S203	H60
1BMM	0,18	0,37	D102	S195	H57
1BN8	0,12	0,16	D150	S18	H193
1BQ3	0,15	0,3	H8	S55	H181
1BRO	0,2	0,17	D228	S98	H257
1BRY	0,21	0,19	D1	S61	H51
1BSI	0,06	0,07	E27	S340	H386
1BU8	0,21	0,2	D176	S152	H263
1BSI	0,06	0,07	E27	S340	H386
1BVN	0,19	0,33	E27	S340	H386
1BYB	0,22	0,17	D275	S464	H335
1C1O	0,2	0,32	E102	S195	H57
1C3L	0,16	0,22	D32	S221	H64
1C4U	0,2	0,4	D102	S195	H57
1C5P	0,18	0,2	D102	S195	H57
1C5R	0,29	0,15	D102	S195	H57
1C5U	0,16	0,09	D102	S195	H57
1C7J	0,31	0,18	E310	S189	H399
1C8Q	0,1	0,23	E27	S340	H386
1C9T	0,24	0,16	D102	S195	H57

1CD3	0,37	0,13	S11	S74	H9
1CEX	0,14	0,17	H188	S120	H188
1CF2	0,22	0,18	E301	S302	H299
1CGI	0,31	0,2	D102	S195	H57
1CGJ	0,29	0,14	D102	S195	H57
1CIY	0,36	0,19	R265	S428	H429
1CL2	0,19	0,28	T148	S130	H103
1CLV	0,05	0,09	E27	S324	H362
1CM5	0,1	0,19	D74	S83	H84
1CNM	0,19	0,15	D39	S224	H69
1CPU	0,09	0,16	E27	S340	H386
1CQ7	0,18	0,15	D222	S139	H143
1CS0	0,58	0,17	D1041	S937	H1039
1CU7	0,2	0,46	D84	S177	H40
1CU9	0,18	0,18	D84	S177	H40
1CUB	0,2	0,32	D175	S120	H188
1CUC	0,18	0,24	D175	S120	H188
1CUS	0,19	0,13	D175	S120	H188
1CUU	0,18	0,23	D175	S120	H188
1CUV	0,19	0,23	D175	S120	H188
1CUW	0,17	0,24	D175	S120	H188
1CUX	0,18	0,27	D175	S120	H188
1CUY	0,14	0,15	D175	S120	H188
1CWD	0,2	0,26	E16	S44	H58
1D3O	0,27	0,12	H71	S26	H71
1D3P	0,18	0,32	D135	S241	H79
1D3Q	0,2	0,47	D135	S241	H79
1D3T	0,2	0,4	D135	S241	H79
1D4B	0,25	0,18	H43	S109	H43
1D8W	0,27	0,19	D231	S177	H229
1D9I	0,18	0,25	D135	S241	H79
1DBI	0,22	0,17	D39	S226	H72
1DE6	0,2	0,2	D231	S177	H229
1DEL	0,28	0,17	D218	S7	H194
1DHK	0,08	0,16	E27	S340	H386
1DMZ	0,22	0,16	Q666	S657	H651
1DT3	0,16	0,31	D201	S146	H258
1DT5	0,19	0,21	D201	S146	H258
1DTQ	0,4	0,15	T107	S191	H198
1DU4	0,14	0,2	D201	S146	H258
1DWP	0,24	0,11	D208	S80	H236
1DYL	0,19	0,3	D167	S219	H145
1E2S	0,2	0,26	E253	S250	H227
1E3D	0,2	0,16	H122	S532	H122
1E4C	0,23	0,19	N29	S31	H77
1E5Q	0,17	0,15	S99	S74	H81
1E5T	0,14	0,11	D641	S554	H680
1E6Q	0,2	0,39	Y295	S309	H228
1EA7	0,13	0,16	D34	S250	H71
1EB9	0,19	0,13	D208	S80	H236
1EG9	0,18	0,3	S310	S347	H261
1EGQ	0,17	0,29	D39	S224	H69
1EHO	0,12	0,1	E324	S197	H437
1EIN	0,07	0,15	D201	S146	H258
1ELE	0,2	0,17	D108	S203	H60
1EVE	0,11	0,16	E327	S200	H440
1EWU	0,24	0,14	H71	S26	H71
1F48	0,14	0,34	E416	S420	H148
1F6V	0,13	0,09	H51	S86	H51
1FFD	0,19	0,22	D175	S120	H188
1FFE	0,18	0,29	D175	S120	H188
