

Gerhard Hofer BSc

**The X-Ray Structure Solution  
of the  
Allergenic Amylases  
Der p 4 and Tri a bA**

**MASTER'S THESIS**

to achieve the university degree of  
Master of Science

Master's degree programme: Biochemistry and Molecular Biomedical Sciences

submitted to

**Graz University of Technology**

Supervisor

Professor Dr. Walter Keller

Institute of Molecular Biosciences

**AFFIDAVIT**

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly indicated all material which has been quoted either literally or by content from the sources used. The text document uploaded to TUGRAZonline is identical to the present master's thesis dissertation.

---

Date

---

Signature

---

## Contents

Acknowledgements.....	v
Introduction .....	1
Allergies.....	1
Definition .....	1
Prevalence.....	2
Current treatment.....	3
Dust mite allergy .....	3
Wheat allergy.....	4
$\alpha$ and $\beta$ Amylases .....	4
$\alpha$ -amylases - Der p 4 .....	4
$\beta$ -amylases - Tri a bA.....	4
Methods.....	6
Expression .....	6
Der p 4.....	6
Tri a bA .....	6
Purification.....	6
Der p 4.....	6
Tri a bA .....	7
SDS-PAGE .....	7
Activity .....	7
Tri a bA .....	7
CD-spectroscopy .....	7
Thermofluor .....	8
Crystallization.....	8
Der p 4.....	8
Tri a bA .....	8
X-ray analysis .....	9

---

Der p 4.....	9
Tri a bA .....	9
Structure refinement .....	10
Der p 4.....	10
Tri a bA .....	10
Epitope prediction .....	10
Results.....	11
Sequence Analysis.....	11
Der p 4.....	12
Tri a bA .....	14
Expression .....	16
Purification.....	19
Tri a bA .....	21
Activity .....	22
Tri a bA .....	22
Thermofluor .....	23
CD-spectroscopy .....	24
Crystallization and Crystal Analysis.....	25
Der p 4.....	25
Tri a bA .....	27
Structure .....	28
Der p 4.....	28
Tri a bA .....	37
Discussion.....	44
Table of Figures.....	46
Bibliography .....	48
Supplemental.....	52

---

## Acknowledgements

This work was supported by the FWF SFB project grant F46 project 04.

I would like to thank Michael Bogdos for flawlessly executing several of the lab tasks of the Tri a bA project (including the Thermofluor and activity experiments) under my supervision in the course of a summer internship.

While this work only shows personal results, the proteins described herein were already under investigation by S. Vrtala's group (Der p 4) and R. Valenta's group (Tri a bA) who have kindly provided the expression plasmids of their respective allergens for this research.

I would also like to acknowledge the people who have made this work possible.

Prof. Walter Keller, for providing me with the opportunity to work on these and many more fascinating projects of his with a near endless supply of freedom and trust.

My parents, who have supported my less than direct approach to higher education with unceasing patience.

And finally my coworkers, foremost Kerstin Fauland, Markus Eder, Judith Wortmann and Michael Bogdos who have advised, amused and motivated me to such a degree, that taking several years to actually write a master's thesis seemed like a good idea.

Thank you.

## Introduction

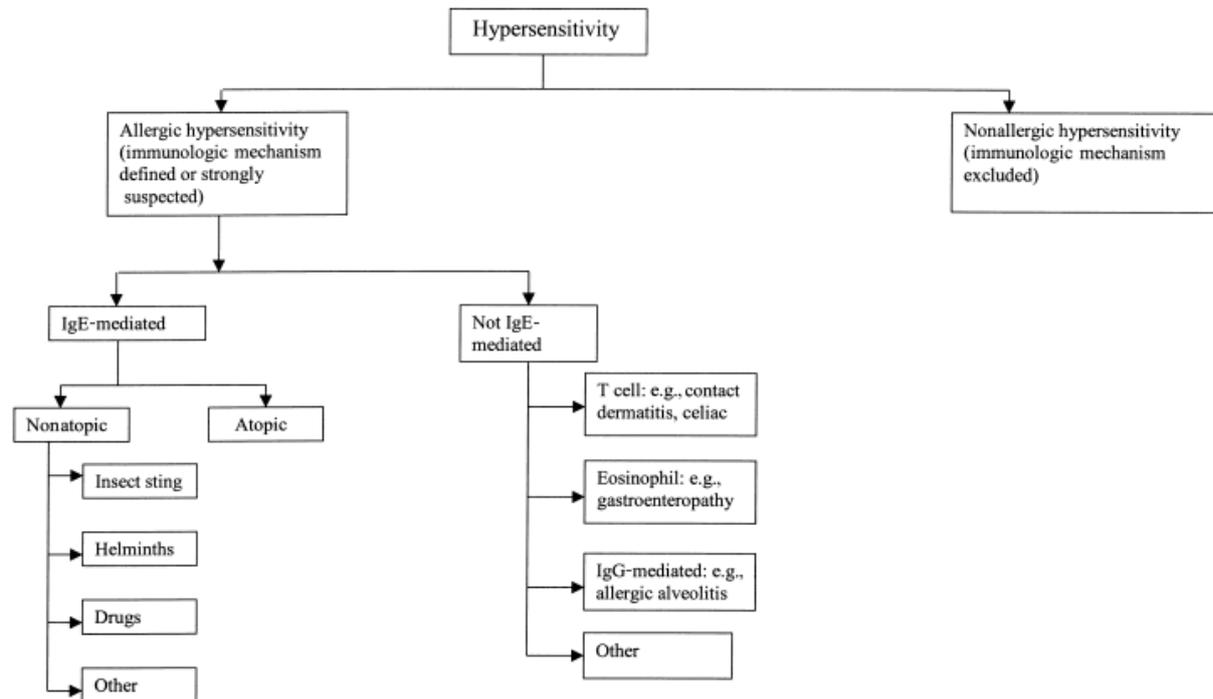
While focusing on the expression, purification, crystallization and structure elucidation of two proteins, this work should be viewed in the context in which it was written. That is to say, from the perspective of a researcher, who is part of a large collection of research groups, working together to further our understanding of the causes of and possible cures to allergies.

As such, the proteins described herein have a medical importance in that they cause allergic reactions in people sensitized to them. This introduction therefore aims to give a very short background about allergens, allergies and why we study them.

## Allergies

### Definition

While there are other forms of hypersensitivity (Figure 1)(1), this work focuses solely on the type I allergies, also called IgE mediated allergies. As this name implies, allergic patients produce IgE antibodies against the allergens in question. Instead of sensibly removing the foreign proteins, as IgG-tagging would, IgE recognition causes basophil release and mast cell degranulation upon contact with the allergen. This reaction causes the well-known symptoms, which range from rhinitis to potentially fatal anaphylactic shock, depending on the severity and location of the reaction.

**Allergy**

Volume 56, Issue 9, pages 813-824

DOI: 10.1111/j.1398-9995.2001.00002.x-i1

Figure 1: Overview of hypersensitivities (1)

## Prevalence

Allergies have been on the rise over the past 30 years (2). There is a strong correlation between the 'western lifestyle' and prevalence for sensitization although no theory has yet been able to give a satisfactory mechanism of causation (3), though many have tried. In 2007 a study comparing children living in a polish city with children living in the nearby polish countryside found that 23% of the rural children were sensitized to at least one allergen, while 64% of the children in the city reacted to at least one allergen in a prick test (4).

A 2014 meta study (5) found a lifetime prevalence of 17% for becoming allergic to food in Europe. The effect this has on the health of the population and the cost to the health system is hard to overestimate.

Even a severe symptom like anaphylaxis for example has an estimated lifetime prevalence between 0.05 % to 2% (6).

Allergy is far from a niche disorder and all trends point towards it becoming more relevant as it is spreading not only in western countries but also in developing countries with their rising industrialization.

## Current treatment

The most common treatment of allergies is one of avoidance. Trying to limit the exposure to the allergen can reduce or remove the symptoms. Given that most allergen sources are hard to avoid, like pollen or house dust, this is usually only step one of the treatment.

The symptoms can be treated by H1-antihistamines, corticosteroids, mast cell stabilizers or anticholinergics, either taken preemptively or applied to the affected area to relieve the symptoms. In severe cases Omalizumab, a humanized monoclonal antibody that blocks the binding of free IgE to both FcεRI and FcεRII, can be injected.

The only treatment targeting the root of the problem currently available is immunotherapy. This involves subjecting the patient to the relevant allergens at levels below the critical dose, or modified to avoid IgE recognition, either by injection or by sublingual application. Over a course of years this treatment has been proven to reduce the symptoms in allergic patients for many common allergies, although the exact mechanism by which this improvement is achieved is still under investigation.

## Dust mite allergy

Being by far one of the most common, the allergy against house and storage dust mites is widespread in Austria both in cities and in the countryside (7). While the subspecies vary, mites can be found living off detritus in human housing and as such provide a source for allergens all year round. The European house dust mite *Dermatophagoides pteronyssinus* breeds especially well in beds, as the body heat and moisture found there counteract the normally too low humidity found in European homes, especially in winter. They feed on organic detritus and excrete their feces in the form of miniscule pellets. The small size of these pellets allow them to get airborne easily, causing every movement in the bed to release a cloud of inhalable allergen containing particles.

The recombinant digestive  $\alpha$ -amylase of *Dermatophagoides pteronyssinus* (Der p 4) was found to bind IgE of 18% of house dust mite allergic people tested (Y. Resch et al, unpublished).

## Wheat allergy

Not to be confused with gluten intolerance, type I wheat allergy has been reported by 0.4% of the population as a self-diagnosed and doctor confirmed condition (8). A study from 2007 showed that at the age of ten 9% of children tested showed anti wheat IgE in their sera (9). With wheat being a staple food in wide parts of the world, even a low percentage of people becoming allergic to it results in a high patient count. Since the majority of wheat allergic patients are cross-reactive to other plants, allergen avoidance means removal of nearly all forms of grain from one's diet.

The wheat  $\beta$ -amylase (Tri a bA) has only recently been found to be allergenic (10), and shows IgE reactivity in 40% of wheat allergic patients tested (S. Pahr et al, unpublished).

## $\alpha$ and $\beta$ Amylases

As a last introductory part, I would like to give a short overview of the chemical significance of the proteins in question. As amylases (E.C.3.2.1) without  $\alpha$ -1,6 activity, both of them hydrolyze the  $\alpha$ -D-(1-4) glycosidic bonds of linear  $\alpha$ -D-(1-4) glycosidically bound polyglucose (i.e. starch). The difference lies in their substrate binding and therefore in the product generated.

### $\alpha$ -amylases - Der p 4

1,4-alpha-D-glucan glucohydrolases (E.C.3.2.1.1) are endo-amylases, meaning that they digest starch by hydrolyzing the chain at any point that leaves an overhang on both sides. Complete digestion therefore creates a mixture of glucose, maltose and maltotriose, provided there are no  $\alpha$ -1,6-linkages, which will not be hydrolyzed. (11)

### $\beta$ -amylases - Tri a bA

1,4-alpha-D-glucan maltohydrolases (E.C.3.2.1.2) successively remove glucose dimers (i.e.  $\beta$ -maltose) from the reducing end of starch, producing a pure  $\beta$ -maltose solution, otherwise only containing the indigestible parts of its substrate.

The protein fold of these two proteins in question is also similar to some degree, the barrel shaped  $(\beta/\alpha)_8$  super structure (TIM-barrel) they share is obvious at first glance. This however is where the similarities end. Whereas Der p 4 has two extra domains, is dependent on  $\text{Ca}^{2+}$  and  $\text{Cl}^-$  for its activity and needs all four possible disulfide bonds to be formed, Tri a bA requires all cysteines to be in their reduced state for its enzymatic activity.

## Methods

### Expression

#### Der p 4

The proteins were expressed in SHuffle strain *E. coli* cells from a pET-17b vector (see supplemental for construct information). Chemically transformed cells were plated on ampicillin agar. Overnight cultures were started by picking single colonies and were grown in 2% glucose LB medium at 30°C. Main cultures were inoculated by adding ONC to a final OD<sub>600</sub> of 0.1 into LB medium containing 0.5% glucose. The expression of Der p 4 was not started as usual with addition of IPTG, instead the culture was grown to an OD<sub>600</sub> of 0.6 to 0.8 before being cooled to 16°C and left shaking overnight.

#### Tri a bA

Essentially as Der p 4 with the following exceptions. Expression was induced by adding IPTG to a final concentration of 0.4 mM when the main cultures reached an OD<sub>600</sub> of 0.6-0.8. Cultures were then either left at 30°C for 4 hours or cooled to 16°C and shaken overnight.

### Purification

#### Der p 4

The *E. coli* pellets were sonicated for lysis in a 100 mM MES pH 6.5, 300 mM NaCl, 5 mM CaCl<sub>2</sub> and 50mM imidazole buffer. The soluble fraction after centrifugation at 15000g for 30 min was applied to an Immobilized Metal Affinity Chromatography (IMAC) column (HisTrap 5ml FF). After washing with 10 column volumes of lysis/washing buffer, the protein was eluted with the same buffer with an increased imidazole concentration (300 mM).

The protein was then further purified by bringing the solution to 40% (v/v) ethanol, centrifuging off the precipitated proteins and buffer changing the supernatant to 50 mM MES pH 6.5, 50 mM NaCl using a micro centrifugation filter.

## Tri a bA

Purification followed the optimized cleanup protocol established for Der p 4 without CaCl<sub>2</sub>. The final buffer was set to pH 6 instead of pH 6.5.

## SDS-PAGE

SDS-PAGEs are either 12 % or 18 % polyacrylamide Tris-glycine and were run at 300 V for 30 to 40 minutes. Visualization was achieved by Coomassie blue staining followed by de-staining in boiling water.

## Activity

### Tri a bA

0.5 µl of the eluate or 5 µl flow-through of the HisTrap purification were mixed with 500 µl of 0.5 % starch in 25 mM MES pH 6 and 20mM NaCl at 25°C. 50 µl samples of these reaction were taken and the reaction stopped by the addition of 50 µl of dinitrosalicylic acid colour reagent (12) and heated to 95°C for 5 minutes. The absorption at 540 nm was measured on a nanodrop-1000, adjusted for t = 0 absorption and graphed as relative absorption to the maximum archived at the end of the reaction.

## CD-spectroscopy

Solutions containing 0.5 mg/ml Tri a bA in 25 mM MES pH 6 and 25 mM NaCl were analyzed in a Jasco J-715 spectropolarimeter at room temperature. The spectra were measured from 260 to 190 nm at 50 nm/min with a response time of 2 sec and 0.2 nm data pitch.

Secondary structure prediction was performed using Dichroweb (13) using the CDSSTR method (14) and datasets 4 and 7 (15).

## Thermofluor

The Fluorescence Thermal Shift Assay (Thermofluor) was set up with triplicates of 5  $\mu$ l Tri a bA solution (5 mg/ml in 50 mM MES pH 6, 50 mM NaCl) mixed with 5  $\mu$ l of SYPRO® orange (pre-diluted 1 : 1000 in pure H<sub>2</sub>O) and 15  $\mu$ l of a multi component buffer system consisting of a molar ratio of 1 : 2 : 2 of L-malic acid, MES and Tris (16) to set the pH without changing the chemical composition of the samples. This mixture gives a nearly linear pH response when two solutions of different pH are mixed (e.g. 1:1 mixing a 1 M pH 6 and a 1 M pH 8 solution of this buffer will give a 1 M buffer very close to pH 7). Varying ratios of 100 mM pH 4 and 100 mM pH 9 buffer stocks were used to sample this pH range.

## Crystallization

### Der p 4

The protein solution (in MES pH 6.5, 150 mM NaCl, 5 mM CaCl<sub>2</sub>) in concentrations 1 to 5 mg/ml was mixed 0.5  $\mu$ l + 0.5  $\mu$ l with the following screens in 96 well microbatch plates and covered with a mixture of 3 parts paraffin to 1 part silicone oil.

Morpheus

Index

JCSG+

PEG/ION

Optimization of the condition JCSG+ F10 ( 1.1 M Na malonate pH 7.0, 0.1 M HEPES pH 7, 0.5 % Jeffamine ED-2001 pH 7.0 ) was performed in 24 well sitting drop setups with 5  $\mu$ l protein solution mixed with 5  $\mu$ l of reservoir solution.

### Tri a bA

For screening the protein solution (in MES pH 6, 50 mM NaCl) in concentrations from 2 to 4 mg/ml was mixed 0.5  $\mu$ l + 0.5  $\mu$ l with the JCSG+ screen in 96 well microbatch plates and covered with a mixture of 3 parts paraffin to 1 part silicone oil.

The condition H3 (0.1 M bis-Tris pH 5.5, 25% PEG 3350) was optimized and repeat crystallization was achieved with 25% PEG 3350 100mM bis-Tris pH 5.0.

### X-ray analysis

#### Der p 4

Diffraction data was collected at the European Synchrotron Radiation Facility ESRF.

#### Tri a bA

Diffraction data was collected at the Elettra synchrotron Trieste.

## Structure refinement

Molecular graphics and analyses were performed with the UCSF Chimera package. Chimera is developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIGMS P41-GM103311), Pymol (The PyMOL Molecular Graphics System, Version 1.7.1 Schrödinger, LLC.) and *Coot* (17).

