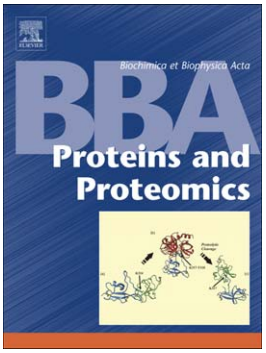


Role of *N*-terminal region of *Escherichia coli* Maltodextrin Glucosidase in folding and function of the protein

Ashutosh Pastor, Amit K. Singh, Prakash K. Shukla, Md. Javed Equbal, Shikha T. Malik, Tej P. Singh, Tapan K. Chaudhuri

PII: S1570-9639(16)30117-0
DOI: doi: 10.1016/j.bbapap.2016.06.008
Reference: BBAPAP 39768

To appear in: *BBA - Proteins and Proteomics*

Received date: 14 January 2016
Revised date: 10 June 2016
Accepted date: 14 June 2016

Please cite this article as: Ashutosh Pastor, Amit K. Singh, Prakash K. Shukla, Md. Javed Equbal, Shikha T. Malik, Tej P. Singh, Tapan K. Chaudhuri, Role of *N*-terminal region of *Escherichia coli* Maltodextrin Glucosidase in folding and function of the protein, *BBA - Proteins and Proteomics* (2016), doi: 10.1016/j.bbapap.2016.06.008

Role of N-terminal region of *Escherichia coli* Maltodextrin Glucosidase in folding and function of the protein

Ashutosh Pastor[a], Amit K. Singh[a,1], Prakash K. Shukla[b], Md. Javed Equbal[a], Shikha T. Malik[a], Tej P. Singh[b], Tapan K. Chaudhuri[a,*]

AUTHOR'S ADDRESS:

[a]Kusuma School of Biological Sciences, Indian Institute of Technology Delhi, New Delhi 110016, India

[b]Department of Biophysics, All India Institute of Medical Sciences, New Delhi 110029, India

[1] Present address: Max Plank Institute of Biochemistry, Am Klopferspitz 18, 82152 Martinsried, Germany

*CORRESPONDING AUTHOR: Dr. Tapan K. Chaudhuri,

Email: tkchaudhuri@bioschool.iitd.ac.in, tapanchaudhuri@hotmail.com, Tel: 91-11-2659-1012, Fax: +91-11-2659-7530

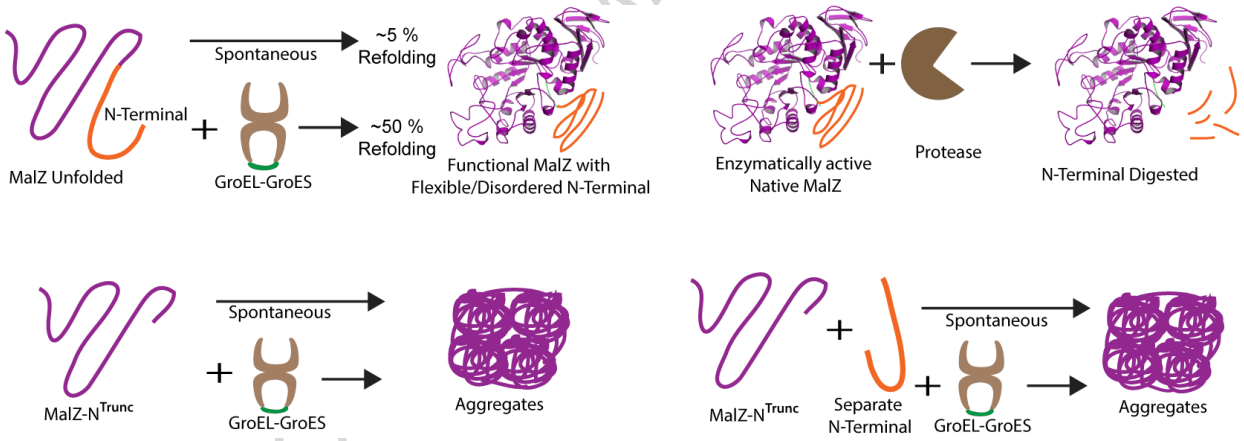
KEYWORDS: Protein Folding, GroEL assisted protein folding, Protein structure and function, Maltodextrin glucosidase, X-ray crystallography

ABBREVIATIONS: MalZ, Maltodextrin glucosidase; MalZ N^{Trunc}, MalZ protein without N-Terminal domain; $MalZ^{NTD}$, N-terminal domain of MalZ; GdnHCl, Guanidine Hydrochloride; IPTG, Isopropyl β-D-1-thiogalactopyranoside

HIGHLIGHTS:

1) MalZ, an aggregation prone protein requires chaperonins GroEL-ES for refolding.
2) Sequence & modelling analysis of MalZ suggest an unstructured N-terminal domain.
3) Crystal structure & limited proteolysis experiments confirm a flexible N-terminal.
4) MalZ N^{Trunc} does not fold in the presence of chaperonins or separate N-terminal.
5) The N-terminal plays an important role in the folding and stability of MalZ.

GRAPHICAL ABSTRACT



ABSTRACT

Maltodextrin glucosidase (MalZ) hydrolyses short malto-oligosaccharides from the reducing end releasing glucose and maltose in *E.coli*. MalZ is a highly aggregation prone protein and molecular chaperonins GroEL and GroES assist in the folding of this protein to a substantial level. The N-terminal region of this enzyme appears to be a unique domain as seen in sequence comparison studies with other amylases as well as through homology modelling. The sequence

and homology model analysis show a probability of disorder in the N-Terminal region of MalZ. The crystal structure of this enzyme has been reported in the present communication. Based on the crystallographic structure, it has been interpreted that the N-terminal region of the enzyme (Met1–Phe131) might be unstructured or flexible. To understand the role of the N-terminal region of MalZ in its enzymatic activity, and overall stability, a truncated version (Ala111-His616) of MalZ was created. The truncated version failed to fold into an active enzyme both in *E.coli* cytosol and *in vitro* even with the assistance of chaperonins GroEL and GroES. Furthermore, the refolding effort of N-truncated MalZ in the presence of isolated N-terminal domain didn't succeed. Our studies suggest that while the structural rigidity or orientation of the N-terminal region of the MalZ protein may not be essential for its stability and function, but the said domain is likely to play an important role in the formation of the native structure of the protein when present as an integral part of the protein.