1FIZ	0,2	0,34	D102	S195	H57
1FKX	0,22	0,16	N289	S291	H258
1FL1	0,18	0,17	H134	S114	H46
1FNI	0,2	0,24	D102	S195	H57
1FPC	0,25	0,18	D102	S195	H57
1FXT	0,2	0,19	H129	S136	H129
1FYR	0,2	0,11	E71	S98	H107
1G1Y	0,26	0,18	T522	S548	H527
1G94	0,05	0,07	E19	S303	H337
1G9H	0,05	0,13	E19	S303	H337
1GCI	0,16	0,21	D32	S221	H64
1GHV	0,18	0,37	D102	S195	H57
1GHZ	0,18	0,2	D102	S195	H57
1GI0	0,18	0,17	D102	S195	H57
1GI3	0,26	0,14	D102	S195	H57
1GI5	0,27	0,15	D102	S195	H57
1GI9	0,32	0,18	D102	S195	H57
1GIU	0,22	0,18	D1	S61	H51

1GJ5	0,17	0,19	D102	S195	H57
1GJB	0,29	0,18	D102	S195	H57
1GS5	0,19	0,26	N248	S33	H34
1GSJ	0,15	0,27	N248	S33	H34
1GU6	0,12	0,14	Q378	S255	H351
1GVZ	0,19	0,39	D102	S195	H57
1GXH	0,24	0,17	W73	S47	H45
1H16	0,1	0,15	D74	S83	H84
1H17	0,09	0,18	D74	S83	H84
1H18	0,11	0,12	D74	S83	H84
1H2W	0,16	0,17	P74	K82	K82
1H2X	0,18	0,2	P74	K82	K82
1H46	0,18	0,28	D74	S43	H42
1H9O	0,21	0,09	E23	S48	H57
1HCG	0,17	0,29	D102	S195	H57
1HCT	0,16	0,17	H201	S158	H201
1HJ9	0,54	0,19	Y234	S93	H91
1HNY	0,12	0,18	E27	S340	H386
1HPL	0,14	0,16	D176	S152	H263
1HQ5	0,32	0,18	H228	S271	H111
1HQG	0,24	0,14	H228	S271	H111
1HQH	0,33	0,2	H228	S271	H111
1HX0	0,05	0,12	E27	S340	H386
1HXH	0,05	0,07	S180	S141	H182
1HYL	0,1	0,18	D102	S195	H57
1I6W	0,19	0,13	D133	S77	H156
1IBX	0,29	0,19	H32	S28	H32
1IC6	0,14	0,19	D39	S224	H69
1IHU	0,07	0,13	E416	S420	H148
1II9	0,15	0,14	E416	S420	H148
1IML	0,25	0,16	H42	S18	H42
1INF	0,2	0,09	Q138	S440	H101
1IOI	0,16	0,17	E79	S142	H166
1IOW	0,17	0,18	S22	S274	H63
1IQF	0,3	0,12	D102	S195	H57
1IZN	0,31	0,2	E221	S171	H152
1J6Y	0,2	0,18	H114	S111	H114
1JAE	0,08	0,12	E27	S324	H362
1JB4	0,28	0,16	Q71	S83	H73
1JB5	0,18	0,16	Q71	S83	H73
1JD7	0,09	0,16	E19	S303	H337
1JD9	0,11	0,26	E19	S303	H337
1JDI	0,19	0,26	N28	S30	H80
1JFH	0,08	0,13	E27	S340	H386
1JFR	0,15	0,19	D177	S131	H209
1JGA	0,19	0,14	H437	S197	H437
1JGB	0,14	0,16	F327	S197	H437
1JKM	0,15	0,25	D308	S202	H338
1JQ7	0,15	0,25	E1251	S1432	H1457
1JU3	0,18	0,28	Y391	S117	H287
1JXJ	0,09	0,19	E27	S340	H386
1JXK	0,1	0,2	E27	S340	H386
1JY1	0,24	0,2	E415	S419	H372
1JYR	0,27	0,15	E71	S98	H107
1JYW	0,18	0,2	HET 1472001	S998	H540
1JYX	0,19	0,17	HET O147	S998	H540
1K0W	0,18	0,14	N28	S30	H80
1K1O	0,3	0,2	D102	S195	H57
1K8Q	0,12	0,16	D324	S153	H353
1KB3	0,11	0,17	E27	S340	H386
1KBB	0,05	0,07	E27	S340	H386
1KBK	0,03	0,07	E27	S340	H386
1KCJ	0,15	0,1	E324	S197	H437