### Der p 4

The diffraction data was integrated using *iMosflm* (18), scaled and merged using *scala* of the *ccp 4* program suite (19). Molecular replacement was achieved using a *Phyre2* (20) model of Der p 4 (based on the human salivary amylase) in the *Phenix* (Python-based Hierarchical ENvironment for Integrated Xtallography (21)) package, which was also used for refinement, with alternating manual adaptation in *Coot* (17).

### Tri a bA

The dataset 'Tri a bA ice' was treated the same as above with the *pdb* model 2XFR (22) of the barley  $\beta$ -amylase used for molecular replacement. 'Tri a bA twinned' was integrated with *XDS* (23) instead of *Mosflm* and the model of 'Tri a bA ice' was used for molecular replacement.

### Epitope prediction

Epitope prediction was carried out by the *EPSVR* server, "an antigenic Epitope Prediction method by using Support Vector Regression (*EPSVR*) with six attributes: residue epitope propensity, conservation score, side chain energy score, contact number, surface planarity score, and secondary structure composition. "(24)

The returned values were visualized by colouring a 50% transparent surface representation with a rainbow spectrum giving red as zero and blue as highest probability.

## Results

### Sequence Analysis

The primary sequences were analyzed for their calculated molecular mass, theoretical pI and the calculated extinction coefficient (to be used in the 280 nm absorption concentration determination), using the ExpASys ProtParam webserver (25). Models for molecular replacement were found using the Basic Local Alignment Search Tool (26) (blastp suite) and multiple sequence alignments were performed by the espresso sub-setting (27) of the T-coffee webserver (28).

## Der p 4

Accession number: Q9Y197

Number of amino acids: 496

Molecular weight: 57149.9 g/mol

Theoretical pl: 6.70

```

      10      20      30      40      50      60
KYHNPFIGN RSVITHLMEW KYDDIGDECE RFLGPYGYGG VQVSPVNEHA ILDRRPWYER

      70      80      90     100     110     120
YQPVSYDIRT RSGDEQQFRR MVKRCNKAGV RIYVDIVLNH MTGAQSGKGT NGHHDGNTL

     130     140     150     160     170     180
QYPGVFPGPN DFHGHESCPT QDLEIHDYTN PKEARNCRSL GLRDLKQQSE YVRQKQVDFL

     190     200     210     220     230     240
NHLIDIGVAG FRSDASTHQW PDDLRSIYSR LHNLNKEFFP ENSQPFIYHE TIYYGGNGIN

     250     260     270     280     290     300
SNEYTSLGRI IEFRFYKEIT NVFRGNNPLH WLKNFGTEWG LVPSGDALVM IDSHDLRVGH

     310     320     330     340     350     360
TGKLGFNINC FEGRLLKAAT AFMLAWNYGV PRVMSSYFWN QIIKDGKDVN DWVGPSPDKN

     370     380     390     400     410     420
GNILSVHPNP DMTCNHEWIC EHRWREIYNM VKFRMIAGQE PVHNWWDNGD YQIAFSRGNR

     430     440     450     460     470     480
AFIAINLQKN QQNLQQKLHT GLPAGTYCDI ISGNLIDNKC TGKSIHVDKN GQADVYVGHD

     490
EFDAFVAYHI GARIVS

```

Ala (A) 20	4.0%	Leu (L) 30	6.0%
Arg (R) 30	6.0%	Lys (K) 22	4.4%
Asn (N) 40	8.1%	Met (M) 9	1.8%
Asp (D) 33	6.7%	Phe (F) 23	4.6%
Cys (C) 9	1.8%	Pro (P) 22	4.4%
Gln (Q) 22	4.4%	Ser (S) 25	5.0%
Glu (E) 23	4.6%	Thr (T) 18	3.6%
Gly (G) 45	9.1%	Trp (W) 12	2.4%
His (H) 25	5.0%	Tyr (Y) 25	5.0%
Ile (I) 34	6.9%	Val (V) 29	5.8%

Ext. coefficient 103750

Abs 0.1% (=1 g/l) 1.815, assuming all pairs of Cys residues form cystines



## Tri a bA

Accession number: P93594

Number of amino acids: 503

Molecular weight: 56611.1

Theoretical pl: 5.24

```

      10      20      30      40      50      60
MAGNMLANYV QVYVMLPLDV VSVDNKFEKG DEIRAQLKKL TEAGVDGVM I DVWVGLVEGK

      70      80      90     100     110     120
GPKAYDWSAY KQVFDLVHEA GLKLQAIMSF HQCGGNVGDV VNIPIPIQWVR DVGATDPDIF

     130     140     150     160     170     180
YTNRGGTRNI EYLTLGVDDQ PLFHGRTAVQ MYADYMASFR ENMKKFLDAG TIVDIEVGLG

     190     200     210     220     230     240
PAGEMRYPST PQSQGWVFPG IGEFICYDKY LEADFKAAAÄ KAGHPEWELP DDAGEYNDTP

     250     260     270     280     290     300
EKTQFFKDNQ TYLTEKGKFF LSWYSNKLIK HGDKILDEAN KVFLGCRVQL AIKISGIHWW

     310     320     330     340     350     360
YRVPNHAAEL TAGYYNLDDR DGYRTIARML TRHHASMNFT CAEMRDSEQS EEAKSAPEEL

     370     380     390     400     410     420
VQQVLSAGWR EGLHVACENA LGRYDATAYN TILRNARPKG INKNGPPEHK LFGFTYLRLS

     430     440     450     460     470     480
NELLEGQNYA TFQTFVEKMH ANLGHDPDVD PVAPLERSKP EMPIEMILKA AQPKLEPFPF

     490     500
DKNTDLPVKD HTDVGDEVLV APV

```

Ala (A) 43	8.5%	Leu (L) 41	8.2%
Arg (R) 20	4.0%	Lys (K) 32	6.4%
Asn (N) 24	4.8%	Met (M) 15	3.0%
Asp (D) 36	7.2%	Phe (F) 22	4.4%
Cys (C) 5	1.0%	Pro (P) 28	5.6%
Gln (Q) 18	3.6%	Ser (S) 17	3.4%
Glu (E) 35	7.0%	Thr (T) 22	4.4%
Gly (G) 41	8.2%	Trp (W) 10	2.0%
His (H) 14	2.8%	Tyr (Y) 23	4.6%
Ile (I) 21	4.2%	Val (V) 36	7.2%

Ext. coefficient 89270

Abs 0.1% (=1 g/l) 1.577, assuming all Cys residues are reduced

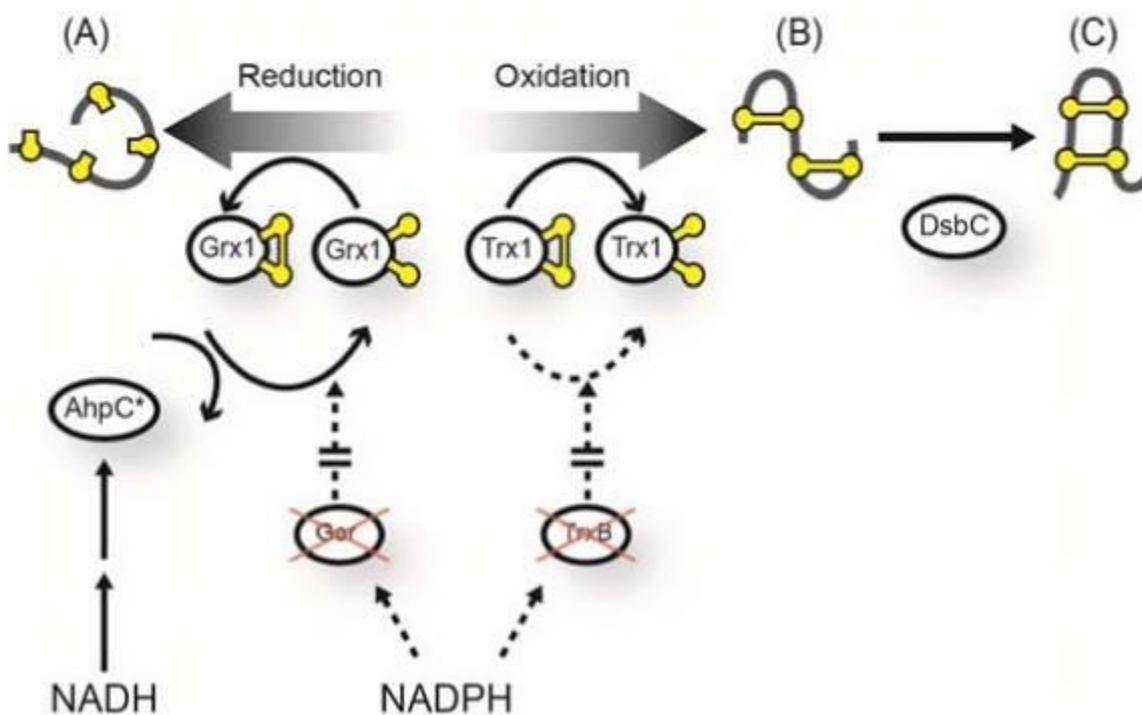


## Expression

Expression of eukaryotic proteins can require a relatively complex approach compared to bacterial proteins. Posttranslational modifications require either the use of eukaryotic expression systems, or specialized bacterial hosts, usually coupled with protocols deviating from the standards.

Neither Der p 4, nor Tri a bA are heavily glycosylated in their native forms. They do however contain several cysteines, which require disulfide bonding for Der p 4, while needing to retain their reduced state in Tri a bA. Both of these conditions have been met using the SHuffle strain (29) of *E. coli* as an expression system.

SHuffle is a special *E. coli* strain with (amongst others) the  $\Delta trxB$ ,  $\Delta gor$  mutations. These mutations of the thioredoxin reductase (*trxB*) and glutathione reductase (*gor*) cause the cytosol of *E. coli* to form and retain disulfide bonds, removing the need to secrete proteins to the periplasm to achieve disulfide bonding. Since most proteins contain more than two cysteines, bonds between the wrong thiol groups are possible and in many cases likely. Therefore cell compartments used to form disulfide bonds contain isomerases specialized on opening and, after allowing for rearrangement, closing disulfide bonds. To achieve this effect, the SHuffle strain is additionally modified to express a copy of DsbC (the *E. coli* periplasmic disulfide bond isomerase) chromosomally without its usual periplasmic targeting sequence, retaining it in the cytosol and allowing disulfide shuffling there, hence the name of this particular strain.



Microb Cell Fact. 2012; 11: 56.

Published online 2012 May 8.

doi: 10.1186/1475-2859-11-56

Figure 4: "Disulfide bond formation in the cytoplasm of SHuffle. Schematic diagram of the redox pathways in the cytoplasm of SHuffle. Dotted lines represent disabled protein interactions due to the deletion of *trxB* and *gor*. Redox state of cysteines (yellow balls) are indicated (oxidized = ball + stick; reduced = ball). (A) Protein is reduced by Grx1 or oxidized by Trx1. (B) Mis-oxidized protein is isomerized to its native correctly folded state (C) by DsbC." (29)

This strain was able to express Der p 4 in its active fold, but only at low expression levels. Even then most of the expressed proteins accumulated in insoluble inclusion bodies, but 2.5 mg of active protein per liter of LB medium were achieved.

Tweaking the expression of the T7 based plasmid to low enough levels required to forego the normal induction with IPTG. Even the basal expression had to be reduced by adding glucose to the medium to avoid inclusion body formation.

The depletion of glucose causes the activation of the adenyl cyclase. This causes high levels of cAMP which binds to the catabolite activator protein (CAP) causing it to bind to the promoter sequence of the lac operon. This does not remove the lactose repressor, but this repressors control is not tight enough to completely suppress the expression of T7 polymerase. The low amount of T7 polymerase thus produced turned out to give high enough expression levels to see a strong overexpression band on an SDS-PAGE, while producing some Der p 4 in its soluble and active form.

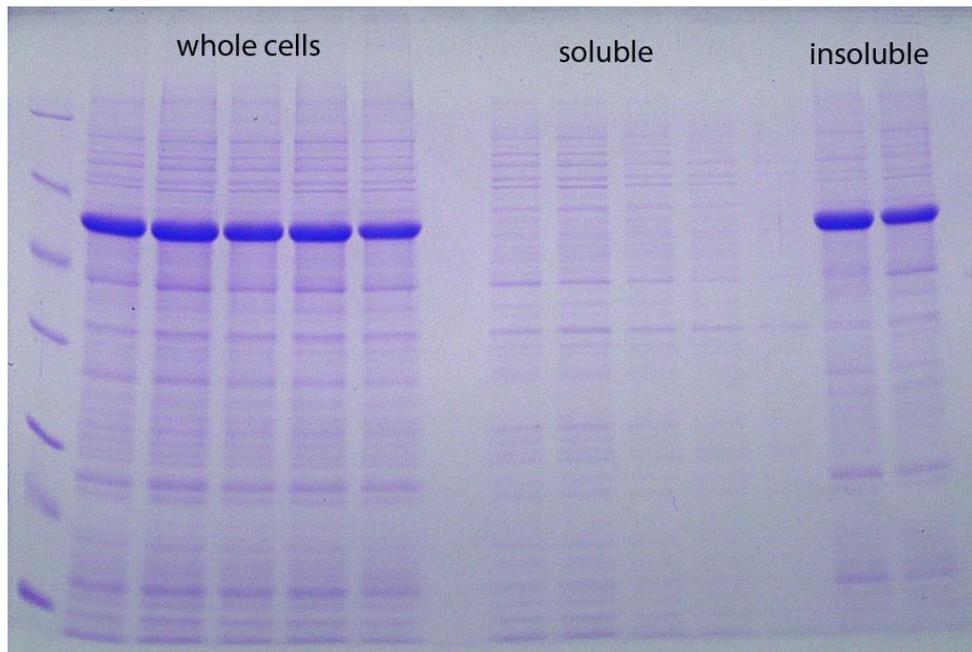


Figure 5: Initial attempt at expressing Der p 4 in BL21 strain *E. coli* resulting in insoluble protein aggregates

## Purification

A two-step purification process was established that gave clean amylase samples for both proteins. First a standard IMAC cleanup removed most other proteins, but as usual retained the *E. coli* proteins prone to binding to an IMAC column. All of the Der p 4 found to be soluble and bind to the IMAC column was also shown to be uniformly disulfide bonded by SDS-PAGE analysis (Figure 6).

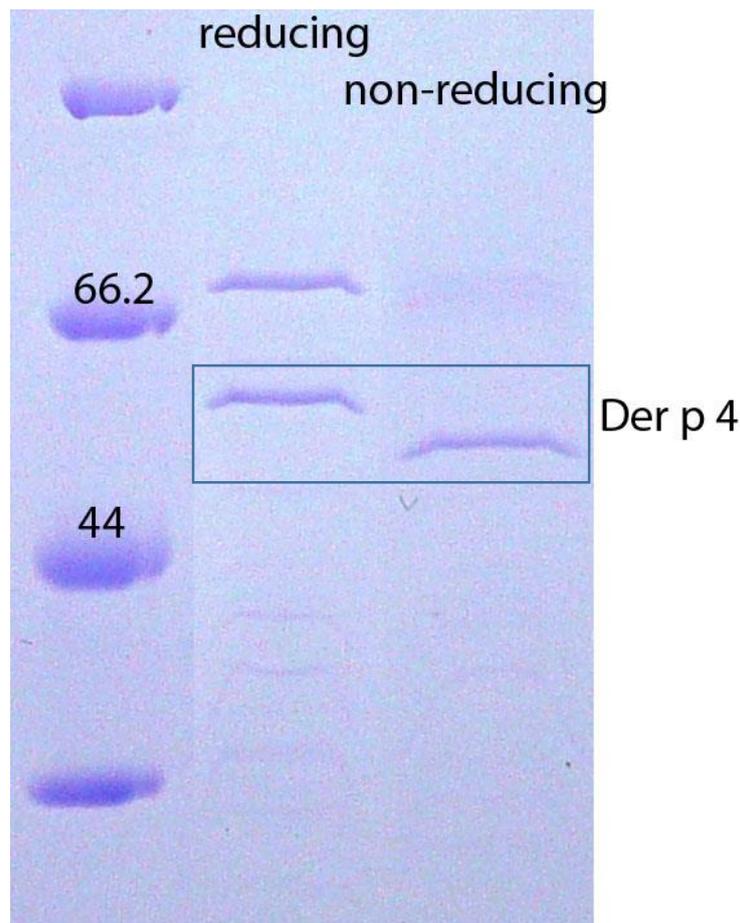
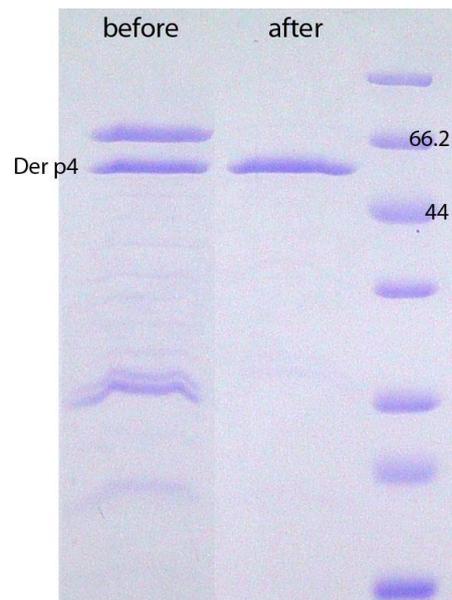


Figure 6: Soluble Der p 4 shows uniform disulfide bonding, causing a single band of equal density to travel further in a non-reducing SDS PAGE

The second cleanup step, bringing the solution to 40 % ethanol to precipitate other proteins, is a shortened purification protocol for specifically cleaning up amylases. That protocol (30) involves bringing the sample to 40 % ethanol on ice, adding glycogen (which will precipitate in ethanol) and centrifuging the resulting amylase-glycogen complex. While this is an effective way of purifying amylases, it failed to deliver pure Der p 4 from an *E. coli* lysate, presumably due to bacterial amylases binding to glycogen as well. Combined with the IMAC step however,

the full glycogen purification was not even necessary, as all proteins co-eluting from the His-Trap column precipitated at the introduction of ethanol to the solution, yielding clean protein at this stage (Figure 7). Stopping the protocol there also prevented the otherwise inevitable introduction of glycogen and its degradation products to the sample.