1. INTRODUCTION

The protein Maltodextrin glucosidase (EC no. 3.2.1.20, Uniprot ID: P21517), expressed by malZ gene, is an enzyme involved in the maltose utilization system in *Escherichia coli* [1]. The major function of this enzyme is to catalyse hydrolysis of short malto-oligosaccharides ranging from maltotriose to maltoheptose, while releasing glucose from the reducing end. The final end products of this reaction are glucose and maltose [2]. The enzyme is known to release maltose directly from maltodextrins longer than maltotriose to a small extent [3]. MalZ can also hydrolyze γ-cyclodextrins containing eight glucosyl residues, however, γ-cyclodextrins are not physiological substrates for this enzyme as they are not transported by *E.coli* [3]. MalZ has been demonstrated to show trans-glycosylase activity which leads to formation of branched oligosaccharides as products [4]. The alternate name commonly used for such enzyme is alpha-

glucosidase. MalZ is a cytosolic enzyme in *E.coli*, and is considered to play a role in regulating the intracellular level of maltotriose, which induces *mal* regulon. Overexpression of malZ results in decreased expression of other genes in the *mal* regulon [2]. In *E.coli*, MalZ is known to counteract the formation of long maltodextrins, which are formed by action of amylomaltase. The mutants where malZ is knocked out produce large amounts of glycogen using maltodextrins as a source of carbon in absence of glycogen synthase [5, 6]. MalZ belongs to the α-amylase family of proteins. The α-amylases are known to play active role in starch metabolism by hydrolysing the α-1,4 and α-1,6 glycosidic linkages as well as performing transglycosylation and hydrolysis of cyclodextins. These enzymes contain an $(\alpha/\beta)_8$ tim-barrel structure which serves as the catalytic domain [7]. While the α-amylases are primarily involved in starch metabolism, the enzymes showing sequence similarity to MalZ are known to be cytosolic enzymes specific to smaller malto-oligosaccharides and play active roles in maltose metabolism in bacterial cells [8, 9]. The microbial α-amylases have been studied extensively for their importance in various industrial applications like food industry, starch industry, textile, detergent, pharmaceutical and paper industry [10].

Maltodextrin glucosidase is also found in other organisms including *Salmonella enterica, Shigella flexneri, Yersinia pestis* and others. The crystal structure of this protein has not been reported in any of these organisms. On the basis of sequence similarity the closest structure known to this protein is alpha-amylase 1 from *Thermoactinomyces vulgaris* [11] having sequence identity of 34% with a sequence coverage of 42-591 amino acids out of 616 total amino acids. The protein in *E.coli* consists of 604 amino acids, two amino acids at N-terminal and 10 amino acids including 6X His-tag at C-terminal appear in the studied protein making the total

number of amino acids to 616. However, no effect of these extra amino acids has been observed on the function of the protein.

It has been reported that MalZ protein is highly aggregation prone and requires assistance of chaperonins GroEL and GroES for its folding when over-expressed in *E.coli* [12]. The co-expression of these chaperonins with MalZ in the *E.coli* cells increases the soluble and functional fraction of MalZ protein in the cytosol by a very significant amount. GroEL and GroES usually help the aggregation prone proteins to fold by encapsulating the protein in the cylindrical cavity of GroEL which is capped by GroES to form a closed chamber [13]. However, MalZ is a large protein having a molecular weight of 70 kDa and cannot be encapsulated within the GroEL-GroES cavity, which can encapsulate protein of size up to ~50 kDa. GroEL assisted folding through a *trans* mechanism [14] has been suggested for the folding of this protein. In this case, it is expected that only a part of the protein might be entering the GroEL cavity and GroES does not encapsulate it, but helps in the release of the protein from the cavity by binding to the *trans* ring of GroEL [15]. The role of any specific part of the MalZ protein in this aggregation prone behaviour and requirement of GroEL and GroES for assisted folding was not clear and the present study is an attempt to answer this question.

The initial sequence analysis and homology model comparisons made it interesting to study the N-terminal of MalZ. However in order to get precise information on its uniqueness among amylase type of proteins and also to understand the reasons for the aggregation prone behaviour and chaperone dependency for folding, structure determination of MalZ using X-ray crystallography was attempted. The crystal structure of MalZ was solved at 3.7Å and has been reported here. The structure of MalZ obtained from crystallography data suggested that the N-terminal region of MalZ (Met1 – Phe131) may either be flexible or unstructured. This was

interesting as the sequence comparison and the homology model data also showed that the N-terminal might be a distinct entity. To find out whether this N-terminal region actually plays any important role in the folding and structural stability of the MalZ protein the N-terminal region (Met1-Phe110) was truncated, and the results suggest that this domain is essentially required within the protein for stabilizing the structure of MalZ in its native functional form and the N-terminal domain does not help to fold MalZ when used *in trans* through co-expression in *E.coli* or *in vitro*.

2. Materials and Methods

2.1 Sequence and homology model analysis

The amino acid sequence of the protein used for our studies was compared to the Uniprot database using NCBI BLAST [16] and the specific domains and their families were identified based on the sequence. Further the homology model was prepared using the I-Tasser server [17] which combines template based prediction with some simulation based refinement. The visualization software Pymol was utilized for visualization and comparisons. The sequence was also analysed using a web server MobiDB [18] for disorder prediction which combines results from multiple servers to generate a consensus on disordered regions.

2.2 MalZ expression and purification

Plasmid PCS19malZ containing the 1.8Kb malZ gene under the control of a T5 promoter and ampicillin resistance marker was a generous gift from Prof. Winfried Boos, (University of Konstanz, Germany). The malZ gene is cloned between sites of restriction endonucleases NcoI and BamHI and it forms a C-terminal 6X His tag from the PCS19 vector. *E.coli* BL-21 cells containing plasmid PCS19 and were grown in Luria Bertani medium with 50 µg/ml ampicillin to $O.D._{600}$ of 0.6. Induction was done by 1mM IPTG. Post induction, the culture was kept at 37º C

for 8 hours. The cells from 1L culture were harvested by centrifugation at 7500g for 15 minutes and temperature was maintained at 4^{o}C from this step onwards. The pellet was re-suspended in 30ml lysis buffer consisting of 20mM sodium phosphate pH7.4, 500mM sodium chloride, 1mg/ml lysozyme and 1mM PMSF. Cell lysis was carried out by sonication with a Branson Sonifier250 (USA). Fifteen cycles of 25 seconds each were given with 1 minute rest phase. AKTA Purifier (GE Healthcare, UK) was used for the purification process. The supernatant was loaded on a 5ml His-Trap column (GE Healthcare, UK) pre-equilibrated with the equilibration buffer containing 20mM sodium phosphate pH7.4 and 500mM sodium chloride. The column was washed extensively by same buffer containing 20mM imidazole for removal of impurities. Resin bound protein is eluted by a linear gradient up to 500mM imidazole in 10 column volumes. MalZ is eluted at imidazole concentration of around 200mM. Fractions eluted were found to contain single bands on SDS PAGE gels stained with Coomassie blue. Protein concentration was measured using Bradford assay (Bio-rad, USA).