1KGU	0,08	0,16	E27	S340	H386
1KGW	0,09	0,15	E27	S340	H386
1KGX	0,15	0,21	E27	S340	H386
1KU6	0,25	0,18	E334	S203	H447
1KW0	0,17	0,31	D151	S273	H271
1KXH	0,04	0,06	E19	S303	H337
1KXQ	0,05	0,11	E27	S340	H386
1KXT	0,07	0,16	E27	S340	H386
1KXV	0,04	0,09	E27	S340	H386
1L0P	0,04	0,04	E19	S303	H337
1L3G	0,47	0,2	I25	S21	H38
1L5V	0,19	0,31	T134	S132	H60
1LBS	0,17	0,24	D187	S105	H224
1LI1	0,13	0,17	S142	S163	H121
1LIZ	0,26	0,17	E116	S94	H120
1LKK	0,26	0,13	E138	S166	H180

1LKL	0,24	0,1	E138	S166	H180
1LLF	0,11	0,17	H1449	S1209	H1449
1LNS	0,18	0,14	D468	S348	H498
1LW6	0,2	0,25	D32	S224	H64
1M33	0,2	0,23	D207	S82	H235
1M3D	0,11	0,19	S51	S31	H10
1M4N	0,15	0,2	H226	S287	H266
1MAX	0,16	0,31	D102	S195	H57
1MEE	0,18	0,23	D32	S221	H64
1MFU	0,1	0,16	E27	S340	H386
1MFV	0,07	0,12	E27	S340	H386
1MGN	0,16	0,31	Fe HEM 154	S92	H93
1MIR	0,19	0,3	N219	S29	H199
1MLA	0,1	0,08	T293	S295	H271
1MMK	0,17	0,38	D151	S273	H271
1MMT	0,17	0,31	D151	S273	H271
1MN6	0,17	0,15	D176	S148	H268
1MNQ	0,19	0,09	D176	S148	H268
1MOA	0,19	0,48	Fe HEM 154	S92	H93
1MPH	0,09	0,13	K8	S20	H9
1MPX	0,22	0,11	D307	S174	H340
1MRJ	0,23	0,18	D1	S61	H51
1MT3	0,2	0,12	D244	S105	H271
1MTW	0,17	0,12	D102	S195	H57
1MU0	0,24	0,17	D244	S105	H271
1MU6	0,2	0,32	D102	S195	H57
1MX1	0,11	0,09	E2354	S2221	H2468
1MX5	0,2	0,32	E1354	S1221	H1468
1N2K	0,19	0,38	E253	S250	H227
1N3J	0,13	0,14	H78	S104	H78
1N5Y	0,47	0,16	T107	S191	H198
1NDQ	0,17	0,25	D32	S221	H64
1NDU	0,18	0,34	D32	S221	H64
1NJT	0,2	0,17	H1157	S1132	H1063
1NJU	0,17	0,3	H457	S432	H363
1NKG	0,18	0,27	E136	S367	H365
1NKK	0,16	0,26	H1157	S1132	H1063
1NKM	0,19	0,42	H157	S132	H63
1NM6	0,19	0,21	D102	S195	H57
1NMN	0,18	0,36	Y138	S132	H94
1NMW	0,2	0,26	I159	T143	P149
1NSB	0,14	0,16	Q137	S439	H100
1NTP	0,16	0,2	H91	S93	H91
1NU8	0,11	0,19	D708	S630	H740
1O2G	0,33	0,15	D102	S195	H57
1O2H	0,33	0,2	D102	S195	H57
1O2I	0,17	0,15	D102	S195	H57
1O2O	0,17	0,14	D102	S195	H57
1O3B	0,29	0,19	D102	S195	H57
1O3C	0,27	0,18	D102	S195	H57
1O3P	0,3	0,19	D102	S195	H57
1O42	0,26	0,15	E18	S46	H60
1O46	0,35	0,18	E18	S46	H60
1O4G	0,3	0,2	E18	S46	H60
1O4I	0,31	0,19	E18	S46	H60
1O4L	0,31	0,11	E18	S46	H60
1O4N	0,28	0,14	E18	S46	H60
1O5D	0,13	0,19	D102	S195	H57
1O5F	0,26	0,19	D102	S195	H57
1O7G	0,16	0,38	S310	S347	H261
1O7N	0,18	0,44	S310	S347	H261
1O7W	0,18	0,44	S310	S347	H261
1OCA	0,16	0,16	D123	S99	H92
1ODO	0,16	0,18	E20	S318	H291