**Figure 7: The ethanol purification step yields pure protein from the IMAC purified sample**

## Tri a bA

While the expression level of soluble Tri a bA was a lot higher than for Der p 4, a high percentage of the soluble and HisTrap cleanable Tri a bA was found to be oxidized (Figure 8 lane 3). All of the intermolecularly as well as most of the intramolecularly disulfide bound amylase was removed by the ethanol precipitation step. The percentage of mis-oxidized protein varied from expression to expression and seems to correlate with expression speed. Likely faster expression produces more misfolded amylase, which then oxidizes due to the cysteines being exposed. Once cleaned, natively folded Tri a bA was highly resistant to oxidization during storage at 4°C.

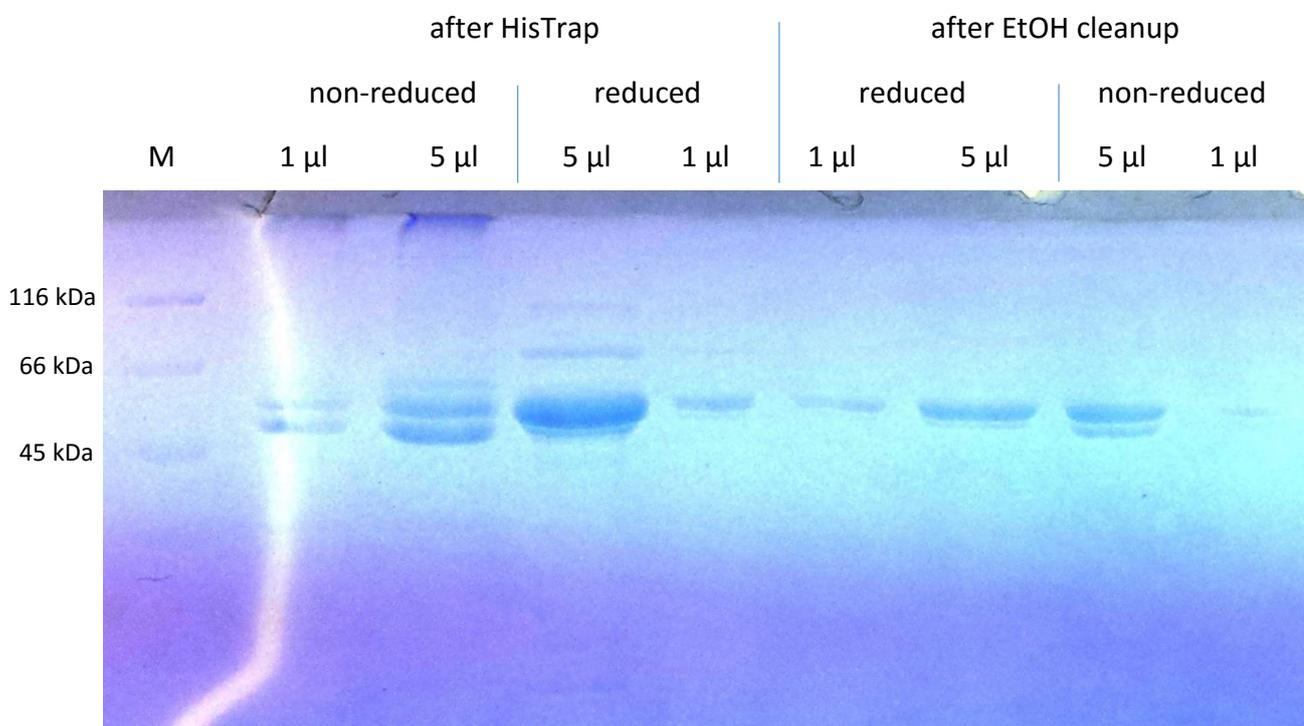


Figure 8: Tri a bA oxidation state after HisTrap and ethanol cleanup

## Activity

### Tri a bA

The activity of the His-Trap accessible amylase was established by a simple starch hydrolysis assay basically as in (12) whereby the new reducing carbohydrate ends created by starch hydrolysis are reacted with dinitrosalicylic acid, the product of which can then be quantified by its absorption at 540 nm.

This method provided a quick method of establishing which fractions of a cleanup contained the highest concentration of active enzyme by revealing activity within minutes and, in its most simple form does not even require photometric measurement, as it produces a clearly visible colour change from yellow to red.

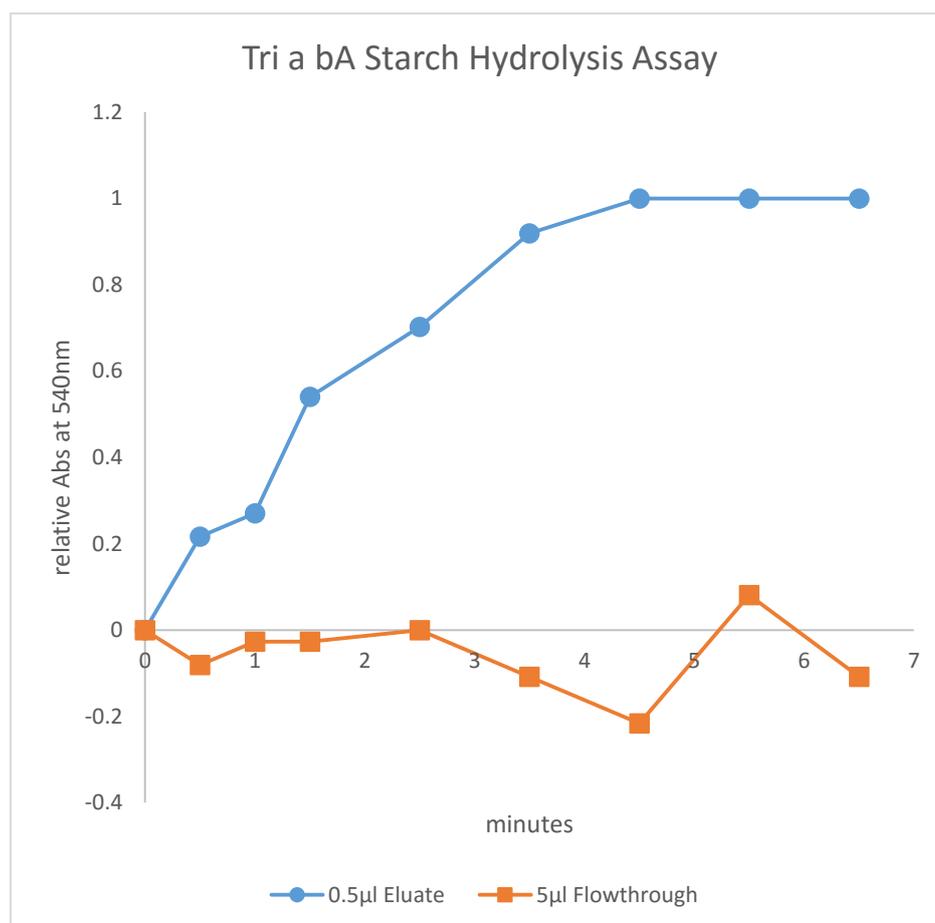


Figure 9: Tri a bA starch hydrolysis assay showing amylase activity in the His-Trap eluate

## Thermofluor

A Fluorescence Thermal Shift Assay (31) (Thermofluor) was used to find the optimal storage and crystallization pH for Tri a bA, using a multi component buffer system (16). The highest melting temperature and therefore highest enthalpic stability was found to be around pH 5.5 which is congruous with the reported optimal activity pH ranges of other eukaryotic  $\beta$ -amylases (32). This pH gives an increased melting temperature of nearly 6 degrees compared to the commonly used pH 7.4, for convenience pH 6 was kept as the standard buffer for further experiments.

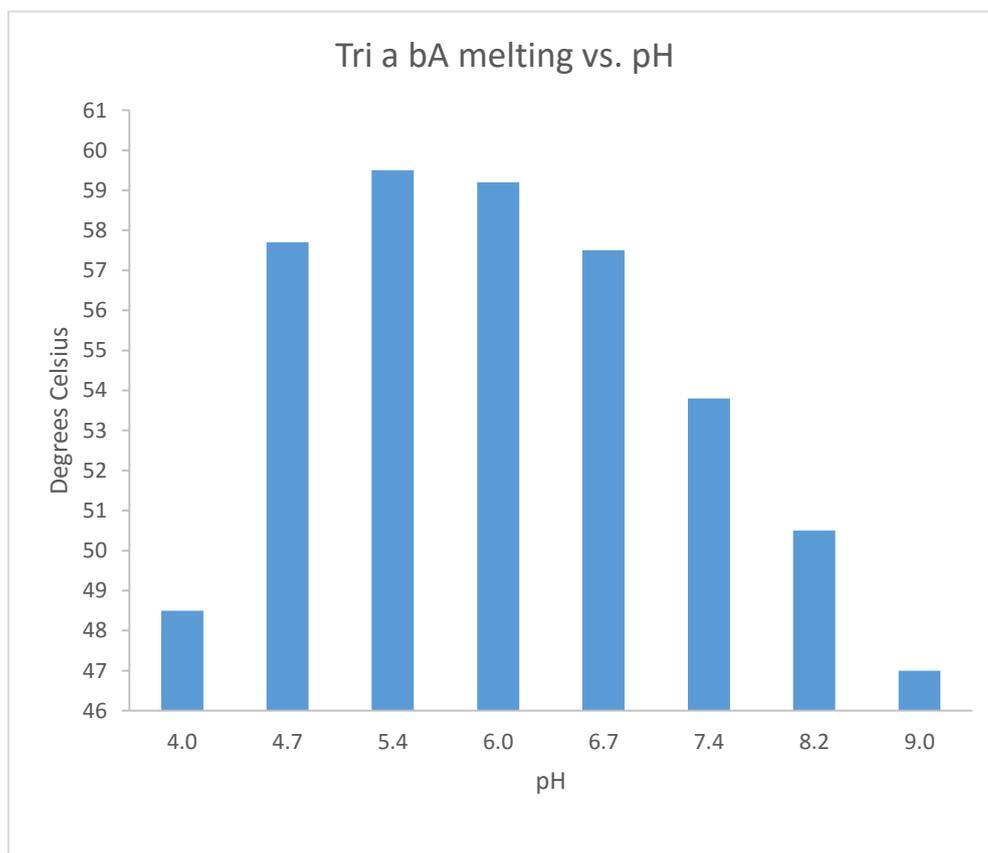


Figure 10: Tri a bA Thermofluor melting results showing highest thermo-stability at pH 5.4

## CD-spectroscopy

The room temperature CD spectrum showed a mostly  $\alpha$  helical fold with some  $\beta$  sheet content. While storage at 4°C for extended periods did not result in any visible or SDS-PAGE discernible changes to the sample, the CD spectrum changed slightly.

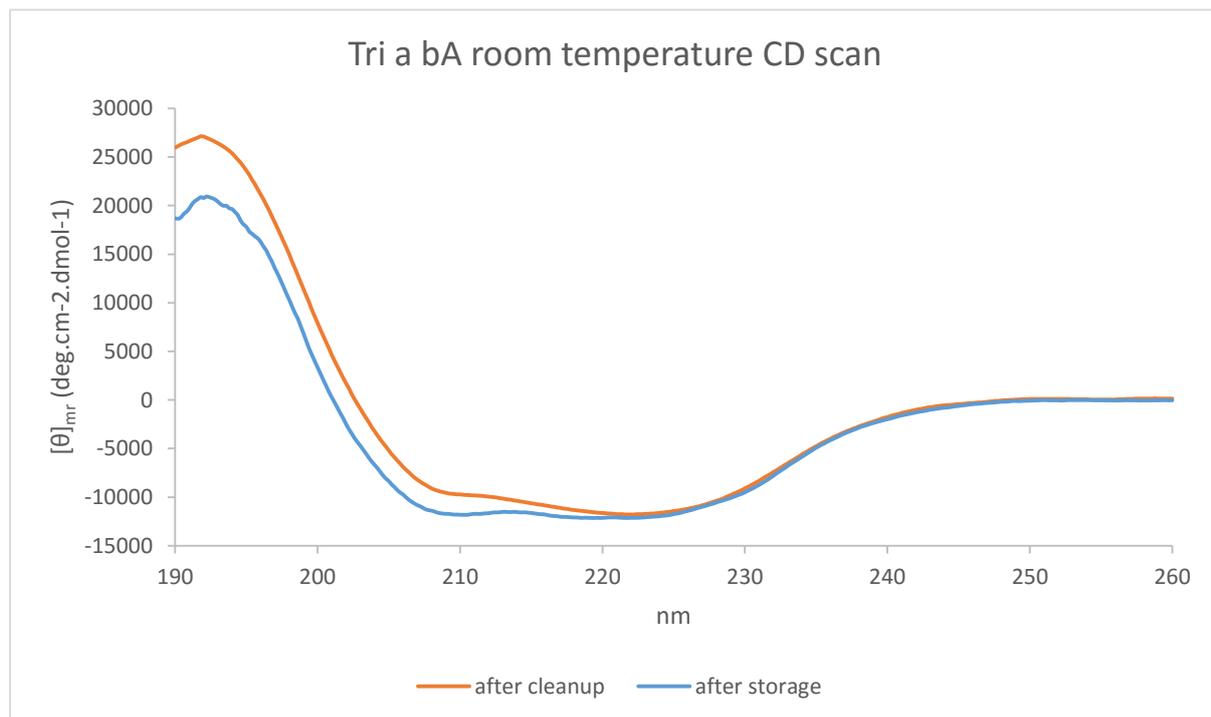


Figure 11: Comparison between freshly purified Tri a bA and a sample stores for 6 months at 4°C

The comparison of these two spectra using the CDSSTR method (14) of Dichroweb (13) using either reference set 4 or 7 was inconclusive (15).

set 4	Helix	Sheet	Turn	Unordered
Tri a bA fresh	38%	18%	18%	26%
Tri a bA stored	39%	16%	18%	27%

set 7	Helix	Sheet	Turn	Unordered
Tri a bA fresh	41%	18%	17%	24%
Tri a bA stored	40%	13%	15%	32%

## Crystallization and Crystal Analysis

### Der p 4

Der p 4 grew needle bundles in most screen conditions containing high concentrations of organic acid salts after two to three weeks. These did not diffract, but optimization screens yielded one drop containing several large single needles, all of which diffracted well at the home source (3.5 Å and better) (Figure 12).

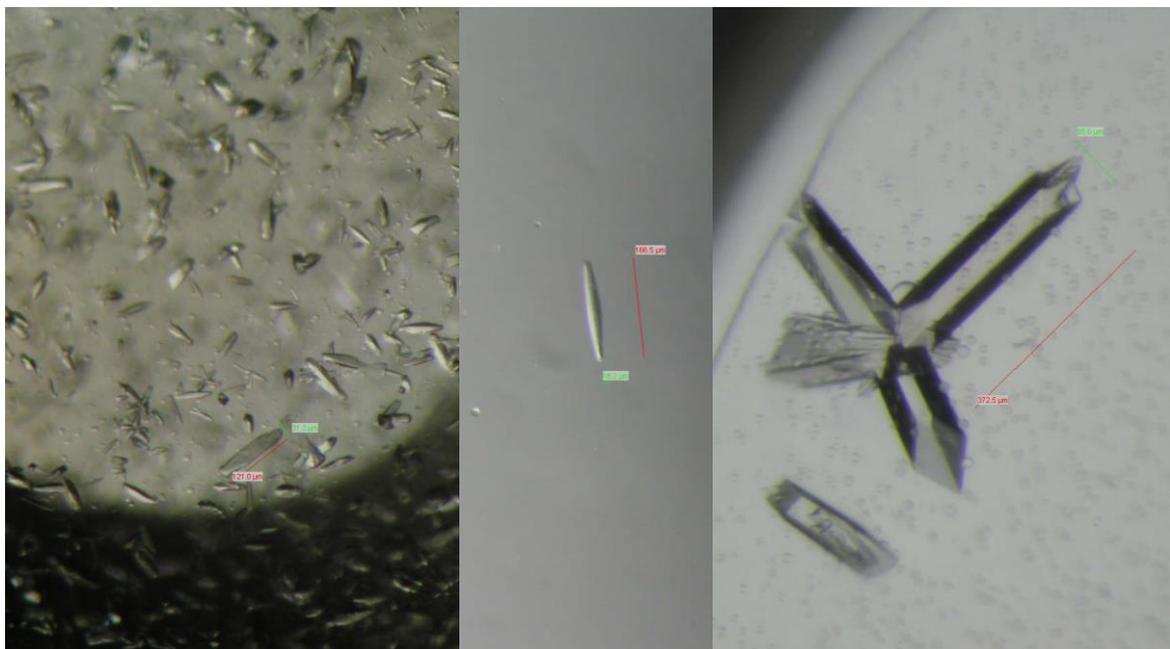


Figure 12: Optimization of Der p 4 crystals. From initial screen and first optimization success to well diffracting crystals

One of the crystals showed smeared spots and a lower diffraction limit than the others upon initial freezing. A 30 second annealing step while mounted on the beamline (blocking the cryo-stream with a piece of paper) caused dramatic improvement in this crystal (Figure 14).