2.3 Crystallization of MalZ, data collection and processing

The pure fractions of MalZ were concentrated with amicon centrifugal concentrators and in this process, the salt and imidazole were also reduced. The protein was concentrated to 8 mg/ml and the final buffer composition was 20mM sodium phosphate, 200mM Sodium chloride and 50mM imidazole. The MalZ solution was mixed with reservoir solution 2µl each, which contained 2.5M ammonium acetate and 0.1M sodium acetate at pH4.6. This solution was used for crystallization using the hanging drop vapour diffusion method. The crystallization set ups were kept at room temperature (298 K). The crystals grew to an approximate size of 0.4×0.1× 0.1 mm^3 with sharp rectangular faces after 10 days. The crystals were stabilized in 2.5M ammonium acetate and 0.1M sodium acetate containing 20% Glycerol (v/v) for data collection at low temperatures. A

single crystal was mounted in a nylon loop and flash - frozen in a stream of nitrogen gas at 100 K. The data were collected on a MAR CCD-225 Scanner (Marresearch, Norderstedlt, Germany) using the beamline, BM14 at the European Synchrotron Radiation Facility (ESRF), Grenoble, France.

2.4 Structure determination and refinement

Structure of MalZ was determined with molecular replacement method using auto-AMoRe [19] from the CCP4 software suite (Collaborative Computational Project, Number 4, 1994). The coordinates of the structure of α-amylase from *Thermoactinomyces vulgaris* (PDB code: 1JI1, [11]) were used as search model. The rotation and translation functions calculated with data in the resolution range of 110.6 – 3.7 Å yielded a unique solution with the first peak being very distinct. The packing arrangement of the molecules in the unit cell for this solution gave no unfavourable intermolecular contacts in space group P4 thus confirming it to be the correct space group and the correct solution. These coordinates were transformed using AMoRe and were subjected to 20 cycles of rigid-body refinement with REFMAC5 [20] The manual model building of the protein was carried out using $|2F_o - F_c|$ Fourier and $|F_o - F_c|$ difference Fourier maps with Graphics Program 'O' [21] on a Silicon Graphics O2 Workstation. The structure was partially refined due to limitation of reflection data **(Table 1)** to values of 0.33/0.34 for R_{cryst}/R_{free} factors.

2.5 Measurement of enzymatic activity and fluorescence in crystallization conditions

Solution of concentrated MalZ protein was mixed in equal amounts of crystallization solution of 2.5M ammonium acetate and 100mM sodium acetate at pH4.6 as used for setting up the crystallization. The crystallization condition was maintained for about 1 hour and 4 hours, respectively, and then the enzymatic activity was measured. It was compared with the enzymatic

activity of MalZ protein in 20mM sodium phosphate with 200mM sodium chloride and 50mM imidazole at pH 7.4. Fluorescence emissions were also recorded at different pH values ranging from 8 to 4.6. The fluorescence scans were done on Agilent Cary Eclipse fluorimeter and the excitation wavelength was kept at 290 nm while the emission was recorded in the range 300-400nm.

2.6 Limited proteolysis of native MalZ.

The proteolytic cleavage sites for trypsin and proteinase K were determined using Expasy peptide cutter server [22] **(Supplementary Table 2)**. Trypsin was purchased from Sigma Aldrich co. LLC (USA) and proteinase K from Promega (USA). MalZ protein at a final concentration of 25µM was incubated with 4µM trypsin and 5µM proteinase K separately in 20mM sodium phosphate buffer at pH 7.5 containing 200mM sodium chloride and 50mM imidazole. The digestion was performed at 30ºC and samples were taken at time points of 30s, 1min, 2min, 5min and 10min. The samples were mixed with SDS gel loading buffer containing PMSF and were heated to 95ºC to inactivate the proteases. These were loaded onto SDS PAGE and the gels were stained using coomassie blue after the completion of run. The bands on gel corresponding to the expected size were cut and an in-gel digestion with trypsin was performed according to previously described protocol [23]. The samples were loaded onto LC-MS and the peptide mass were compared to the trypsin digestion profile of MalZ.

2.7 Cloning of MalZ N^{Trunc} and MalZ N-terminal domain

A PCR based selective amplification of the malZ gene in the region of amino acids Ala111 – His616 was done, thus truncating the gene sequence corresponding to the first 110 amino acids (Met1 - Phe110). Plasmid PCS19 containing malZ gene was used as a template for the PCR

reaction. PCR was run on a Bio-Rad (USA) thermocycler and Pfu polymerase (Merck-Genei) was used for amplification. Thirty amplification cycles were used with an annealing temperature of 64^{o}C. The PCR product was eluted from the gel using a gel extraction kit (Advanced microdevices, India). PCS19 and the extracted PCR product was double digested using Fermentas (Thermo Fisher Scientific, USA) fast digest NcoI and BamHI at 37^{o}C for 20 minutes. The amplified gene was ligated back in the plasmid PCS19 containing a C-terminal 6X-His tag using T4 DNA ligase (Fermentas). The DH5α colonies obtained on ampicillin containing plates were inoculated in Luria Bertani broth and the cloned plasmid was purified by a plasmid miniprep kit (Qiagen, Netherlands).The plasmid was double digested to confirm the presence of cloned construct. Over-expression of the protein in BL-21 cells was confirmed and compared to the full length MalZ protein. The full length MalZ was cloned for a different study into pACYC-Duet1 vector at *Nco*I and *Bam*HI restriction sites to get the recombinant construct pACYC-D-malZ. To get a construct with only N-terminal domain of MalZ, the plasmid pACYC-D-malZ was digested with *Eco*RV (New England Biolabs) as there was an EcoRV site at the desired location for truncation. Another EcoRV site was present in the plasmid downstream and hence a self-ligation was done. The recombinant clone was selected on LA plate containing chloramphenicol. This construct was referred to as pACYC-D-malZ^{NTD}.

2.8 Solubility analysis by fractionation.

BL-21 cells transformed with plasmids containing pCS19 malZ and pCS19 malZ N^{Trunc} were grown and harvested. The pellet from 10 ml culture with O.D.$_{600}$ about 1.2 were re-suspended to 10 ml of lysis buffer containing 20mM sodium phosphate with 500mM sodium chloride and 100mM imidazole at pH 7.4. Cells were lysed by sonication and the lysate was centrifuged at 15000g for one hour. The supernatant was collected and the pellet was re-suspended in equal

amount of buffer after washing it twice. The lysate, supernatant and pellet fractions were loaded on SDS PAGE gel.

2.9 Co-expressions of different plasmids

Competent cells were prepared from the *E.coli* BL-21 cells containing the PCS19 malZ and malZ-N^{Trunc} plasmids. These sets of cells were further transformed with pGro7 vector (Takara Inc.) containing both groEL and groES genes under control of an arabinose inducible araC promoter and chloramphenicol resistant marker. The cells were grown in ampicillin (50µg/ml) and chloramphenicol (20µg/ml) containing LB media. Induction was done with 1mM IPTG and 2mg/ml arabinose. The constructs of PCS19 malZ-N^{Trunc} and pACYC-D-$MalZ^{NTD}$ were co-transformed in BL-21(DE3) cells. The co-transformed cells were grown on LB media at 37^{o}C with 50µg/ml ampicillin and 20µg/ml chloramphenicol for selection. The culture was kept for 8 hours after induction by 1mM IPTG.