1ODS	0,18	0,28	D269	S181	H298
1OGY	0,28	0,1	Fe HEM	S76	H79
1OH9	0,14	0,15	N248	S33	H34
1OIL	0,2	0,33	E264	S87	H286
1OPK	0,27	0,18	E157	S183	H192
1OPL	0,27	0,18	E157	S183	H192
1ORV	0,17	0,13	D708	S630	H740
1ORW	0,14	0,16	D708	S630	H740
1OSE	0,1	0,23	E27	S340	H386
1OXM	0,13	0,17	D175	S120	H188
1OYT	0,2	0,19	D102	S195	H57
1OZF	0,07	0,14	S443	S445	H469
1OZG	0,1	0,16	S443	S445	H469
1OZH	0,1	0,12	S443	S445	H469
1P13	0,29	0,12	E172	S200	H214

1P3C	0,16	0,16	D97	S171	H47
1P7V	0,14	0,14	D39	S224	H69
1P7W	0,14	0,15	D39	S224	H69
1P8M	0,27	0,11	H228	S271	H111
1P8N	0,19	0,16	H228	S271	H111
1P8O	0,29	0,1	H228	S271	H111
1P8R	0,3	0,14	H228	S271	H111
1PEK	0,17	0,09	D39	S224	H69
1PFR	0,19	0,27	E238	S114	H241
1PFX	0,17	2	S100	S93	H101
1PIF	0,07	0,2	E27	S340	H386
1PIG	0,09	0,17	E27	S340	H386
1PIJ	0,43	0,18	R26	S67	H24
1PIY	0,17	0,3	E238	S114	H241
1PJ6	0,19	0,27	Y259	S271	H225
1PMB	0,2	0,47	HEM154	S92	H93
1PMM	0,2	0,18	S269	S271	H241
1PN5	0,15	0,1	G149	S151	H150
1PPI	0,15	0,29	E27	S340	H386
1Q0Z	0,17	0,25	D248	S102	H276
1Q35	0,23	0,2	D205	S278	H267
1Q3X	0,15	0,14	D532	S633	H483
1Q45	0,36	0,18	E184	S341	H279
1Q4N	0,08	0,13	E27	S340	H386
1QA0	0,2	0,35	D90	S183	H46
1QAD	0,26	0,15	E22	S47	H56
1QE3	0,22	0,09	E310	S189	H399
1QFM	0,17	0,17	D641	S554	H680
1QHM	0,13	0,17	D1074	S1083	H1084
1QIE	0,22	0,19	E327	S200	H440
1QIF	0,21	0,19	E327	S200	H440
1QIG	0,18	0,09	E327	S200	H440
1QLW	0,17	0,28	E230	S206	H298
1QNO	0,17	0,32	D195	S158	H198
1QNQ	0,12	0,16	D195	S158	H198
1QNR	0,14	0,21	D195	S158	H198
1QO7	0,2	0,31	Q86	S148	H97
1QQU	0,18	0,2	D102	S195	H57
1QWK	0,14	0,19	N157	S174	H187
1QZ3	0,19	0,25	D252	S155	H282
1QZQ	0,29	0,19	E415	S419	H372
1R1D	0,19	0,25	D192	S 93	H222
1R1P	0,2	0,1	E71	S97	H106
1R1Q	0,25	0,19	E71	S97	H106
1R1S	0,08	0,05	E71	S97	H106
1R8W	0,26	0,18	D312	S7	H368
1R9D	0,26	0,2	D312	S7	H368
1R9M	0,11	0,08	D708	S630	H740
1RKX	0,26	0,18	S61	S40	H17
1RLA	0,24	0,17	H228	S271	H111
1S01	0,16	0,21	D32	S221	H64
1S6H	0,16	0,31	D102	S195	H57
1S83	0,19	0,43	D102	S195	H57
1SC9	0,2	0,19	E207	S80	H235
1SCA	0,17	0,28	D32	S221	H64
1SCD	0,18	0,2	D32	S221	H64
1SCN	0,2	0,37	D32	S221	H64
1SHC	0,29	0,18	H27	S29	H27
1SI0	0,21	0,15	D205	S278	H267
1SLV	0,17	0,18	D102	S195	H57
1SLX	0,26	0,19	D102	S195	H57
1SMD	0,07	0,1	E27	S340	H386
1SPR	0,25	0,18	E16	S44	H58
1SPS	0,22	0,2	E16	S44	H58
1SQT	0,2	0,23	D96	S197	H45