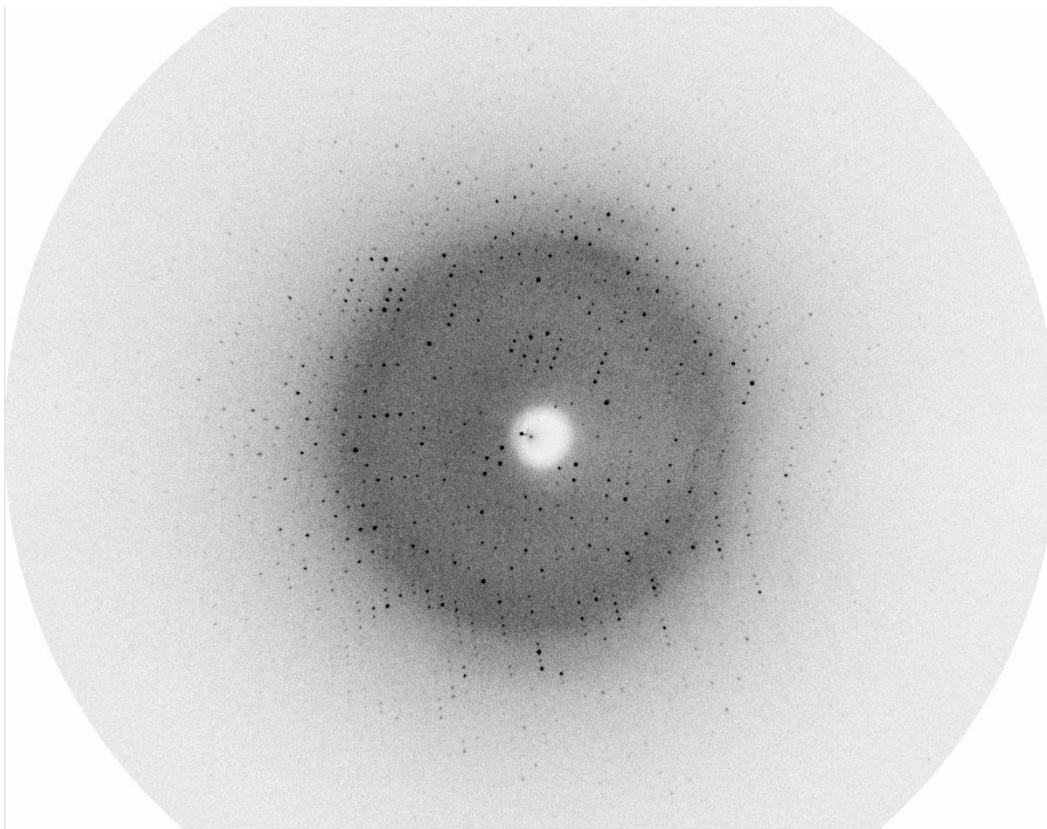


Figure 13: Der p 4 crystal diffraction at home source

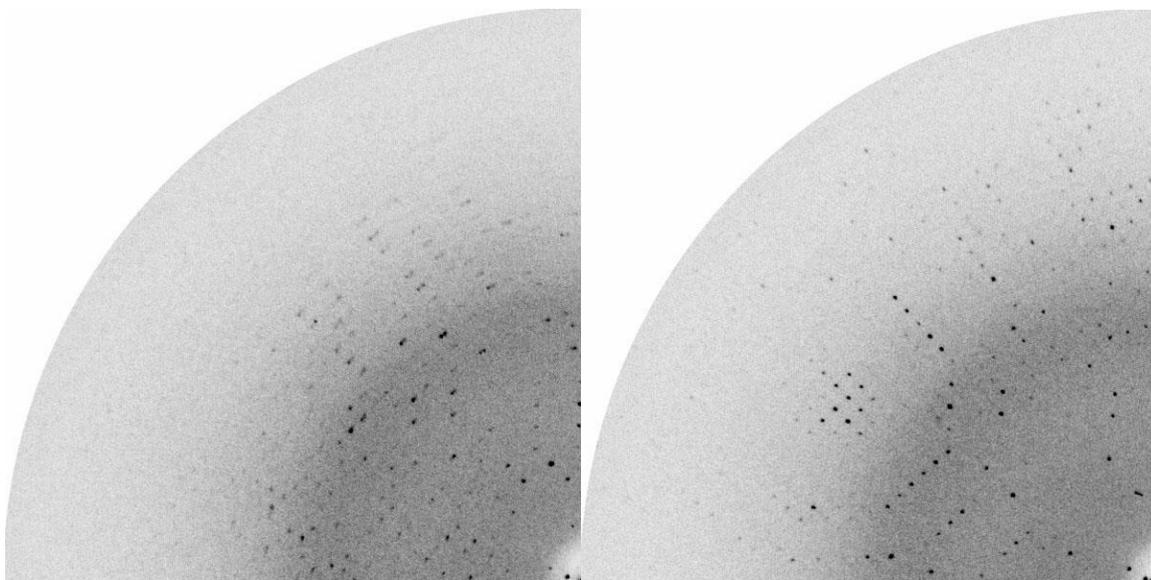


Figure 14: Der p 4 crystal diffraction before and after annealing

## Tri a bA

Crystallized quickly and reproducibly from its optimized buffer in 25% PEG 3350 100 mM bis-Tris pH 5.0. Crystal growth could be observed within several hours of the crystallization set up. Twinning was observed in most crystals. The structure was solved using a crystal that showed 9% merohedral twinning with a twin law of  $h, -k, -l$  (named 'Tri a bA twin'), as well as one crystal showing no twinning, but severe ice rings (named Tri a bA ice) and a slightly lower diffraction limit (2.0 Å compared to 1.8 Å).

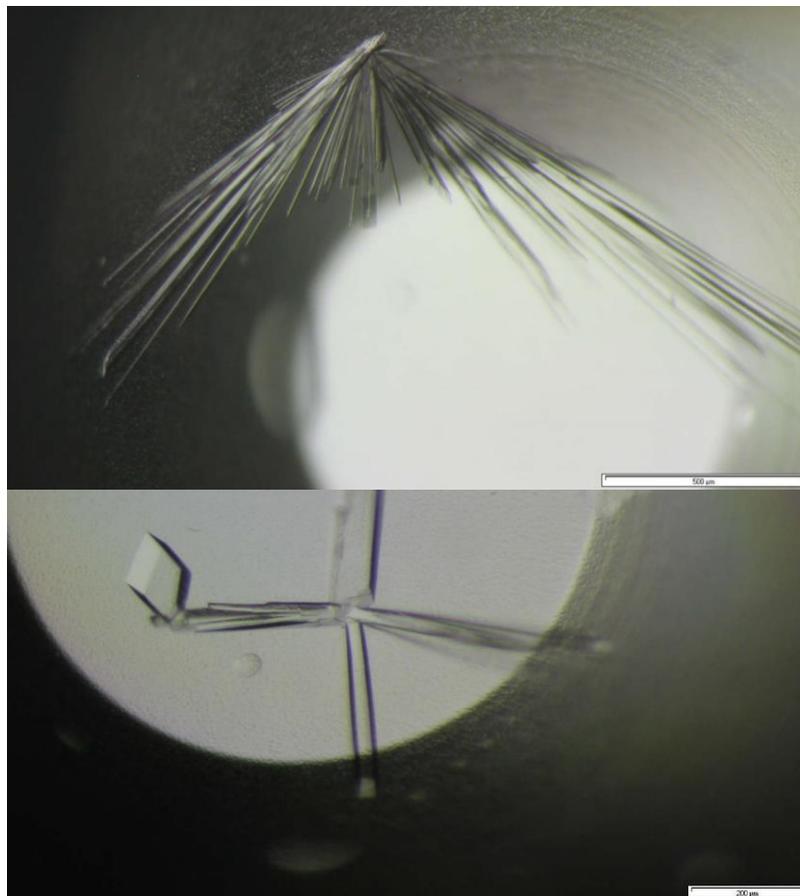


Figure 15: Tri a bA crystals grown in the optimized condition of 25 % PEG 3350, 100 mM Bis-Tris pH 5.0

## Structure

## Der p 4

The structure of Der p 4 was solved to 1.5 Å. Refinement achieved an R factor of 0.14 and an R free factor of 0.16. The calcium and chloride binding sites are clearly visible, as are the four disulfide bonds. A summary of the data merging and refinement statistics is found in the following table:

Der p 4	
Wavelength (Å)	0.9793
Resolution range (Å)	45.6 - 1.5 (1.554 - 1.5)
Space group	P 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>
	77.65 112.68 123.64
Unit cell	90 90 90
Total reflections	969194 (92138)
Unique reflections	168656 (16472)
Multiplicity	5.7 (5.6)
Completeness (%)	97.27 (92.71)
Mean I/sigma(I)	11.48 (2.41)
Wilson B-factor	16.54
R-merge	0.07948 (0.7227)
R-meas	0.08722
CC1/2	0.998 (0.557)
CC*	0.999 (0.846)
R-work	0.1367 (0.2381)
R-free	0.1624 (0.2654)
Number of non-hydrogen atoms	9445
macromolecules	8433
ligands	14
water	998
Protein residues	996
RMS(bonds)	0.01
RMS(angles)	1.24
Ramachandran favored (%)	97
Ramachandran outliers (%)	0.097
Clashscore	1.7
Average B-factor macromolecules	22.2
ligands	21.1
solvent	34.6
	31.1



Figure 16: Cartoon representation of Der p 4. Calcium ion shown as blue sphere.

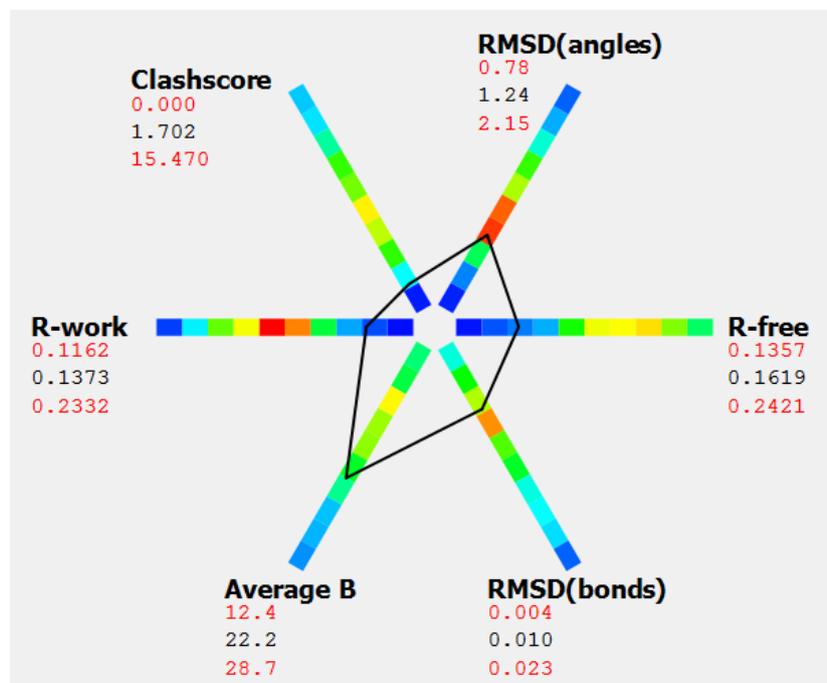


Figure 17: Der p 4 graph showing histograms of the distribution of statistics from 618 PDB entries of similar resolution (38)

### Calcium binding site

Eukaryotic  $\alpha$ -amylases require the complexation of one  $\text{Ca}^{2+}$  ion per molecule to retain their fold and function. Its function is presumed to be the stabilization of the interaction between the two domains creating the active site cleft. The calcium is bound by an  $\text{O}_8$  square antiprism formed by the sidechains of asparagine 99 and aspartate 164, the backbone carbonyl groups of arginine 155 and histidine 198, as well as 3 water molecules.

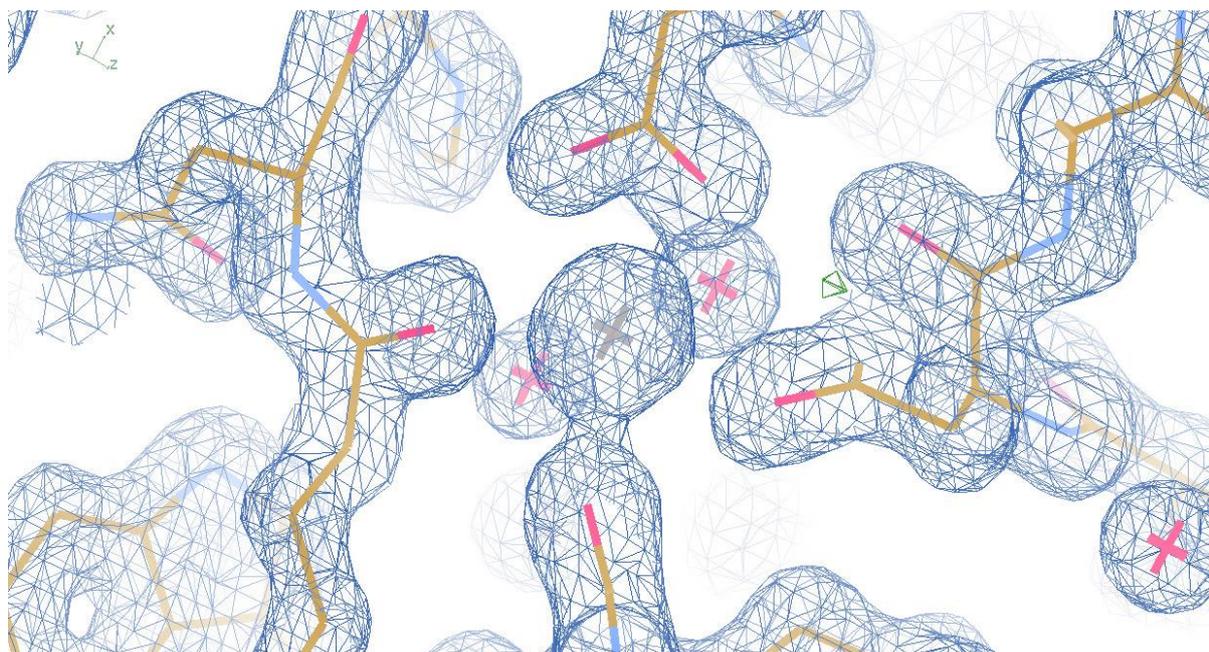


Figure 18: Calcium binding site of chain A as shown in Coot at 1.5 rmsd contour cutoff of the  $2 F_{\text{obs}} - F_{\text{calc}}$  map

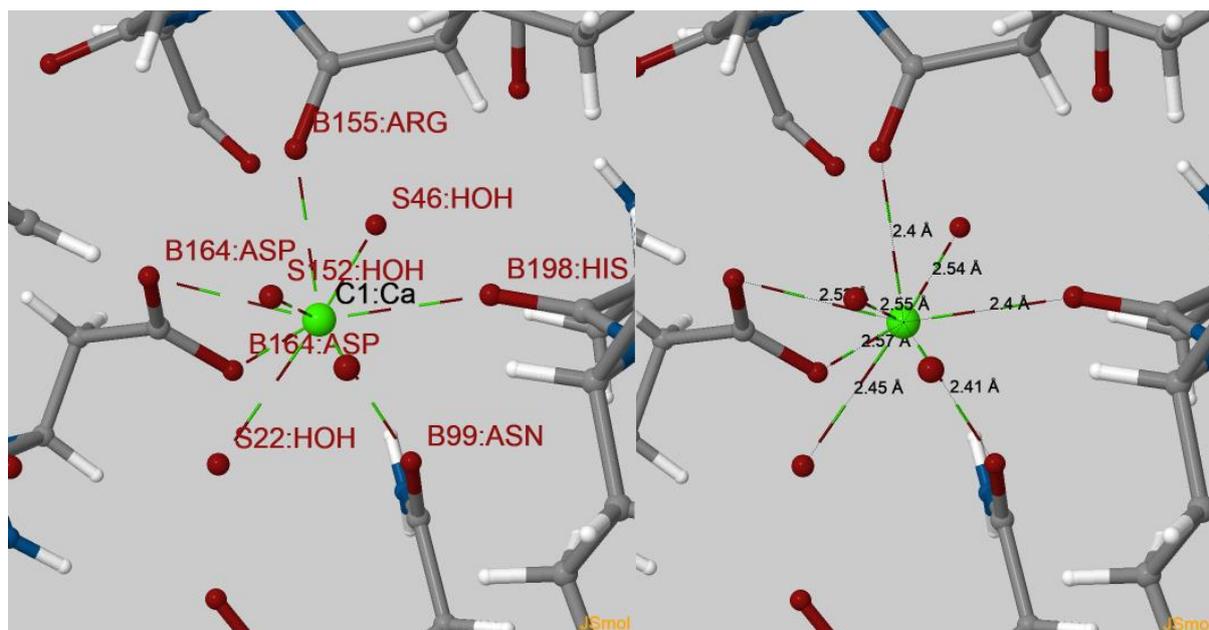


Figure 19: Calcium binding site of chain B shown in JSmol after analysis at the CheckMyMetal server (33)

The calcium binding sites of chain A and B were analyzed by the CheckMyMetal server (33) with the results and short explanation of the categories given below.