2.10 Solubility determination by enzymatic activity assay.

MalZ enzymatic activity is monitored with the substrate p-nitrophenyl-α-D maltoside. Upon hydrolysis of the substrate by MalZ, p-nitrophenol is released, which gives a yellow colour and its quantification is done by measuring absorbance at 405nm [24]. The substrate was used at a final concentration of 0.5mM. The enzymatic activity of MalZ was measured from cell lysates where MalZ, MalZ-N^{Trunc}, MalZ N-terminal were over-expressed and where they were co-expressed with GroEL and GroES. The cells were lysed in 20mM sodium phosphate buffer containing 500mM Sodium chloride, 0.5mM PMSF and 1mg/ml lysozyme through sonication and activity assay was done from the whole cell lysates and the supernatant fractions which gave similar results. *E.coli* cell lysate without plasmid were taken as a control for background subtraction.

2.11 Purification of MalZ-N^{Trunc} and MalZ N-terminal

Purification of MalZ-N^{Trunc} was done in the denatured state. The over-expression conditions were the same as described above. A pellet of 1L culture was collected and cell lysis done as mentioned above. The lysate was centrifuged at 12000g in an Eppendorf table top centrifuge. The pellet was resuspended in 20mM phosphate buffer containing 1M urea and triton X-100 for pellet washing. It was again centrifuged and the pellet was resuspended in 20mM phosphate pH7.4 containing 8M urea for denaturation. The pellet was homogenized and left for 15-20 minutes to allow the denaturation of inclusion bodies. The solution was again centrifuged and the supernatant was loaded on a 5ml His-Trap column pre-equilibrated with 20mM sodium phosphate and 8M urea. After sample loading the column was washed for 10 column volumes of buffer having 20mM sodium phosphate pH 6.3 and 8M urea. Further, elution was done in 10 column volumes with a gradient of 0-500mM imidazole in buffer similar to equilibration buffer. The purity was checked on SDS PAGE. BL-21 DE3containing pACYC-D-$malZ^{NTD}$ were grown on LB media containing 20µg/ml chloramphenicol for selection. The culture was kept for 6 hours after induction by 1mM IPTG. About 50ml of the culture was taken and the cells were pelleted at 4^{o}C. The cell pellets were then re-suspended in 4ml of 20mM sodium phosphate buffer containing 200mM Sodium chloride, and lysed by sonication using a Branson sonifier 250 and 10 cycles at 1 minute intervals were sufficient to break the cells. The lysate was then centrifuged at 20,000g in a Sorvall RC6 plus centrifuge (Thermo Scientific) and the supernatant was collected. The samples were heated at different temperatures up to 80^{o}C for 5 mins. The initial temperature was kept at 50^{o}C and samples were incubated for 5 min, then centrifuged and supernatant collected and an aliquot of that sample was run on the SDS gel, the remaining supernatant was again treated at 55^{o}C for five minutes and in the same manner for higher

temperatures. The protein could be seen in the soluble fraction up to 65^{o}C and on reaching 70^{o}C it starts to aggregate.

2.12 *In vitro* refolding of MalZ and MalZ-N^{Trunc}

Refolding of purified MalZ and MalZ-N^{Trunc} were studied in terms of regain of enzymatic activity. MalZ was denatured by 4M GdnHCl for 20 minutes before setting up the refolding reaction while MalZ-N^{Trunc} was also transferred to buffer containing 4 M GdnHCl by buffer exchange. The buffer used for this experiment consisted of 20mM sodium phosphate, 10mM potassium chloride and 10 mM magnesium chloride at pH 7.4. Denatured MalZ and MalZ-N^{Trunc} were at an initial concentration of 5µM and these were finally diluted 100 times leaving the residual concentration of GdnHCl to 0.04M. The refolding was done at a final concentration of MalZ variants at 0.05µM. MalZ and MalZ-N^{Trunc} were diluted in one set containing refolding buffer only and two sets of buffers containing GroEL. After a 15 minute incubation 0.2µM GroES and 5mM ATP were added to the fractions containing 0.1µM GroEL. The reaction mixture was kept for 2 hours at room temperature and then MalZ enzymatic activity was measured. The *in vitro* refolding experiment of MalZ N^{Trunc} with isolated MalZ N-terminal was performed with same refolding buffer as above. Multiple sets of refolding conditions containing different stoichiometric ratios of isolated MalZ N-terminal and truncated MalZ protein were used. The ratio of MalZ N^{Trunc}: MalZ N-terminal from 1:1 to 1:100 was analysed and GroEL was taken in an equimolar ratio to the isolated N-terminal part. The refolding buffer consisted of the isolated MalZ N-terminal and GroEL was added initially and then GroES and ATP were added after 15 minutes of addition of MalZ N^{Trunc} to the refolding solutions. The refolding reaction was allowed to proceed for 1 hour at room temperature and then enzymatic activity was measured.

2.13 The comparison of enzymatic activity after limited proteolysis.

The native MalZ protein was subjected to limited proteolysis by trypsin as mentioned previously. The digested product was taken out at 15min, 30min, 45 min and 60 min and was loaded on gel and analysed for enzymatic activity according to the mentioned protocol. The enzymatic activity readings were measured after a 7 min incubation with the substrate. The band intensities on gel were quantified and compared by Bio-Rad Image lab software. Hydrophobic interaction chromatography was used for the separation of the trypsin digestion products. Limited proteolysis was set for 45 minutes at 30ºC and ammonium sulphate to a final concentration of 1.2M was added to the reaction mixture. The mixture was loaded on a manually packed 1ml source15ISO column (G.E. Healthcare, U.K.) pre-equilibrated with 20mM sodium phosphate buffer pH7.4 containing 1.2M ammonium sulphate. Elution was done in 10 column volumes by gradually decreasing the ammonium sulphate content in the buffer. The eluted fractions were analysed on SDS PAGE, the concentration was measured by Bradford assay and enzymatic activity was measured as mentioned previously.

2.14 Circular dichroism of MalZ N-terminal domain

Purified MalZ N-terminal domain was diluted to a final concentration of 5µM in 10mM sodium phosphate buffer containing 20mM sodium chloride and having pH 7.4. The circular dichroism spectra was recorded in the range of 200-300nm on a Jasco J-815 CD polarimeter. The full length MalZ was used at a concentration of 4µM. The CD spectrum of full length MalZ was recorded in 10mM sodium phosphate containing 20mM sodium chloride and 1mM imidazole. The spectra were analysed for secondary structure prediction using the inbuilt algorithm in the instrument's software.