1SR3	0,26	0,17	E116	S94	H120
1SUE	0,13	0,08	D32	S221	H64
1SVN	0,19	0,24	D32	S221	H64
1SVR	0,09	0,16	H165	S167	H165
1T60	0,12	0,17	S142	S163	H121
1T61	0,13	0,18	S142	S163	H121
1TA2	0,2	0,24	D102	S195	H57
1TAH	0,18	0,18	D263	S87	H285
1TBN	0,14	0,18	H117	S146	H117
1TBO	0,19	0,17	H140	S120	H140
1TCC	0,2	0,41	D187	S105	H224
1THM	0,11	0,16	D38	S225	H71
1TIB	0,2	0,25	D201	S146	H258
1TK3	0,16	0,15	D708	S630	H740

1TME	0,19	0,32	S119	S240	H122
1TMQ	0,17	0,2	E27	S324	H362
1TNG	0,19	0,41	D102	S195	H57
1TNH	0,17	0,18	D102	S195	H57
1TNN	0,22	0,19	H46	S65	H46
1TOM	0,08	0,17	D102	S195	H57
1TS3	0,17	0,17	Q136	S111	H141
1U2Y	0,06	0,14	E27	S340	H386
1U30	0,04	0,11	E27	S340	H386
1U33	0,08	0,19	E27	S340	H386
1U43	0,19	0,33	D8	S35	H42
1U8E	0,16	0,2	D708	S630	H740
1UA3	0,08	0,16	E27	S340	H386
1UFO	0,15	0,11	D183	S113	H217
1UKP	0,18	0,1	D275	S464	H335
1UTM	0,2	0,27	D102	S195	H57
1UTN	0,17	0,15	D102	S195	H57
1UTO	0,53	0,17	Y234	S93	H91
1UTP	0,21	0,18	D102	S195	H57
1UTQ	0,2	0,25	D102	S195	H57
1UVT	0,07	0,09	D102	S195	H57
1UWB	0,16	0,39	T107	S191	H198
1UWK	0,19	0,35	H501	S487	H444
1V5L	0,19	0,19	H66	S22	H66
1VCQ	0,19	0,3	D167	S219	H145
1VFR	0,19	0,12	E54	S81	H82
1VH4	0,13	0,14	E187	S243	H215
1VK9	0,17	0,15	S94	S96	H74
1VLQ	0,15	0,24	D274	S188	H303
1VZ2	0,13	0,09	D641	S554	H680
1VZ3	0,14	0,09	D641	S554	H680
1W15	0,33	0,16	H404	S400	H420
1WMD	0,14	0,22	D30	S255	H68
1WME	0,14	0,17	D30	S255	H68
1WMF	0,16	0,19	D30	S255	H68
1XNT	0,1	0,2	D59	S57	H150
1XZC	0,18	0,25	D175	S120	H188
1XZH	0,22	0,19	D175	S120	H188
1XZI	0,17	0,19	D175	S120	H188
1XZK	0,12	0,17	D175	S120	H188
1XZL	0,19	0,17	D175	S120	H188
1XZM	0,17	0,16	D175	S120	H188
1YJB	0,15	0,33	D32	S221	H64
1ZFP	0,25	0,12	E71	S98	H107
2BSP	0,11	0,1	D150	S18	H193
2CBP	0,32	0,18	C79	S87	H84
2CPU	0,1	0,15	E27	S340	H386
2CUT	0,15	0,18	D175	S120	H188
2HLC	0,19	0,22	D102	S195	H57
2HMY	0,16	0,35	E47	S294	H292
2PFL	0,1	0,15	D74	S83	H84
2PKC	0,15	0,24	D39	S224	H69
2PRK	0,14	0,21	D39	S224	H69
2RGF	0,13	0,16	H49	S18	H49
2RMC	0,2	0,19	H126	S133	H126
2SHP	0,28	0,16	E17	S44	H53
2SNI	0,15	0,16	D32	S221	H64
2TEC	0,18	0,17	D38	S225	H71
2YAS	0,19	0,19	D207	S80	H235
3CPU	0,11	0,18	E27	S340	H386
3GCT	0,19	0,36	D102	S195	H57
3PFL	0,08	0,17	D74	S83	H84
3TEC	0,15	0,17	D38	S225	H71
5CHA	0,19	0,36	D102	S195	H57
5RLA	0,22	0,16	H228	S271	H111
5YAS	0,18	0,28	D207	S80	H235
8AAT	0,29	0,18	D222	S139	H143