Metal	Occupancy	B factor (env.)(33)	Ligands	Valence(34)	nVECSUM(35)	Geometry (33),(36)	gRMSD(°)(33)	Vacancy(33)
Ca B	1	15.9 (16.2)	O <sub>8</sub>	1.9	0.02	Square Antiprism	10.6°	0
Ca A	1	24.4 (24.1)	O <sub>8</sub>	1.9	0.04	Square Antiprism	10.5°	0

Occupancy	Occupancy of ion under consideration
B factor (env.)(33)	Metal ion B factor, with valence-weighted environmental average B factor in parenthesis
Ligands	Elemental composition of the coordination sphere
Valence(34)	Summation of bond valence values for an ion binding site. Valence accounts for metal-ligand distances
nVECSUM(35)	Summation of ligand vectors, weighted by bond valence values and normalized by overall valence. Increase when the coordination sphere is not symmetrical due to incompleteness.
Geometry(33),(36)	Arrangement of ligands around the ion, as defined by the NEIGHBORHOOD algorithm
gRMSD(°)(33)	R.M.S. Deviation of observed geometry angles (L-M-L angles) compared to ideal geometry, in degrees
Vacancy(33)	Percentage of unoccupied sites in the coordination sphere for the given geometry

### Chloride binding site

The eukaryotic  $\alpha$ -amylases can bind one  $\text{Cl}^-$  ion ( $\text{Br}^-$  is also accepted) per molecule. This binding is not necessary for activity but increases the turnover rate several times over.

The chloride ion binding site refined well with chloride set at occupancy 1 with a B-factor of 19.7, which is comparable to its surrounding. It shows bidentate binding by two arginines (Arg 192 and Arg 332), with two ordered water molecules on one side and isoleucine 251 on the other.

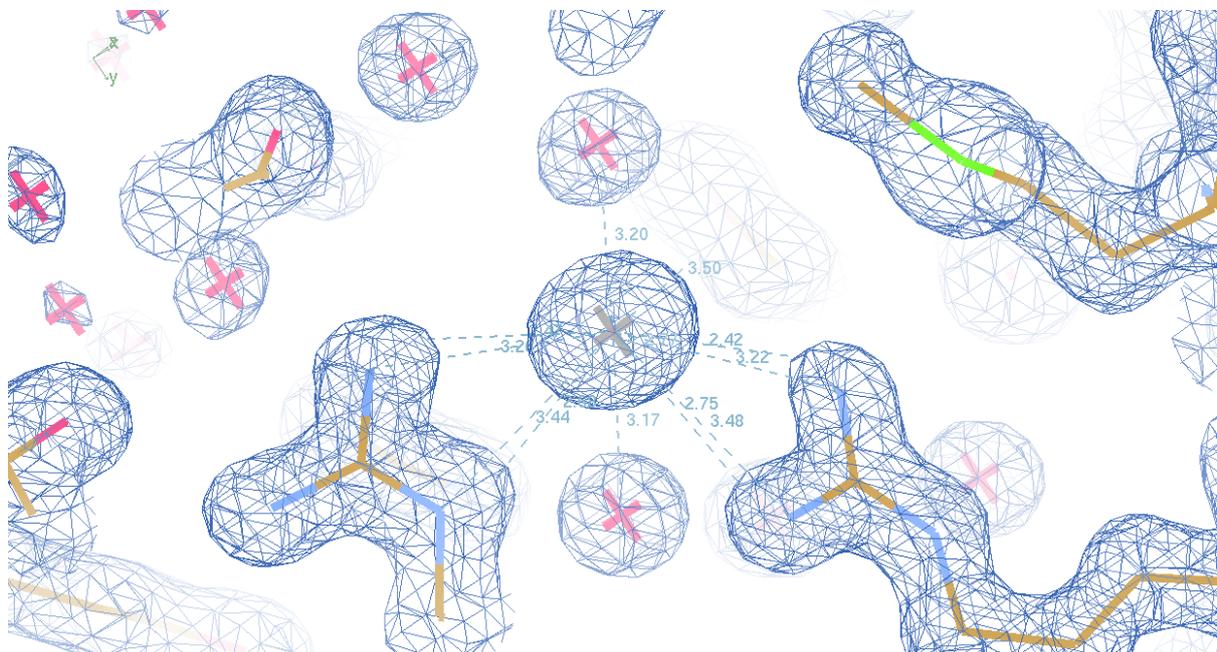
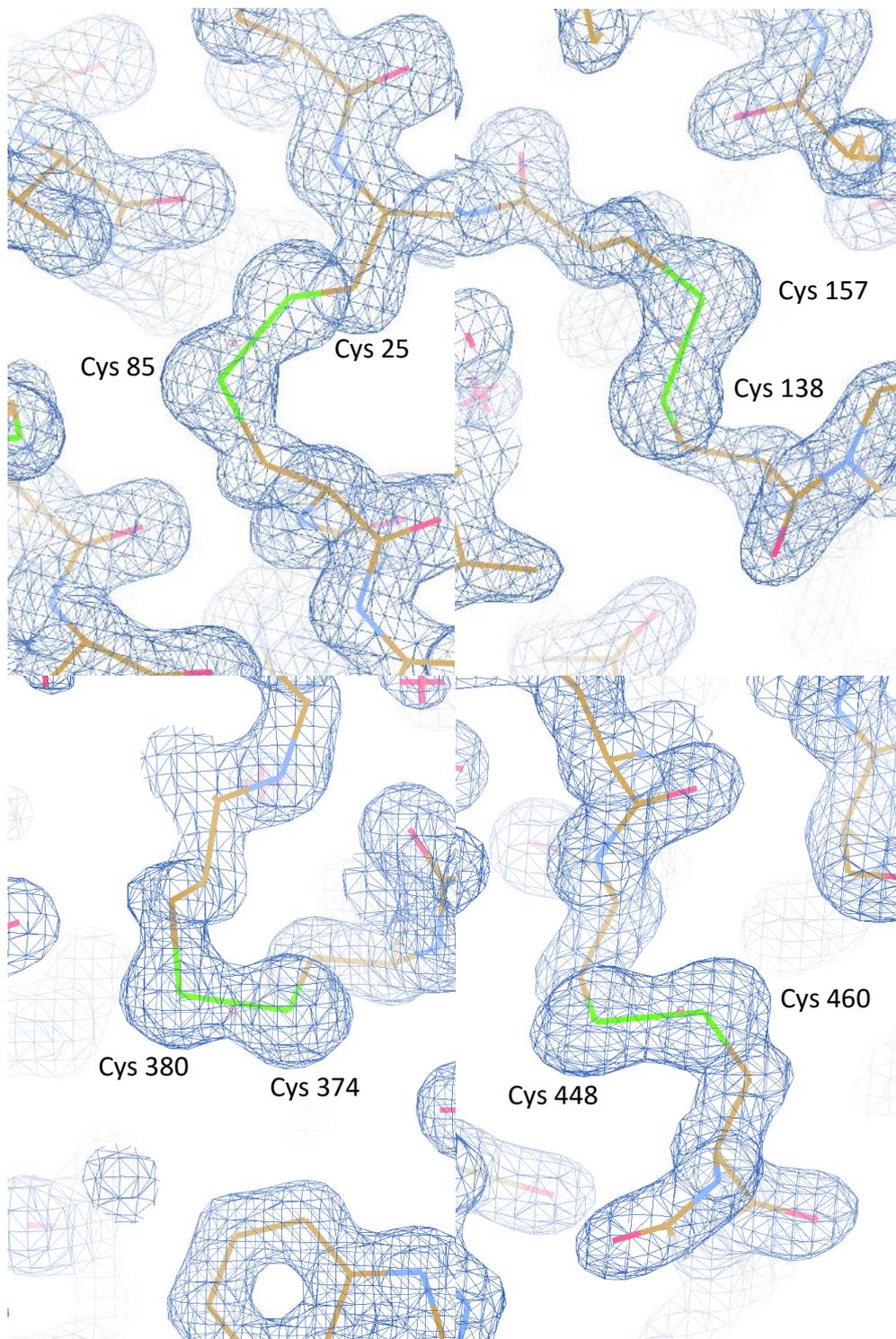


Figure 20: Chloride ion binding site in chain A shown in Coot at 1.5 rmsd contour cutoff of the  $2 F_{\text{obs}} - F_{\text{calc}}$  map. It is bound by arginine 192 (on the left) and arginine 332 (on the right) and two water molecules.

### *Disulfide bonds*

All four expected disulfide bonds can clearly be seen in the electron density map, showing the complete oxidization expected, as predicted by the other eukaryotic  $\alpha$ -amylase structures solved.



**Figure 21:** The four disulfide bonds of Der p 4 as shown in Coot at 1.5 rmsd contour cutoff of the  $2 F_{\text{obs}} - F_{\text{calc}}$  map

### *Active Site*

The comparison of the active site of Der p 4 and the human pancreatic  $\alpha$ -amylase shows that the three catalytic residues (Asp 194, Glu 230 and Asp 295 in Der p 4) overlay well.

The chloride ion is held directly below the active site and is presumed to polarize the water molecule used in the hydrolyzation of the glycoside bond.

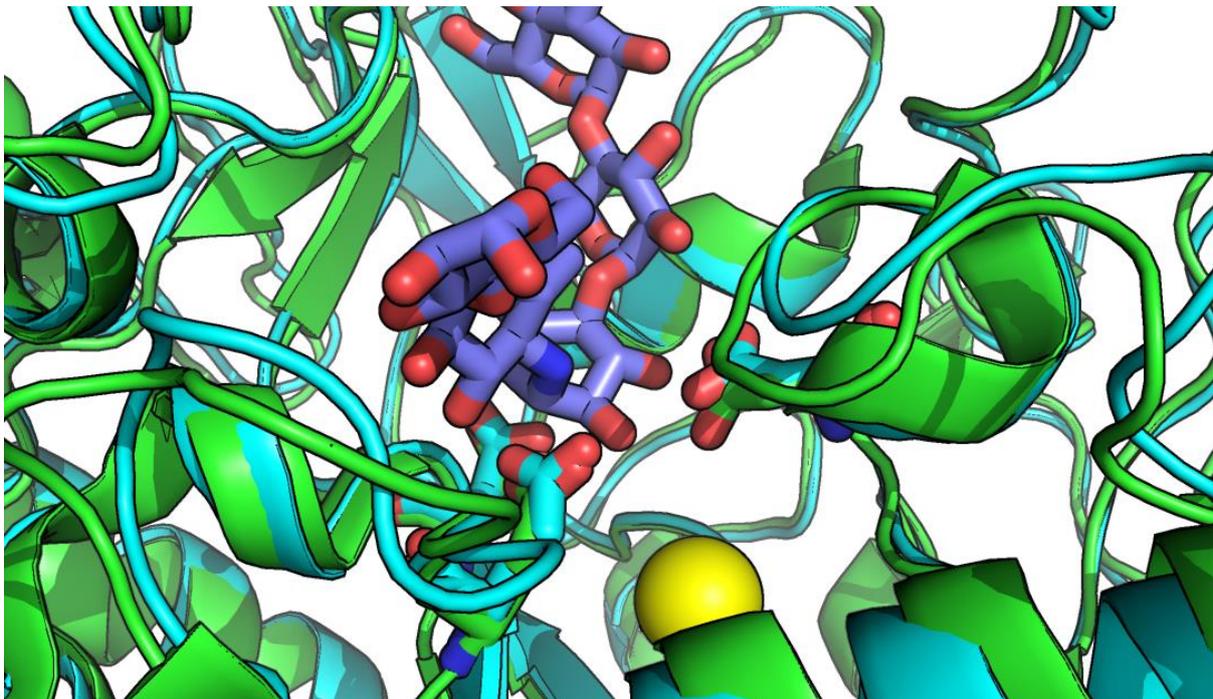


Figure 22: The active site of Der p 4 (teal) overlaid with the human pancreatic  $\alpha$ -amylase (green) complexed with acarbose (blue) from the pdb model 3BAY (39). The chloride ion from Der p 4 is shown as a yellow sphere.

*Epitope prediction*

Epitope prediction by the EPSVR server (24) shows four high likelihood patches for Der p 4. Three of them surround the active site cleft while the fourth is located on the beta sandwich domain. Interestingly no epitopes are predicted on the flipside of the protein.

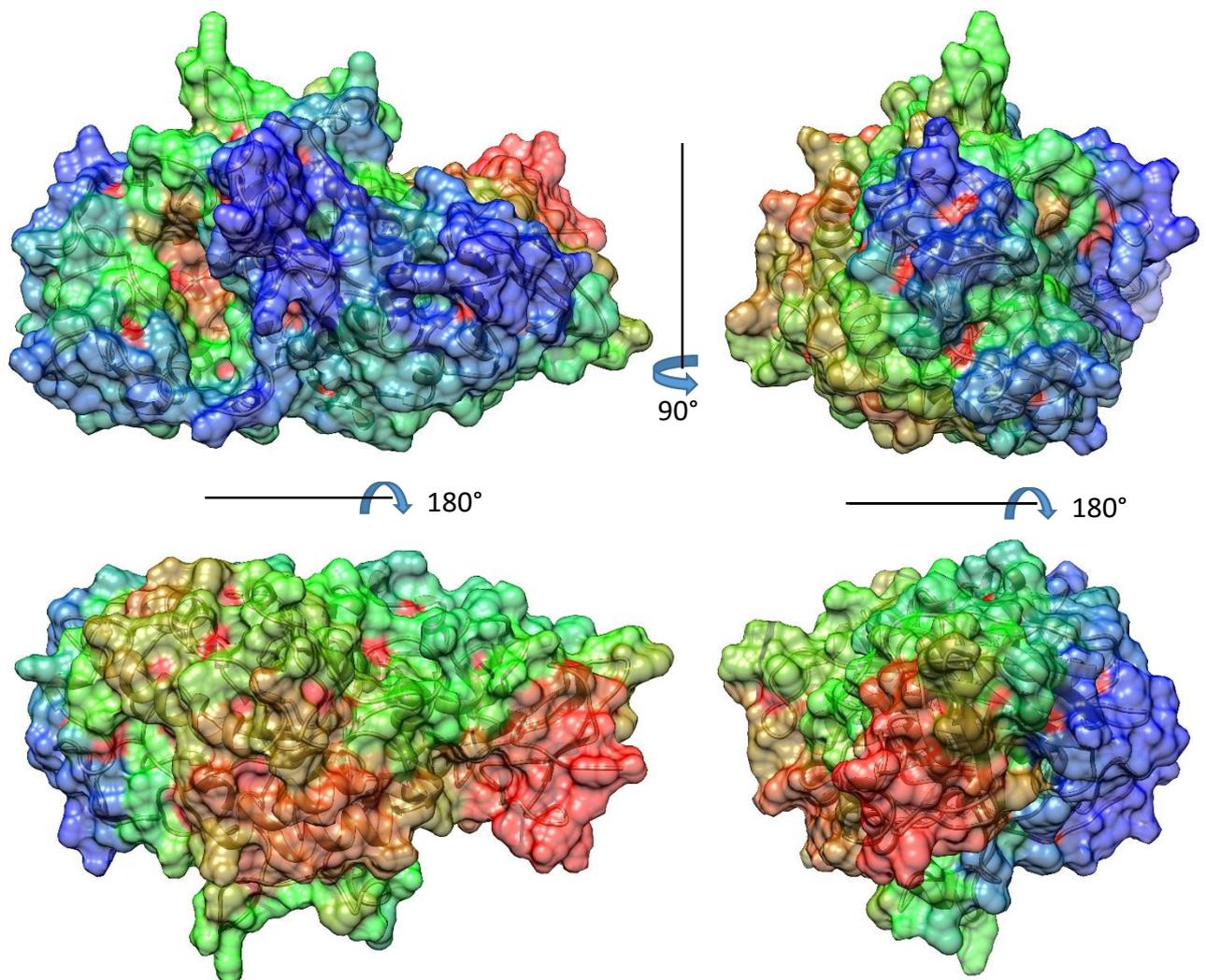


Figure 23: Der p 4 epitope prediction by the EPSVR server (24). Rainbow colouring from red (0) to blue (highest) score.