3 RESULTS

3.1 Sequence comparison of MalZ with other amylases

The crystal structure of MalZ protein was not available and sequence comparison studies were carried out initially to get some insight into the structural properties of this protein. The sequence of MalZ protein was compared with proteins in the Uniprot database to find sequence similarities with known protein families. Protein BLAST was run using the NCBI server [16] and the results obtained showed that the region corresponding to amino acids Met1-Phe110 is an E-set superfamily type domain which usually consists of two β sheets with three or four β strands in antiparallel direction in each sheet (**Figure 1a**). This domain is present in some proteins including some amylases like maltogenic alpha-amylase 1 from *Thermoactinomyces vulgaris* [11], however, the function of this domain is not very clear. There is no significant sequence similarity of the MalZ N-terminal domain with these E-set domains in other amylase family proteins whose structure is available in PDB (**Supplementary Table 1**). The region of protein corresponding to the amino acids Gln120-Glu540 represents Alpha amylase superfamily and most of the functional sites including the active sites are present in this region. The sequence comparison studies show that the active site of MalZ should consist of three residues Asp338 as nucleophile, Glu375 as proton donor and Asp450 as transition state stabilizer. These have been determined by the similarity of residues in active sites of other alpha amylases and have been mentioned in the Uniprot database accession number 21517.

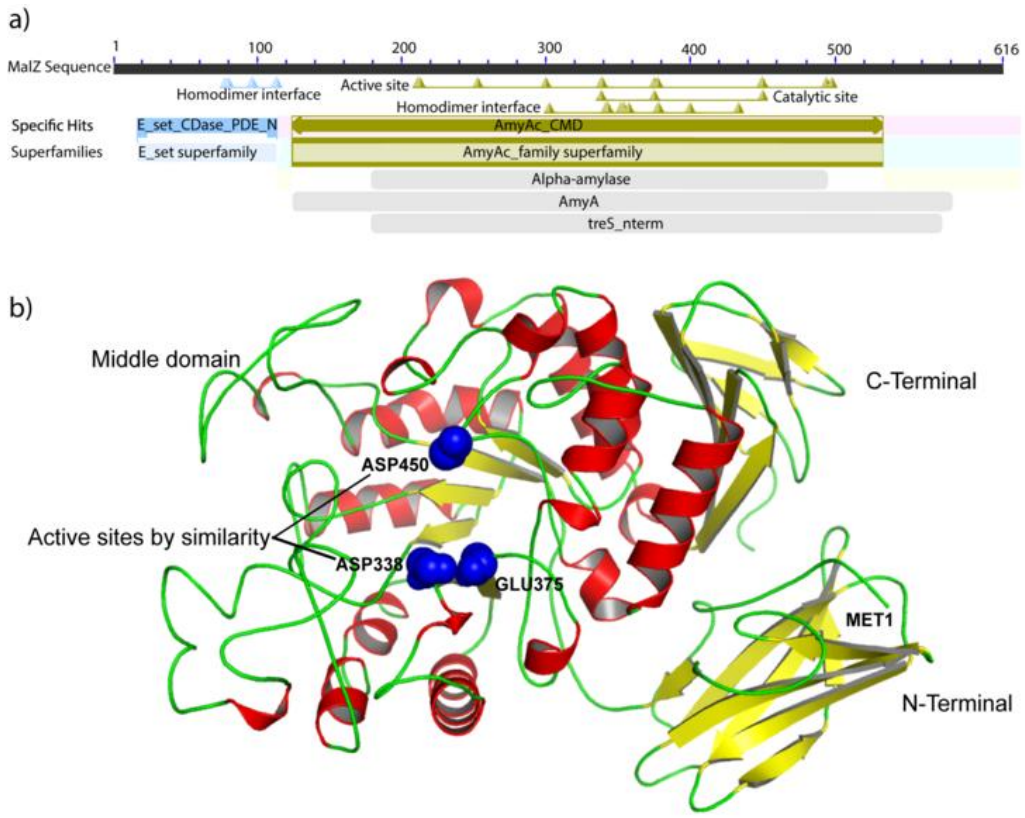


Figure 1: **Sequence comparison of MalZ and homology model**
a) NCBI Protein BLAST results of MalZ with the non-redundant protein database identify the different regions of the protein. The first 120 amino acids have similarity to the E-Set Super family while the rest of the protein shows high similarity with the Amylase super family of proteins; **b)** The homology model of MalZ shows a distinct N-terminal domain which also corresponds to the BLAST results and the structure of the N terminal shows an E-Set immunoglobulin like domain with two β-sheets of antiparallel β-strands.

3.2 Homology model of MalZ

Homology model of MalZ was prepared using I-tasser server [17]. The Ramachandran plot for the homology model shows 71.3% residues in most favoured regions, 23.4% residues in additionally allowed regions and 2.8% and 2.5% residues respectively in generously allowed and disallowed regions. The C-score for the predicted model is 0.38, which shows that the model is statistically reliable. The Z-score for the alignment with the major templates was significant, it was 5.64 with the α-amylase 1 from *Thermoactinomyces vulgaris* with a 34% sequence identity and 91% sequence coverage [25] and 8.41 with *Bacillus stearothermophilus* neopullanase with a

33% sequence identity and a coverage higher than 80% [26]. The model shows that the N-terminal domain and the C-terminal domain are present as distinct regions in the protein (**Figure 1b**). The middle part which corresponds to the amylase super family region comprises the major domain of the protein containing the active sites. The N terminal region in the model primarily consists of β-sheets. While it is not very accurate to predict the inter-domain interactions from the homology model, however the idea derived was that a separate domain exists. Also, no significant sequence homology was found to the N-terminal region alone when its sequence (Met1-Phe131) was separately compared to all the proteins having structural data in the PDB.

3.3 Prediction of flexibility or disorder in the N-terminal domain

The homology model prediction server predicted two structures with C-Scores of 0.38 and 0.33 and these two models were aligned and compared. Comparison of these models and models from other servers indicated that while the overall topology of the molecule remains the same, there is significant variation in the N-terminal region **(Figure 2a)**. Based on this it was assumed that the N-terminal might exist in different conformations. Further, a disorder prediction server, MobiDB [19] was utilized to analyse the disorder in the protein based on the sequence information. The server gives a consensus results based on multiple disorder prediction algorithms from Dis-EMBL [27], ESpritz [28], IUPred [29] and JRONN [30]. The result indicated that the region between amino acids Ile35 to Ile68 is having a high disorder propensity and it was predicted in consensus by all algorithms **(Figure 2b)**. The analysis of the results obtained from the homology modelling together with the comparison of protein sequences indicated a role of this terminal in the stability and the function of the MalZ protein. Further, an attempt was also made to get more precise information on the structure of the protein by determining the crystal structure of MalZ.

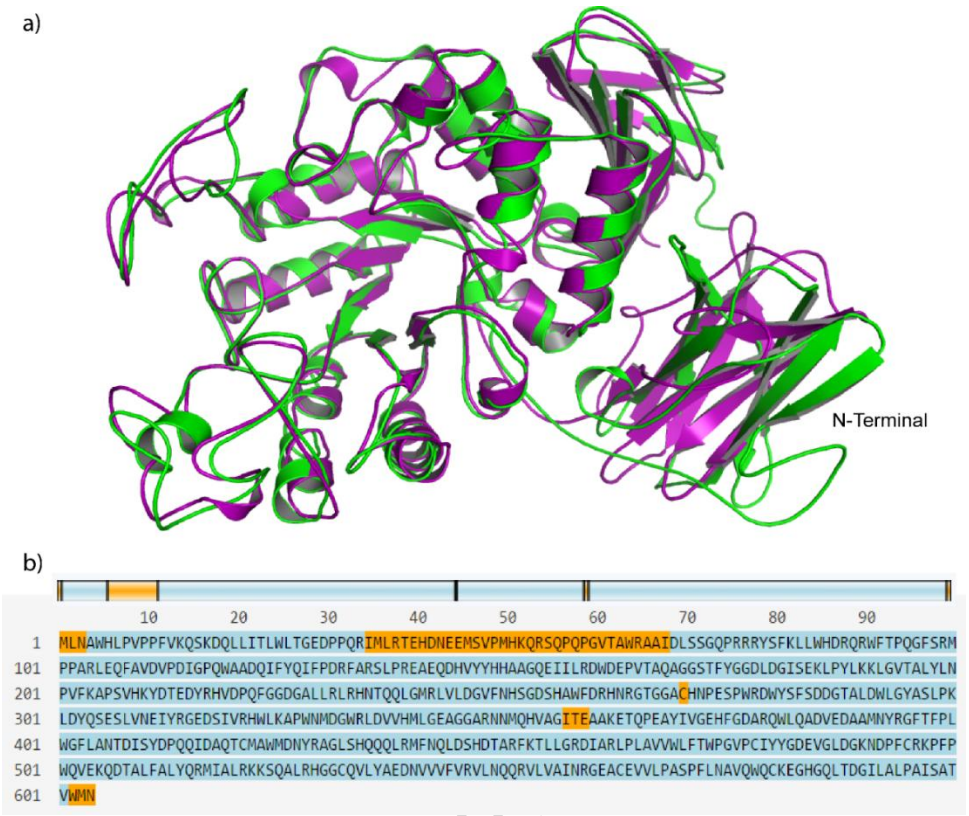


Figure 2: Prediction of disorder in the N-terminal region
a) Comparison of two predicted homology models of MalZ protein showing the difference of conformation in the N-Terminal domain. The rest of the protein shows significant alignment; **b)** the results of disorder prediction servers showing a disorder prone region in the N-terminal between amino acids Ile36 to Ile68.

3.4 Crystallization of MalZ

The purified samples of MalZ were dissolved in the solvent containing 20mM sodium phosphate, 200mM sodium chloride and 50mM imidazole at a concentration of 20mg/ml. A reservoir solution was prepared with 2.5M ammonium acetate and 0.1M sodium acetate at pH 4.6. The protein solution was mixed with the reservoir solution in equal parts and kept for crystallization at room temperature for 10 days. Although the crystals of MalZ appeared to be good morphologically **(Figure 3b)**, the quality of diffraction was poor and the resolution was restricted to 3.7Å. The repeated attempts to improve the quality of crystals did not yield better results. The reflection data were processed with DENZO and SCALEPACK from the HKL

package [31]. The crystals belong to tetragonal space group P4 with unit cell dimensions, a = b = 110.6Å, c = 69.5 Å. The presence of one molecule per asymmetric unit gave a crystal volume per protein mass (Vm) of approximately 3.0 $Å^3Da^{-1}$ corresponding to a solvent content of 59%. The results of data collection and processing are given in **Table 1**

3.5 Crystal structure of MalZ

Structure of MalZ was solved and refined at 3.7 Å resolution. The coordinates of the structure of MalZ have been deposited in the PDB with accession number 5BN7. Although the refinement of the structure was not very rigorous due to the limitations of reflection data, the gross features of the structure were clear **(Figure 3d)**. The Ramachandran plot of the structure showed that 67% residues were present in the most favoured regions while 31% residues were found in the additionally allowed regions. There were 2% residues in the generously allowed regions. There was no electron density for the N-terminal region up to residue number Phe131 in the protein chain and hence it could not be defined in the crystal structure. In order to confirm whether the N-terminal region was present in the protein structure and disordered or it was altogether absent from the protein chain, the protein samples from the crystals were examined by dissolving the crystals in water. The SDS-PAGE from these samples **(Figure 3c)** showed a single band and had a molecular weight identical to that of the intact MalZ protein confirming that the protein molecule in the crystals corresponded to that of the intact protein.