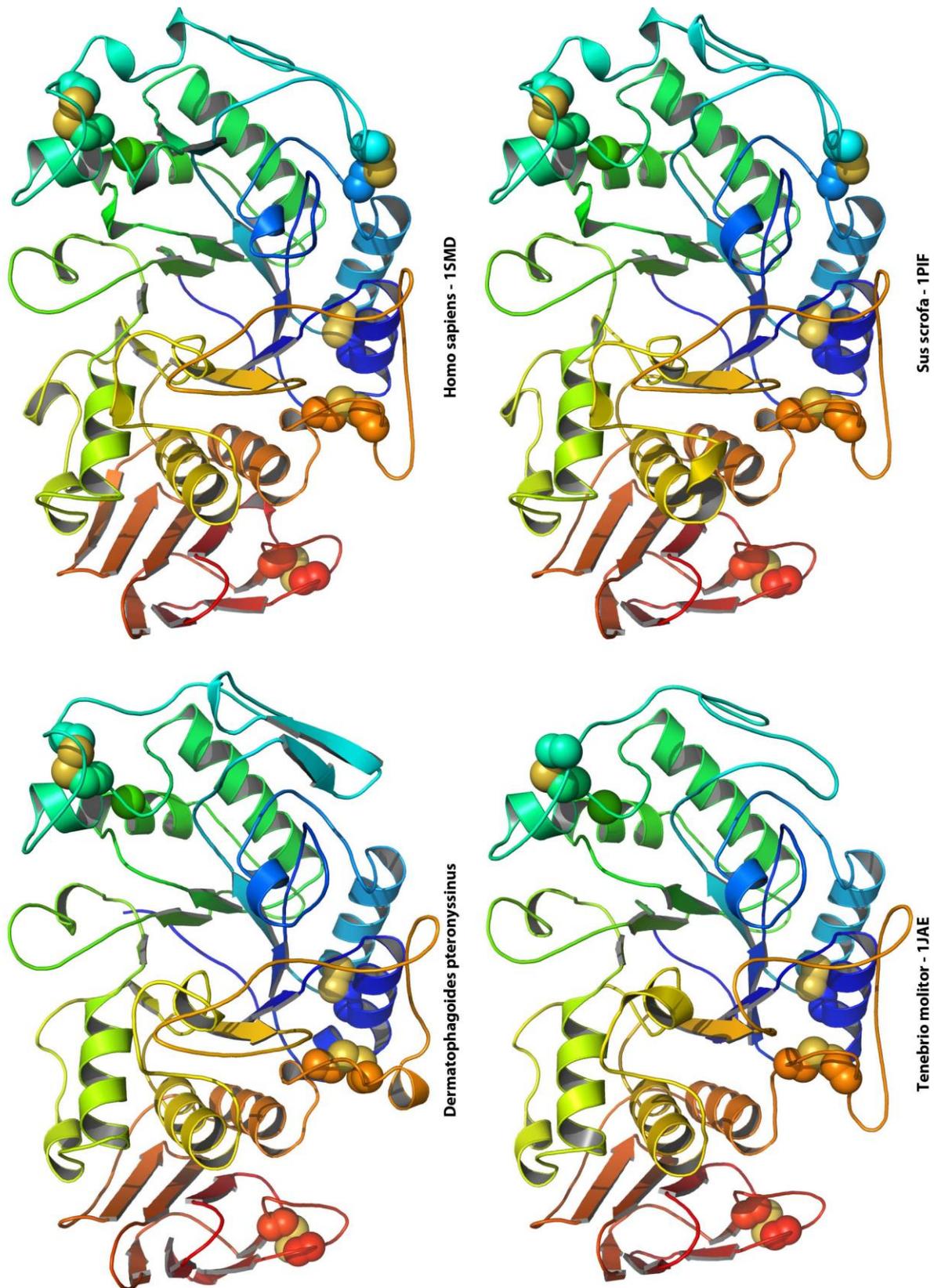


Figure 24: Cartoon representation of Der p 4, showing cysteines as spheres, compared to three other known eukaryotic  $\alpha$ -amylases

## Tri a bA

The structure of Tri a bA was solved to 1.8 Å from dataset 'Tri a bA twinned' and 2.0 Å from dataset 'Tri a bA ice'. Refinement achieved an R factor of 0.17 and an  $R_{\text{free}}$  factor of 0.20 (0.16 and 0.18 for Tri a bA ice). A summary of the data merging and refinement statistics is found in the following table:

	Tri a bA twinned	Tri a bA ice
Wavelength (Å)	0.9717	0.9717
Resolution range (Å)	47.12 - 1.8 (1.864 - 1.8)	27.06 - 2.0 (2.071 - 2.0)
Space group	P 4 <sub>3</sub>	P 4 <sub>3</sub>
Unit cell	94.515 94.515 66.447	93.87 93.87 65.86
Unit cell	90 90 90	90 90 90
Total reflections	362797 (36000)	147817 (14537)
Unique reflections	54377 (5417)	35923 (3829)
Multiplicity	6.7 (6.6)	3.8 (3.8)
Completeness (%)	100.00 (100.00)	92.55 (97.05)
Mean I/sigma(I)	5.05 (2.28)	6.93 (3.76)
Wilson B-factor	17.02	14.93
R-merge	0.2428 (0.679)	0.1203 (0.2861)
R-meas	0.2635	0.1394
CC1/2	0.977 (0.73)	0.986 (0.906)
CC*	0.994 (0.918)	0.996 (0.975)
R-work	0.1697 (0.2850)	0.1594 (0.1763)
R-free	0.2029 (0.3078)	0.1847 (0.2034)
Number of non-hydrogen atoms	4639	4482
macromolecules	3922	3922
water	717	560
Protein residues	487	487
RMS(bonds)	0.011	0.007
RMS(angles)	1.19	0.99
Ramachandran favored (%)	98	99
Ramachandran outliers (%)	0	0
Clashscore	3.23	2.85
Average B-factor	21.5	19.5
macromolecules	19.2	18.2
solvent	34.1	29

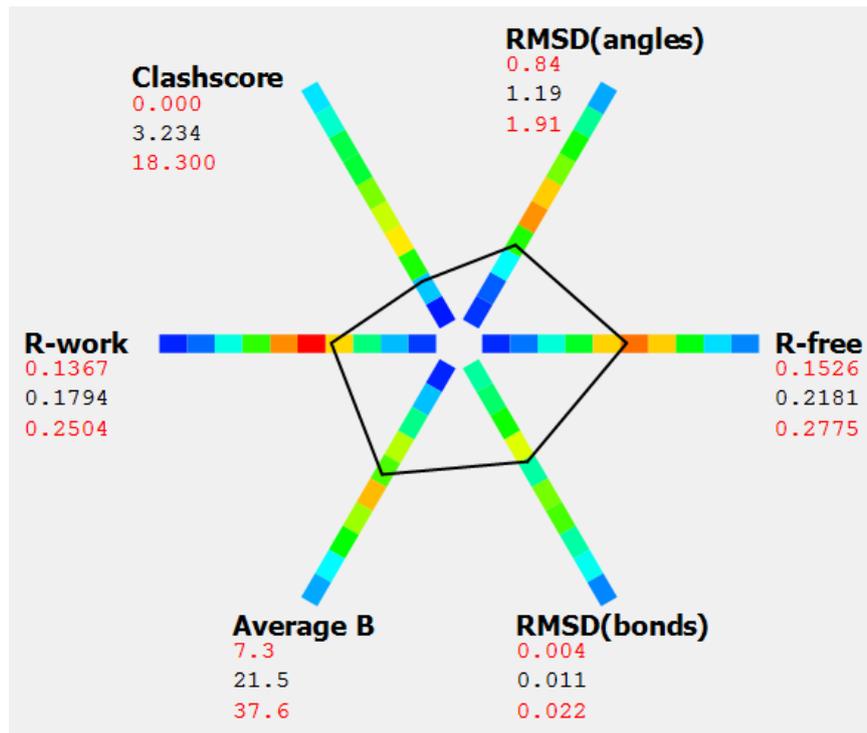


Figure 25: Tri a bA twinned graph showing histograms of the distribution of statistics from 736 PDB entries of similar resolution (38)

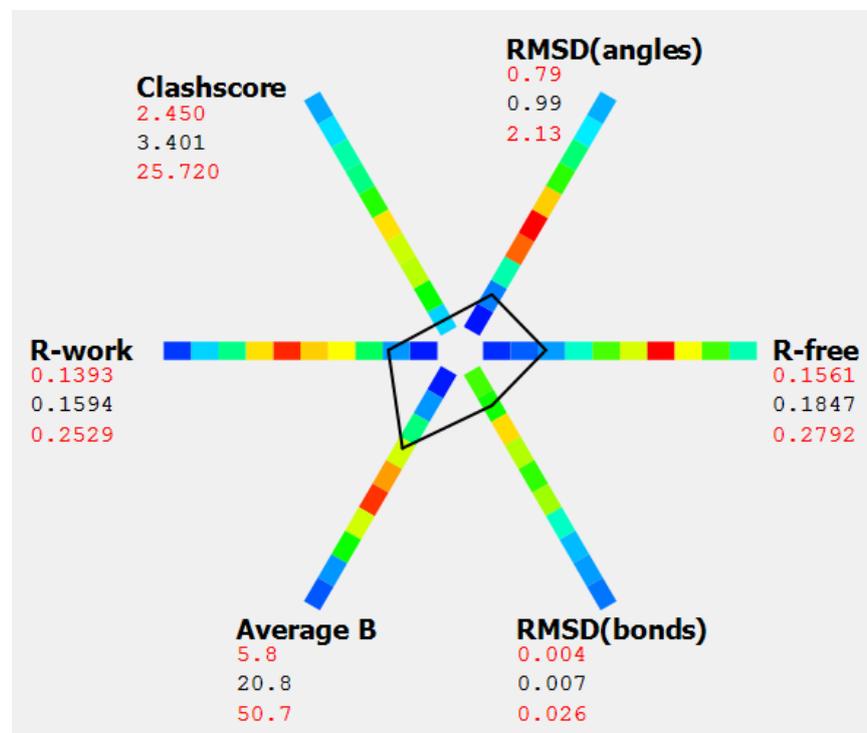


Figure 26: Tri a bA ice graph showing histograms of the distribution of statistics from 734 PDB entries of similar resolution (38)

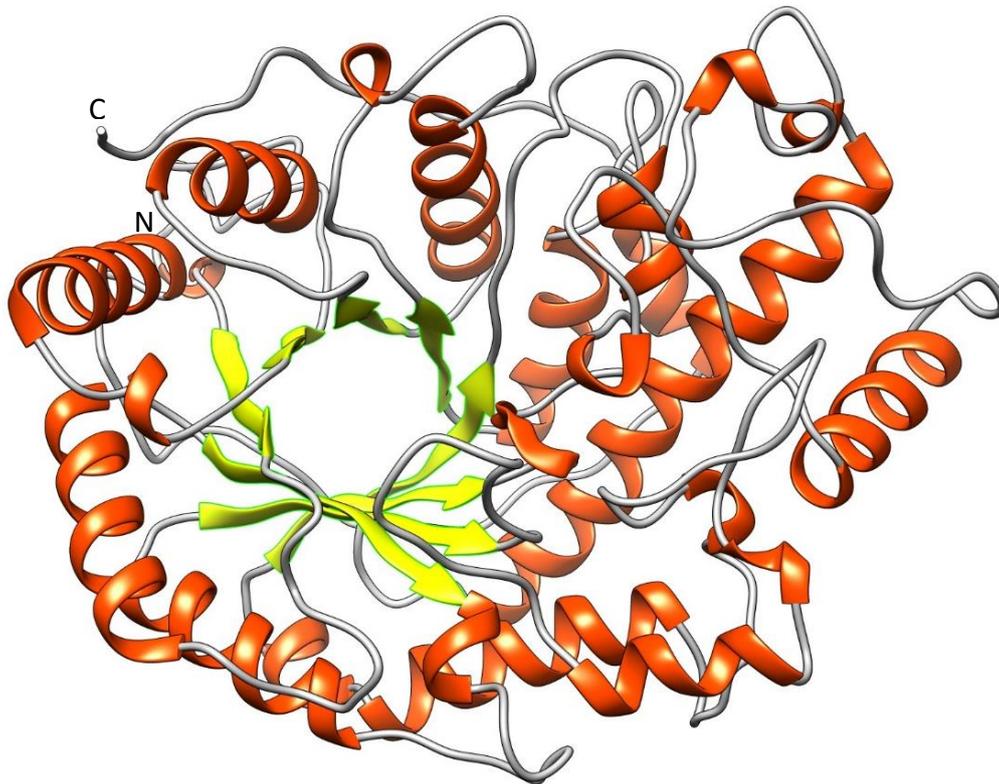


Figure 27: Tri a bA in cartoon representation showing the  $\beta$ -barrel forming the active site from the substrate accessible side.

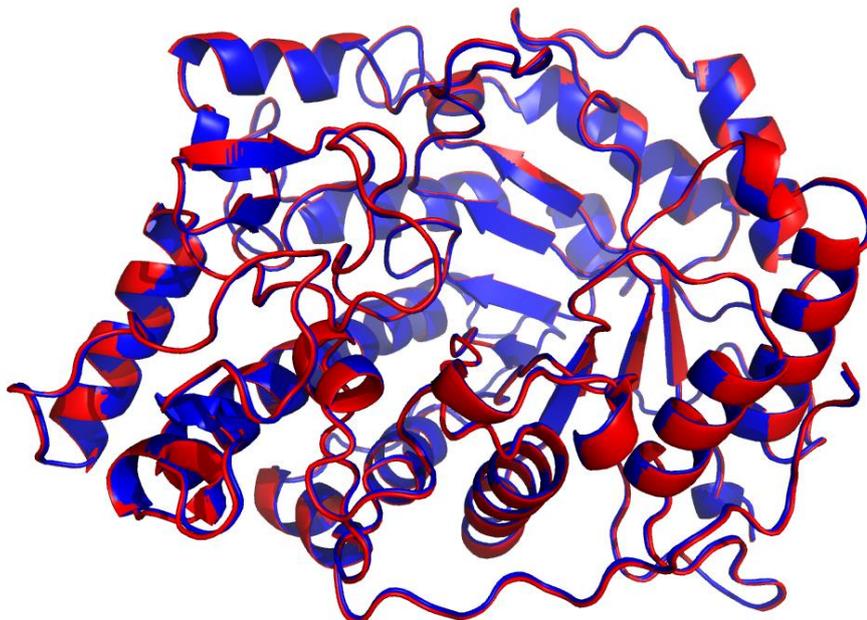


Figure 28: Comparison of Tri a bA ice (blue) and Tri a bA twinned (red). The structures align with an rmsd of 0.256

### Active Site

A comparison with the published active site of the barley  $\beta$ -amylase (PDB model 2XFF (37)) shows, that the catalytic residues as well as the substrate binding residues are overlapping neatly.

Asp 99 is shifted by 1 Å, but since it is positioned on the long substrate binding loop, a slight shift in a substrate free crystal structure does not imply a biological difference.