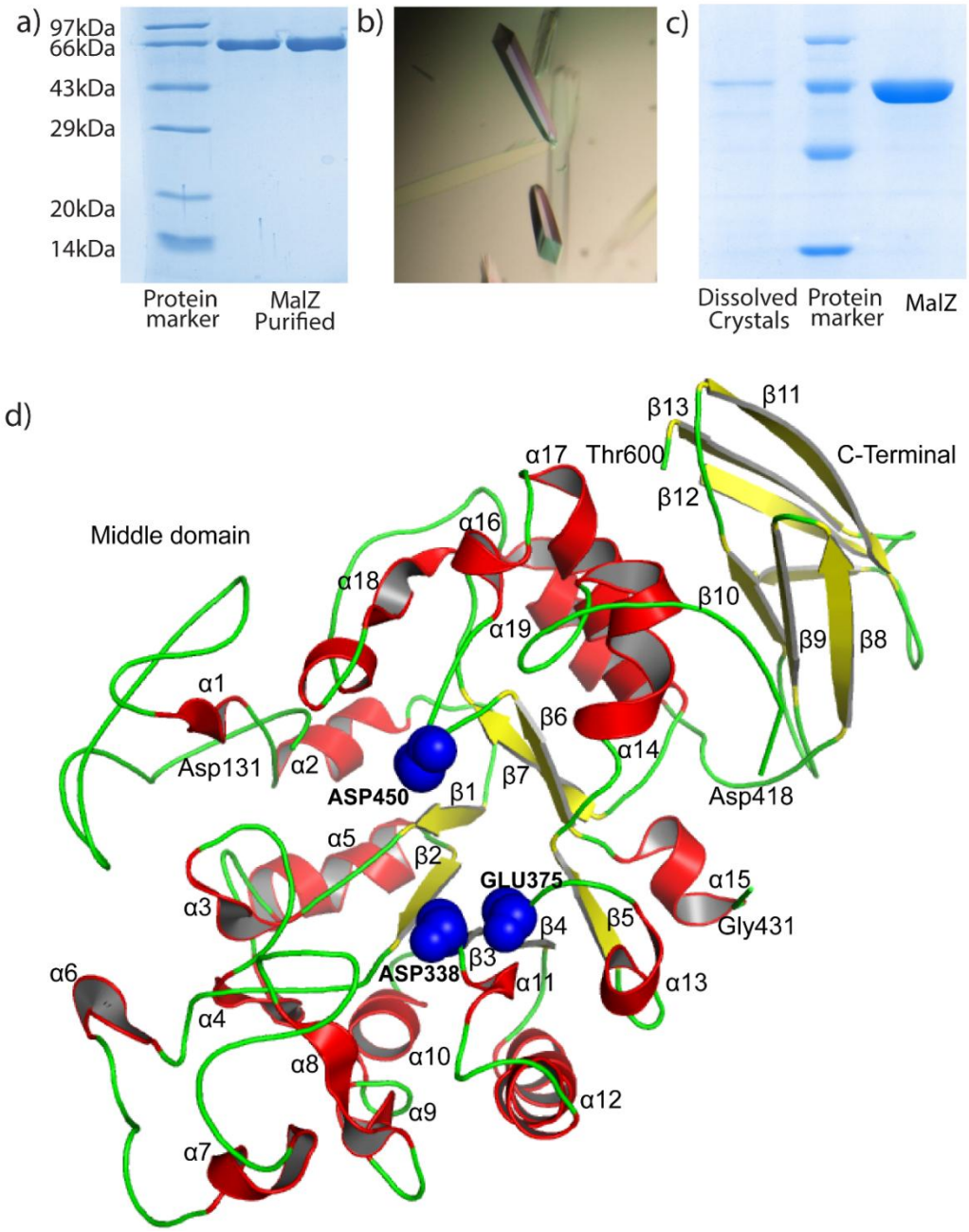


Figure 3: **MalZ purification, Crystallization and structure determination**
a) Purified MalZ protein. SDS PAGE showing a band of size 70 kDa; **b)** Maltodextrin glucosidase crystals after 10 days of growth at room temperature; **c)** Gel showing the presence of intact MalZ protein in the crystals. The left lane consists of protein sample from dissolved crystals, while the right lane shows purified intact MalZ for comparison. The middle lane has medium range protein markers. The gel image shows MalZ protein present in the crystals to be intact in terms of molecular weight; **d)** Structure of MalZ solved at 3.7Å resolution. The first 130 residues from N-terminus were not observed in the electron density. The density for residues

419-430 was also not observed. The secondary structure elements, α-helices and β-strands are labelled.

Table 1. Data collection and refinement statistics

PDB	5BN7
Space group	P4
Unit-cell parameters (Å)	
a=b	110.5
c	69.5
Number of molecules in the asymmetrical unit	1
V_m (Å^3/Da)	3.01
Solvent Content (%)	59.53
Resolution range (Å) 110.61-3.7	
Number of total reflections	31009
No. of unique reflections	6156
Overall completeness (%)	70.4
Completeness in the highest shell (3.80 – 3.70 Å)	73.9
R_{sym}(%)	9.7
R_{sym}in the highest shell	72.2
I / σ	2.5
I / σ in the highest shell	1.2
R_{cryst}(%)	32.9
R_{free} (5 % data) (%)	34.4
Number of protein atoms	3618
R. m. s. deviations	
Bond length (Å^2)	0.015
Bond angles (^{o})	2.4
Dihedral angles	22.3
Mean B factor (Å^2)	
Wilson B factor	88.0
Main chain atoms	131.93
Side chains and water molecules	130.2
Overall	131.0
Ramachandran plot statistics	
Residues in the most favored regions (%)	66.4
Residues in the additionally allowed regions (%)	30.7
Residues in the generously allowed regions (%)	2.3
Residue in the disallowed regions	0.5

This led us to infer that the N-terminal region (Met1 to Phe131) of MalZ is either highly flexible or it adopted several conformations that existed statistically as reported for other proteins not showing electron density in certain regions [32]. It might also be possible that the domain is in unfolded state or can be in different orientations. All these states might have resulted in the absence of electron density. It may also be mentioned here that the middle region of MalZ contains the active site residues **(**Asp338 as nucleophile, Glu375 as proton donor and Asp450 as transition state stabilizer**)** of the protein as identified from similar studies of various other α-amylases. This central region appears to be a α, β Tim-barrel structure, which is also a common structural feature of several other amylases [33]. This type of structure is also found in many other substrates of the chaperonins GroEL and GroES [34, 35]. Hence, it may correlate to the importance of GroEL and GroES for the assistance in the folding of MalZ protein to a native functional state.

3.6 MalZ protein retains its conformational identity under crystallization conditions

The N-terminal region of MalZ protein was not visible in the electron density map of the molecule, and it was assumed to be statistically disordered or flexible. Since the crystals were prepared with a condition having pH4.6 which is a much lower pH than the working pH for the protein ~ 7.4, it could have been suspected to be a reason for the non-uniform structure of the N-terminal. In order to address this issue, the MalZ protein was subjected to exact crystallization conditions and the enzymatic activity and fluorescence emission were measured. The enzymatic activity was almost similar in crystallization conditions as to the physiological pH conditions when the samples were kept in these conditions for 1 to 4 hours **(Figure 4a).** The N-terminal region of the protein contains five tryptophan residues out of total 22 tryptophan residues present

in the MalZ protein which is almost a fourth part of the total number. So it was expected that a difference could be observed in tryptophan fluorescence emission intensity of the protein in case the N-terminal region changed / deleted due to the lowering of pH under crystallization conditions. However, there was no significant change in the tryptophan fluorescence intensity as well as the wavelength maxima in the crystallization conditions **(Figure 4b).** These results suggested that the N-terminal might be intrinsically disordered and the role of lower pH or crystallization conditions in altering properties of N-terminal region could be ruled out.