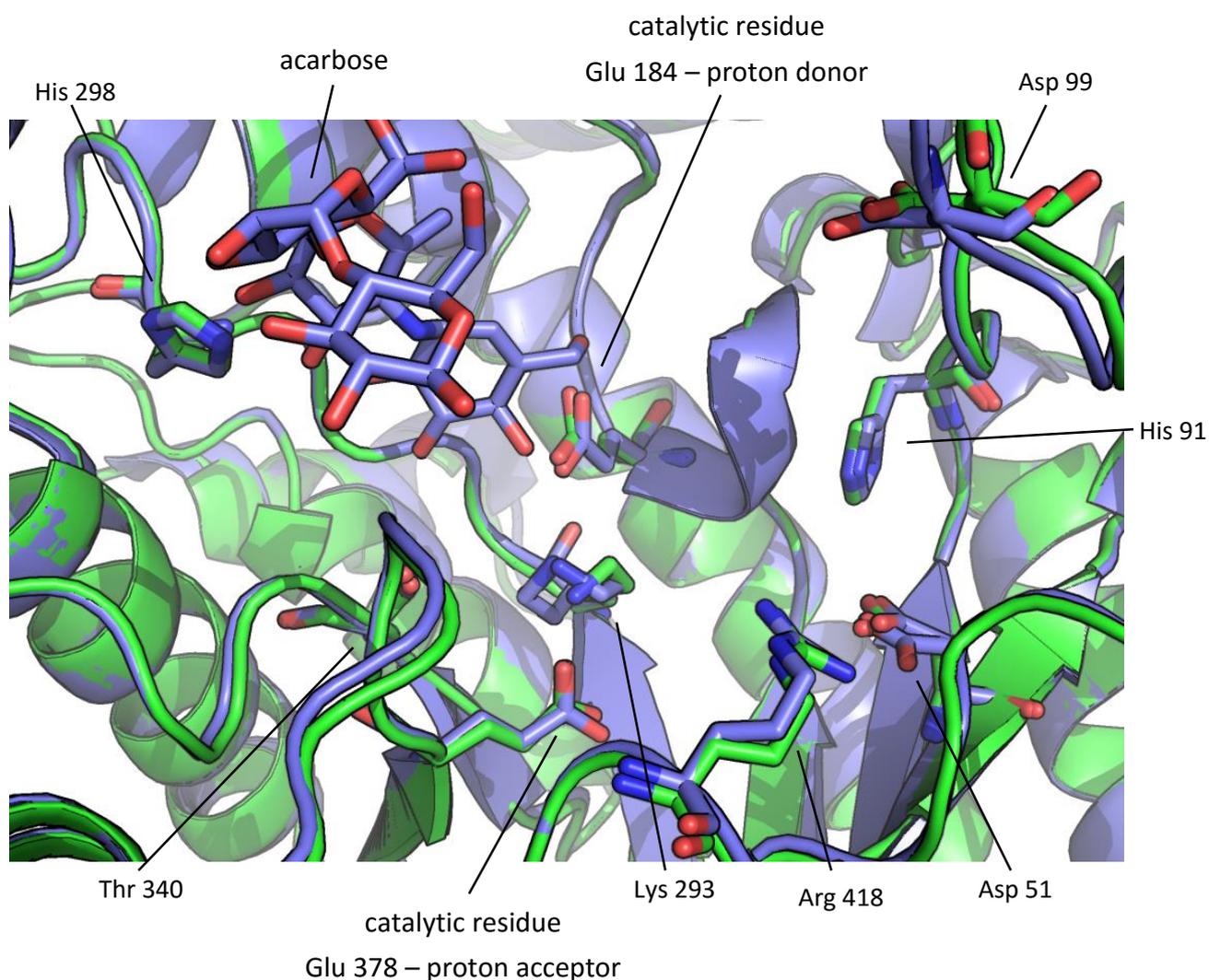


Figure 29: Overlay of the active site residues of Tri a bA (green) and barley  $\beta$ -amylase in complex with acarbose 2XFF (blue). The catalytic residues Glu 184 and Glu 378 as well as the seven residues annotated as substrate binding by the UniProt database are shown as stick models.

### *Cysteines*

The five cysteines of Tri a bA were found to be in their reduced state. This is in line with the previously published  $\beta$ -amylases. Intramolecular disulfide bonding is not possible in the native fold as there are no neighboring cysteines and although not all cysteines are completely inaccessible from the solvent, they are not available for intermolecular disulfide bonds because of steric hindrance.

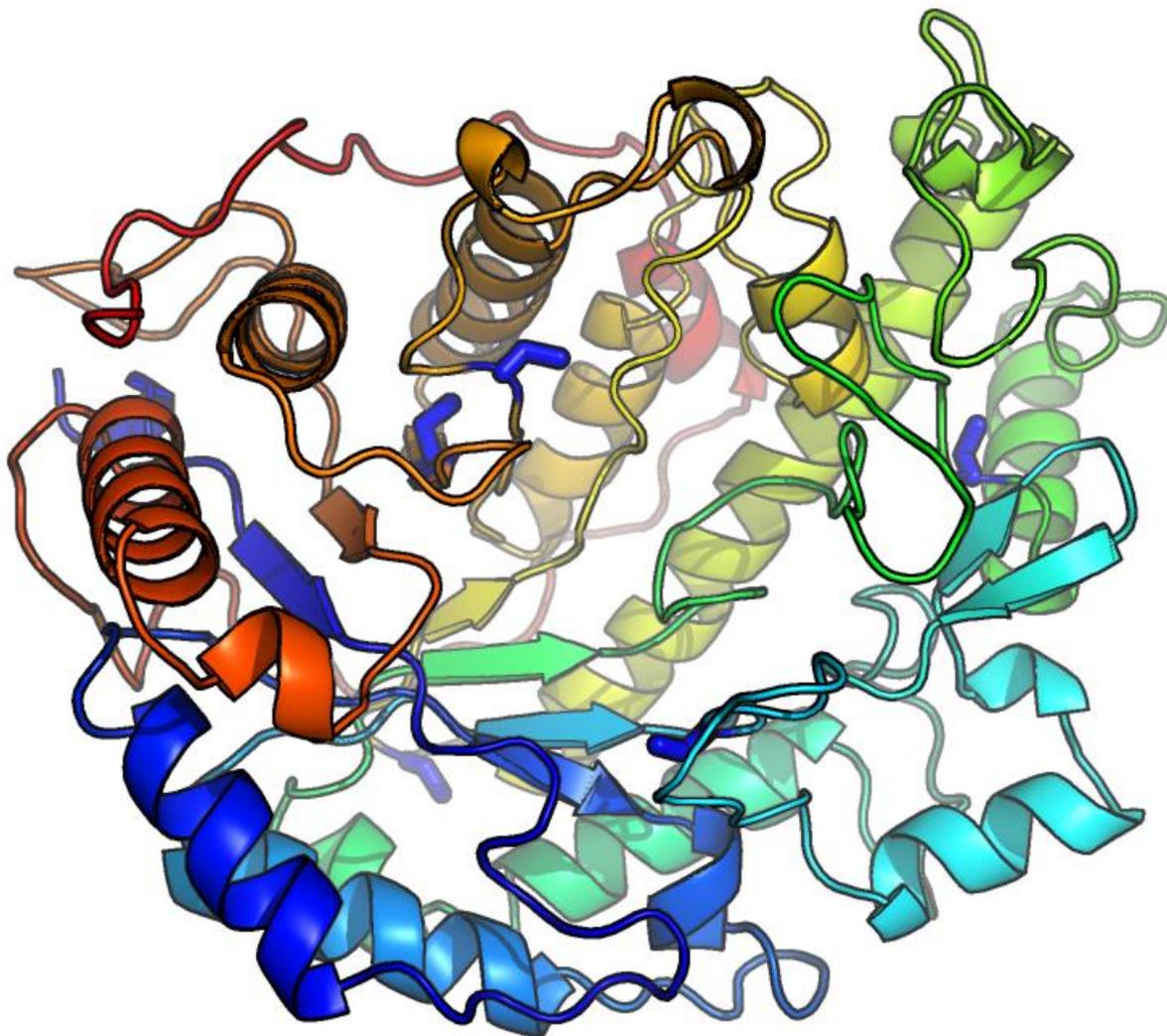


Figure 30: Tri a bA in spectrum coloured cartoon representation (N to C termini from blue to red) with the five cysteines shown as blue sticks.

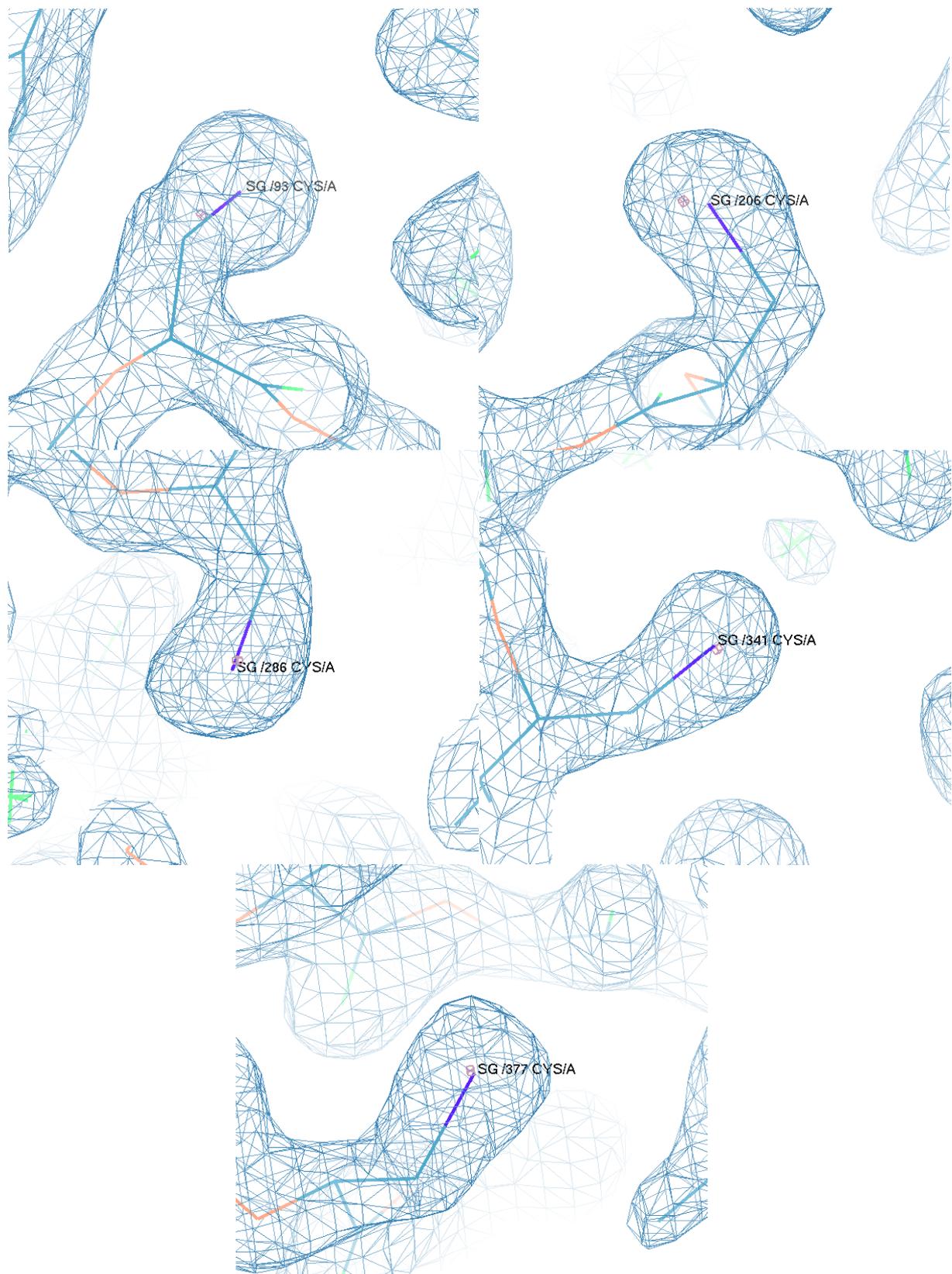


Figure 31: The five cysteines of Tri a bA ice as shown in Coot at 1.5 rmsd contour cutoff of the  $2 F_{\text{obs}} - F_{\text{calc}}$  map

*Epitope prediction*

Epitope prediction by the EPSVR server (24) shows one large high likelihood patch encompassing one full side of the protein and one small patch on the reverse side of the active site.

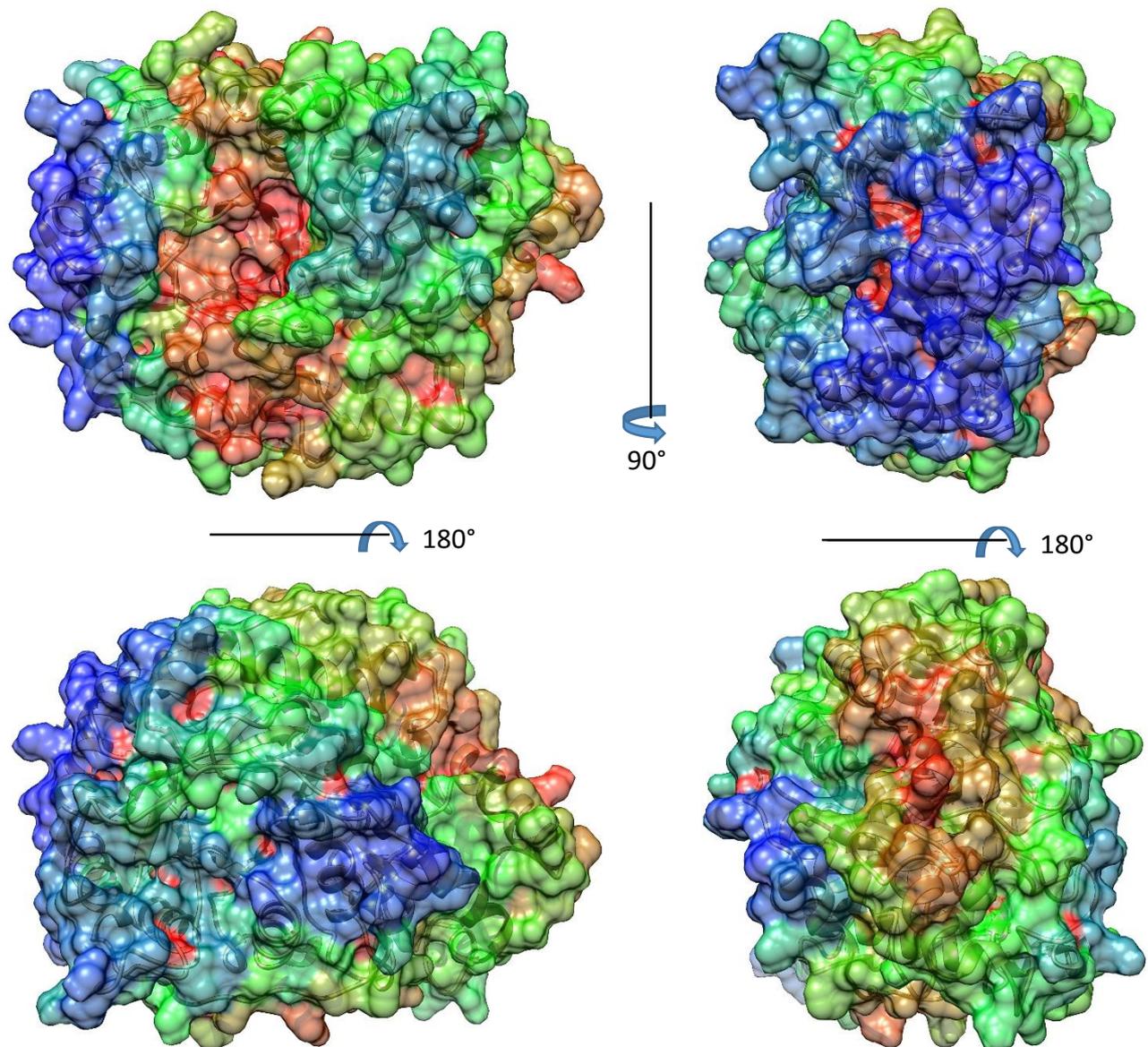


Figure 32: Tri a bA epitope prediction by the EPSVR server (24) in rainbow colouring from red (0) to blue (highest) score.

## Discussion

In line with the relatively new field of molecular allergology, this work sheds light on a small group of allergens, on how to express and handle them, as well as revealing their structure.

It shows the versatility of SHuffle strain *E. coli* in expressing active eukaryotic proteins with cysteine content. Not only did this strain oxidize Der p 4 to its full activity, sufficient Tri a bA was produced completely reduced, which points to the strain being perfectly suited to express proteins of unknown disulfide bond status since it clearly does not indiscriminately oxidize the proteins it expresses.

The use of a better tunable expression system than the T7 system used here is advisable since the folding/oxidization speed for Der p 4 was far lower than the expression, causing the need for extended optimization trials to achieve a sensible yield.

For storage, and most of all for diagnostic and treatment applications of the often conformational epitope dependent IgE based allergies, the correct protein fold is essential. Eukaryotic amylases can be expressed in their native fold in *E. coli* with this protocol, removing the need for more time consuming eukaryotic expression systems.

Last but not least, the crystal structures elucidated in this project add to the constantly increasing knowledge of allergens, their structure and the differences compared to non-allergic proteins. They therefore complement the structural databases and bring us closer to possibly understanding the reason why some proteins illicit allergic responses while others do not.

### *Further Research*

With the natively folded and active proteins being expressed and characterized, the allergen's resistance to external influences like the different stages of digestion or long time storage under various conditions can be assessed.

They can be, and to some degree have already been, used to determine which allergic patients produce IgE against these specific enzymes, allowing a more precise understanding of their disease. Should it prove sensible, they could even be included in patient specific immunotherapy based on a personalized mix of recombinant allergens, should this form of treatment be established.

The predicted antibody epitopes will be compared to other epitope predictions and can provide a starting point for modifying the proteins to create hypoallergenic mutants.

## Table of Figures

Figure 1: Overview of hypersensitivities (1) .....	2
Figure 2: Multiple sequence alignment of four solved $\alpha$ -amylases created using T-Coffee (28). Der p 4, porcine pancreatic $\alpha$ -amylase (1DHK), human salivary $\alpha$ -amylase (1C8Q) and <i>Tenebrio molitor</i> larval $\alpha$ -amylase (1TMQ). The alignment quality is given by colour from blue (bad) to red (good). Disulfide bonds are denoted by corresponding numbers, residues involved in ion binding by the chemical symbol of the ion bound and catalytic residues by stars.....	13
Figure 3: Multiple sequence alignment of four solved $\alpha$ -amylases created using T-Coffee (11). Tri a bA, barley $\beta$ -amylase (2XFF), soybean $\beta$ -amylase (1Q6C) and sweet potato $\beta$ -amylase (1FA2). The alignment quality is given by colour from blue (bad) to red (good).....	15
Figure 4: "Disulfide bond formation in the cytoplasm of SHuffle. Schematic diagram of the redox pathways in the cytoplasm of SHuffle. Dotted lines represent disabled protein interactions due to the deletion of <i>trxB</i> and <i>gor</i> . Redox state of cysteines (yellow balls) are indicated (oxidized = ball + stick; reduced = ball). (A) Protein is reduced by Grx1 or oxidized by Trx1. (B) Mis-oxidized protein is isomerized to its native correctly folded state (C) by DsbC." (29) .....	17
Figure 5: Initial attempt at expressing Der p 4 in BL21 strain <i>E. coli</i> resulting in insoluble protein aggregates.....	18
Figure 6: Soluble Der p 4 shows uniform disulfide bonding, causing a single band of equal density to travel further in a non-reducing SDS PAGE .....	19
Figure 7: The ethanol purification step yields pure protein from the IMAC purified sample .	20
Figure 8: Tri a bA oxidization state after HisTrap and ethanol cleanup .....	21
Figure 9: Tri a bA starch hydrolysis assay showing amylase activity in the His-Trap eluate ...	22
Figure 10: Tri a bA Thermofluor melting results showing highest thermo-stability at pH 5.4	23
Figure 11: Comparison between freshly purified Tri a bA and a sample stores for 6 months at 4°C.....	24
Figure 12: Optimization of Der p 4 crystals. From initial screen and first optimization success to well diffracting crystals.....	25
Figure 13: Der p 4 crystal diffraction at home source .....	26
Figure 14: Der p 4 crystal diffraction before and after annealing .....	26
Figure 15: Tri a bA crystals grown in the optimized condition of 25 % PEG 3350, 100 mM Bis-Tris pH 5.0 .....	27
Figure 16: Cartoon representation of Der p 4. Calcium ion shown as blue sphere. ....	29
Figure 17: Der p 4 graph showing histograms of the distribution of statistics from 618 PDB entries of similar resolution (39) .....	29

---

Figure 18: Calcium binding site of chain A as shown in Coot at 1.5 rmsd contour cutoff of the $2 F_{obs}-F_{calc}$ map.....	30
Figure 19: Calcium binding site of chain B shown in JSmol after analysis at the CheckMyMetal server (34) .....	30
Figure 20: Chloride ion binding site in chain A shown in Coot at 1.5 rmsd contour cutoff of the $2 F_{obs}-F_{calc}$ map. It is bound by arginine 192 (on the left) and arginine 332 (on the right) and two water molecules. ....	32
Figure 21: The four disulfide bonds of Der p 4 as shown in Coot at 1.5 rmsd contour cutoff of the $2 F_{obs}-F_{calc}$ map .....	33
Figure 22: The active site of Der p 4 (teal) overlaid with the human pancreatic $\alpha$ -amylase (green) complexed with acarbose (blue) from the pdb model 3BAY (40). The chloride ion from Der p 4 is shown as a yellow sphere. ....	34
Figure 23: Der p 4 epitope prediction by the EPSVR server (24). Rainbow colouring from red (0) to blue (highest) score.....	35
Figure 24: Cartoon representation of Der p 4, showing cysteines as spheres, compared to the three other known eukaryotic $\alpha$ -amylases.....	36
Figure 25: Tri a bA twinned graph showing histograms of the distribution of statistics from 736 PDB entries of similar resolution (39).....	38
Figure 26: Tri a bA ice graph showing histograms of the distribution of statistics from 734 PDB entries of similar resolution (39) .....	38
Figure 27: Tri a bA in cartoon representation showing the $\beta$ -barrel forming the active site from the substrate accessible side. ....	39
Figure 28: Comparison of Tri a bA ice (blue) and Tri a bA twinned (red). The structures align with an rmsd of 0.256.....	39
Figure 29: Overlay of the active site residues of Tri a bA (green) and barley b-amylase in complex with acarbose 2XFF (blue). The catalytic residues Glu 184 and Glu 378 as well as the seven residues annotated as substrate binding by the UniProt database are shown as stick models.....	40
Figure 30: Tri a bA in spectrum coloured cartoon representation (N to C termini from blue to red) with the five cysteines shown as blue sticks. ....	41
Figure 31: The five cysteines of Tri a bA ice as shown in Coot at 1.5 rmsd contour cutoff of the $2 F_{obs}-F_{calc}$ map.....	42
Figure 32: Tri a bA epitope prediction by the EPSVR server (24). Rainbow colouring from red (0) to blue (highest) score.....	43

---

---

## Bibliography

1. Johansson, S. G., J. O. Hourihane, J. Bousquet, C. Brujnzeel-Koomen, S. Dreborg, T. Haahtela, M. L. Kowalski, N. Mygind, J. Ring, P. van Cauwenberge, M. van Hage-Hamsten, and B. Wüthrich. 2001. A revised nomenclature for allergy. An EAACI position statement from the EAACI nomenclature task force. *Allergy* 56: 813–824.
2. Beasley, R., J. Crane, C. K. Lai, and N. Pearce. 2000. Prevalence and etiology of asthma. *J. Allergy Clin. Immunol.* 105: S466–S472.
3. Ring, J., U. Krämer, T. Schäfer, and H. Behrendt. 2001. Why are allergies increasing? *Curr. Opin. Immunol.* 13: 701–708.
4. Majkowska-Wojciechowska, B., J. Pełka, L. Korzon, A. Kozłowska, M. Kaczała, M. Jarzębska, T. Gwardys, and M. L. Kowalski. 2007. Prevalence of allergy, patterns of allergic sensitization and allergy risk factors in rural and urban children. *Allergy Eur. J. Allergy Clin. Immunol.* 62: 1044–1050.
5. Nwaru, B. I., L. Hickstein, S. S. Panesar, A. Muraro, T. Werfel, V. Cardona, A. E. J. Dubois, S. Halken, K. Hoffmann-Sommergruber, L. K. Poulsen, G. Roberts, R. Van Ree, B. J. Vlieg-Boerstra, and a. Sheikh. 2014. The epidemiology of food allergy in Europe: A systematic review and meta-analysis. *Allergy Eur. J. Allergy Clin. Immunol.* 69: 62–75.
6. Lieberman, P., C. A Camargo, K. Bohlke, H. Jick, R. L. Miller, A. Sheikh, and F. E. R. Simons. 2006. Epidemiology of anaphylaxis: findings of the American College of Allergy, Asthma and Immunology Epidemiology of Anaphylaxis Working Group. *Ann. Allergy. Asthma Immunol.* 97: 596–602.
7. Riedler, J., W. Eder, G. Oberfeld, and M. Schreuer. 2000. Austrian children living on a farm have less hay fever, asthma and allergic sensitization. *Clin. Exp. Allergy* 30: 194–200.
8. Vierk, K. A., K. M. Koehler, S. B. Fein, and D. A. Street. 2007. Prevalence of self-reported food allergy in American adults and use of food labels. *J. Allergy Clin. Immunol.* 119: 1504–1510.
9. Matricardi, P. M., A. Bockelbrink, K. Beyer, T. Keil, B. Niggemann, C. Grüber, U. Wahn, and S. Lau. 2008. Primary versus secondary immunoglobulin e sensitization to soy and wheat in the Multi-Centre Allergy Study cohort. *Clin. Exp. Allergy* 38: 493–500.
10. Šotkovský, P., J. Sklenář, P. Halada, J. Cinová, I. Šetinová, A. Kainarová, J. Goliáš, K. Pavlásková, S. Honzová, and L. Tučková. 2011. A new approach to the isolation and characterization of wheat flour allergens. *Clin. Exp. Allergy* 41: 1031–1043.
11. De Souza, P. M., and P. D. O. e Magalhães. 2010. Application of microbial  $\alpha$ -amylase in industry - a review. *Brazilian J. Microbiol.* 41: 850–861.

- 
12. Bernfeld, P. 1955. [17] Amylases,  $\alpha$  and  $\beta$ . In *Methods in Enzymology* vol. 1. 149–158.
  13. Whitmore, L., and B. A. Wallace. 2008. Protein secondary structure analyses from circular dichroism spectroscopy: Methods and reference databases. *Biopolymers* 89: 392–400.
  14. Compton, L. A., and W. C. Johnson. 1986. Analysis of protein circular dichroism spectra for secondary structure using a simple matrix multiplication. *Anal. Biochem.* 155: 155–167.
  15. Sreerama, N., and R. W. Woody. 2000. Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Anal. Biochem.* 287: 252–260.
  16. Newman, J. 2004. Novel buffer systems for macromolecular crystallization. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 60: 610–612.
  17. Emsley, P., B. Lohkamp, W. G. Scott, and K. Cowtan. 2010. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 66: 486–501.
  18. Leslie, A. G. W., H. R. Powell, R. J. Read, and J. L. Sussman. 2007. Evolving Methods for Macromolecular Crystallography NATO Science Series. *Physics (College. Park. Md).* 245: 41–51.
  19. Winn, M. D., C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. W. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. a. Potterton, H. R. Powell, R. J. Read, A. Vagin, and K. S. Wilson. 2011. Overview of the CCP4 suite and current developments. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 67: 235–242.
  20. Kelley, L. A., S. Mezulis, C. M. Yates, M. N. Wass, and M. J. E. Sternberg. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* 10: 845–858.
  21. Adams, P. D., P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, and P. H. Zwart. 2010. PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 66: 213–221.
  22. Rejzek, M., C. E. Stevenson, A. M. Southard, D. Stanley, K. Denyer, A. M. Smith, M. J. Naldrett, D. M. Lawson, and R. a Field. 2011. Chemical genetics and cereal starch metabolism: structural basis of the non-covalent and covalent inhibition of barley  $\beta$ -amylase. *Mol. Biosyst.* 7: 718–730.
  23. Kabsch, W. 2010. Xds. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 66: 125–132.
  24. Liang, S., D. Zheng, D. M. Standley, B. Yao, M. Zacharias, and C. Zhang. 2010. EPSVR and EPMeta: prediction of antigenic epitopes using support vector regression and multiple server results. *BMC Bioinformatics* 11: 381.

- 
25. Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., and A Bairoch. 2005. Protein Identification and Analysis Tools on the ExPASy Server. *Proteomics Protoc. Handbook, Humana Press* 571–607.
26. S. Altschul, W. Gish, and W. Miller. 1990. Basic Local Alignment Search Tool. *J Mol Biol.* 215: 403–410.
27. Armougom, F., S. Moretti, O. Poirot, S. Audic, P. Dumas, B. Schaeli, V. Keduas, C. Notredame, and C. Notredame. 2006. Espresso: Automatic incorporation of structural information in multiple sequence alignments using 3D-Coffee. *Nucleic Acids Res.* 34.
28. Di Tommaso, P., S. Moretti, I. Xenarios, M. Orobittg, A. Montanyola, J. M. Chang, J. F. Taly, and C. Notredame. 2011. T-Coffee: A web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Res.* 39: W13–W17.
29. Lobstein, J., C. A Emrich, C. Jeans, M. Faulkner, P. Riggs, and M. Berkmen. 2012. SHuffle, a novel Escherichia coli protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. *Microb. Cell Fact.* 11: 56.
30. A. Loyter, M. Schramm. 1962. The glycogen-amylase complex as a means of obtaining highly purified alpha-amylases. *Biochim. Biophys. Acta* 65: 200–206.
31. Ericsson, B. M. Hallberg, G. T. DeTitta, N. Dekker, and P. Nordlund. 2006. Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Anal. Biochem.* 357: 289–298.
32. Lundgard, R., and B. Svensson. 1987. The four major forms of barley  $\beta$ -amylase. Purification, characterization and structural relationship. *Carlsberg Res. Commun.* 52: 313–326.
33. Zheng, H., M. D. Chordia, D. R. Cooper, M. Chruszcz, P. Müller, G. M. Sheldrick, and W. Minor. 2014. Validation of metal-binding sites in macromolecular structures with the CheckMyMetal web server. *Nat. Protoc.* 9: 156–70.
34. Brown, I. D. 2009. Recent developments in the methods and applications of the bond valence model. *Chem. Rev.* 109: 6858–6919.
35. Müller, P., S. Köpke, and G. M. Sheldrick. 2003. Is the bond-valence method able to identify metal atoms in protein structures? *Acta Crystallogr. - Sect. D Biol. Crystallogr.* 59: 32–37.
36. Kuppuraj, G., M. Dudev, and C. Lim. 2009. Factors governing metal-ligand distances and coordination geometries of metal complexes. *J. Phys. Chem. B* 113: 2952–2960.
37. Rejzek, M., C. E. Stevenson, A. M. Southard, D. Stanley, K. Denyer, A. M. Smith, M. J. Naldrett, D. M. Lawson, and R. A. Field. 2011. Chemical genetics and cereal starch metabolism:

structural basis of the non-covalent and covalent inhibition of barley  $\beta$ -amylase. *Mol. Biosyst.* 7: 718–730.

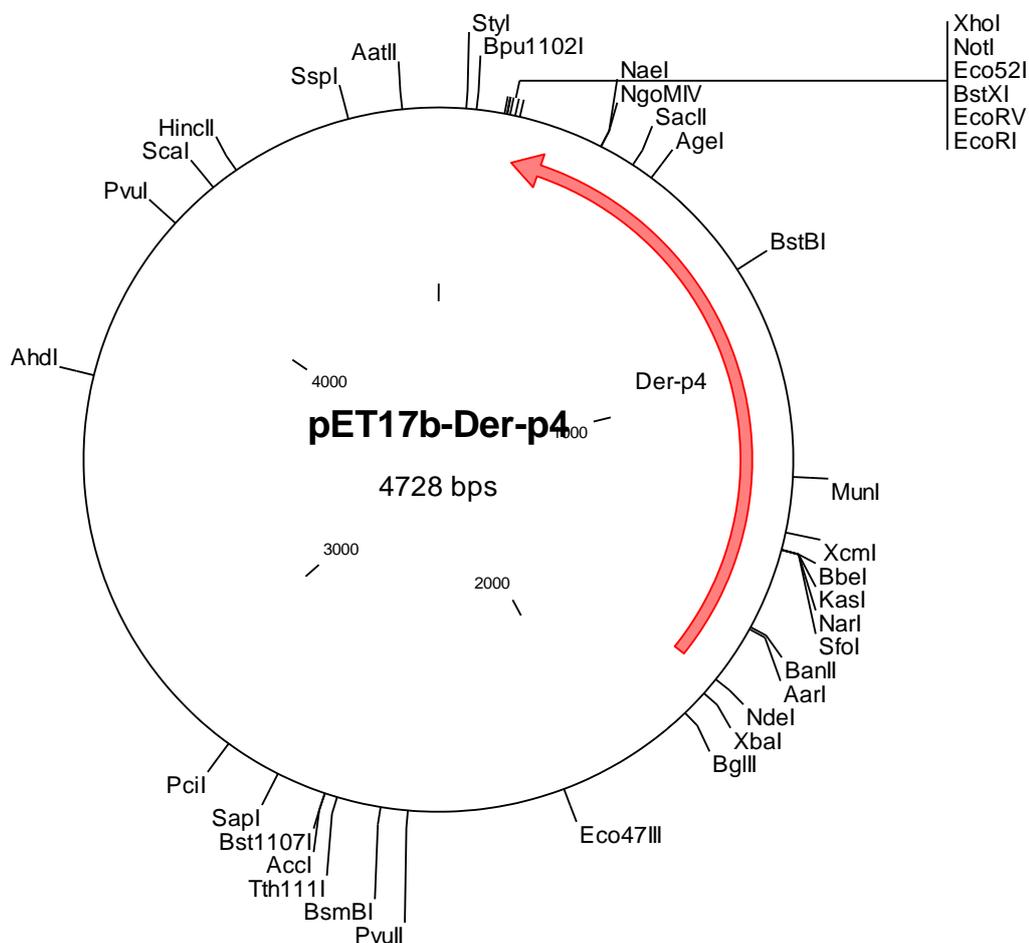
38. Urzhumtseva, L., P. V. Afonine, P. D. Adams, and A. Urzhumtsev. 2009. Crystallographic model quality at a glance. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 65: 297–300.

39. Maurus, R., A. Begum, L. K. Williams, J. R. Fredriksen, R. Zhang, S. G. Withers, and G. D. Brayer. 2008. Alternative catalytic anions differentially modulate human  $\alpha$ -amylase activity and specificity. *Biochemistry* 47: 3332–3344.

## Supplemental

**Plasmids are delivered dry or in ddH<sub>2</sub>O. Please use directly or resuspend the DNA in 20 ul autoclaved ddH<sub>2</sub>O. If needed, 0.2ul plasmid DNA can be used to transform E.coli.**

**Sequence chromatogram will be send by email.**



Synthetic gene was cloned into EcoRI/NdeI digested pET17b.

Open reading frame orientation as illustrated. Plasmid map can be reconstructed using the gene bank style txt file provided (sent by e-mail) and free software of **pDRAW32**: <http://www.acaclone.com> or directly viewed using **Clone Manager 5** or newer version (<http://www.scied.com>).