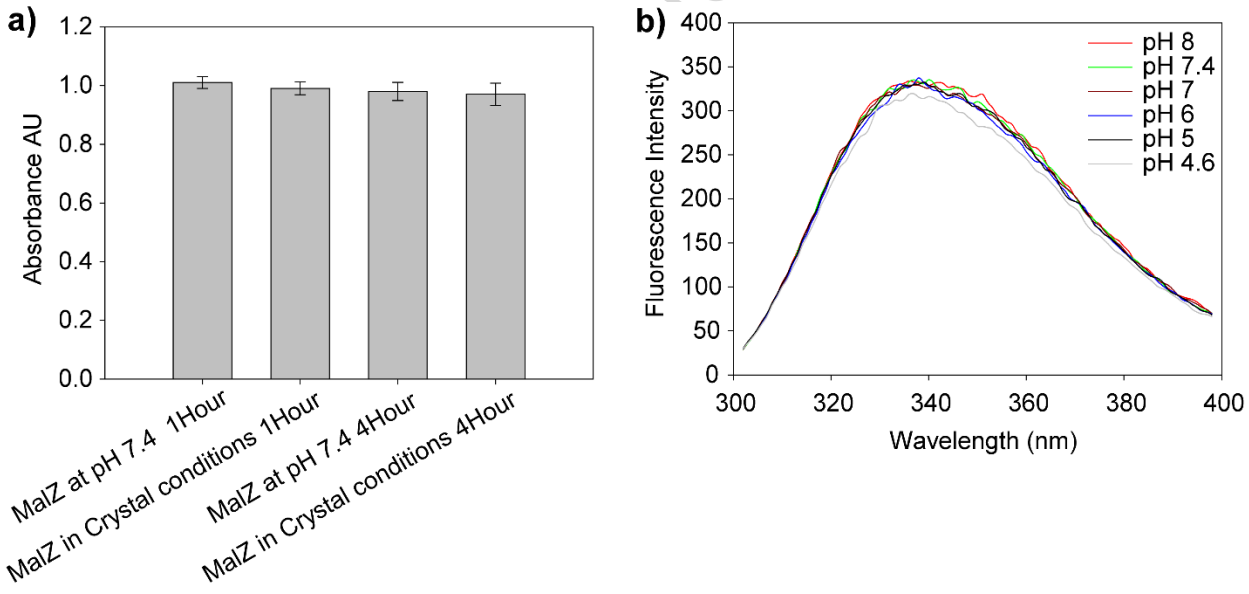


Figure 4. Stability of MalZ protein at crystallography conditions
a)Acitivity assay of MalZ after incubation for 1 hour and 4 hour in the conditions used for preparing crystals; **b**) Fluorescence spectra of MalZ after 30 minutes incubation of MalZ native protein at different pH ranging from 8 to 4.6. The figure shows that there is no effect of the crystallization conditions in the overall conformation of the protein and the enzymatic activity and fluorescence spectra remains largely similar in these conditions.

3.7 Limited proteolysis cleaves the N-Terminal of MalZ

Limited proteolysis was utilized to find the flexible or unstructured sites in the protein [36, 37]. The assumption was that if the N-terminal site would be flexible at the linker or unstructured, it

would be digested faster than the folded part of the protein. The native and functional state of MalZ was subjected to limited proteolysis by trypsin **(Figure 5a)** as well as proteinase K **(Figure 5b)**. The sites of cleavage were identified by Expasy peptide cutter. MalZ contains substantial number of truncation sites for both these enzymes **(Supplementary table 2).**The gel run after proteolysis for different time durations show a prominent band at 54 kDa, which is expected to be remaining after a truncation of about 130 amino acids. This was analysed by the molecular weight analysis program Image lab of Bio-rad.

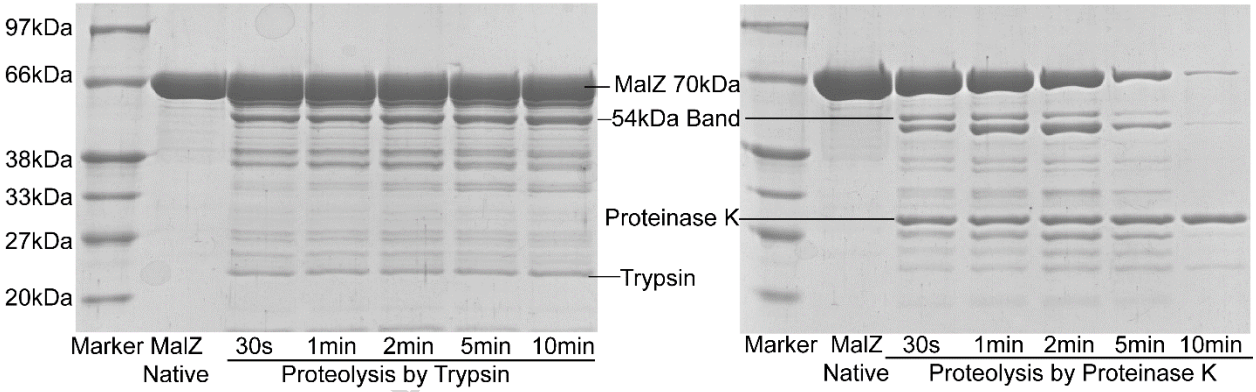


Figure 5: Limited proteolysis of MalZ
Limited proteolysis of MalZ by Trypsin and Proteinase K. The gels show a prominent band around 54 kDa which is the MalZ protein left after digestion of about 130 amino acids. It corresponds to the result as it is expected to be the N-terminal region itself that got digested.

The 54kDa band obtained after limited proteolysis was cut from the gel and further analysed on mass spectroscopy after an in-gel digestion by trypsin. The peptides obtained after digestion do not match to the N-terminal region but match to the remaining MalZ corresponding to the middle and C-terminal region **(Table 2)**. These results confirmed that the N-terminal is getting digested during limited proteolysis. This supports our assumption that the N-terminal domain is either flexible or unstructured.

1	MLNAWHLPVP PFVKQSKDQL LITLWLTGED PPQRIMLRTE HDNEEMSVPM

51	HKQRSQPQPG	VTAWRAAIDL	SSGQPRRRYS	FKLLWHDRQR	WFTPQGFSRM
101	PPARLEQFAV	DVPDIGPQWA	ADQIFYQIFP	DRFARSLPRE	AEQDHVYYHH
151	AAGQEIILRD	WDEPVTAQ**AG**	**GSTFYGGDLD**	**GISEK**LPYLK	**KLGVTALYLN**
201	**PVFK**APSVHK	YDTEDYR**HVD**	**PQFGGDGALL**	**R**LRHNTQQLG	MR**LVLDGVFN**
251	**HSGDSHAWFD**	**R**HNRGTGGAC	HNPESPWRDW	YSF**SDDGTAL**	**DWLGYASLPK**
301	**LDYQSESLVN**	**EIYRGEDSIV**	**R**HWLK**APWNM**	**DGWRLDVVHM**	**LGEAGGARNN**
351	**MQHVAGITEA**	**AKETQPEAYI**	**VGEHFGDAR**Q	WLQADVEDAA	MNYRGFTFPL
401	**WGFLANTDIS**	**YDPQ**QIDAQT	CMAWMDNYR**A**	**GLSHQQQLRM**	**FNQLDSHDTA**
451	**R**FKTLLGRDI	ARLPLAVVWL	FTWPGVPCIY	YGDEVGLDGK	NDPFCR**KPFP**
501	**WQVEKQDTAL**	**FALYQR**MIAL	RKKSQALRHG	GCQVLYAEDN	VVVFVRVLNQ
551	QR**VLVAINR**G	EACEVVLPAS	PFLNAVQWQC	KEGHGQLTDG	ILALPAISAT
601	VWMN				

Table 2. Peptide mass fingerprinting of band obtained after limited proteolysis
The 54kDa band obtained after limited proteolysis experiment was analysed by peptide mass fingerprinting after an in-gel digestion by trypsin. The bold red letters in the amino acid sequence of MalZ denotes the peptides identified in the excised band. The absence of N-terminal peptides here clearly shows that the N-terminal was truncated in the 54kda band obtained after limited proteolysis.

3.8 N-Terminal truncation affects the enzymatic activity/solubility of MalZ *in vivo*.

The N-terminal truncation affects the solubility of MalZ. This could be clearly observed by fractionation analysis on a SDS PAGE gel (**Figure 6a**). It is evident that while the native MalZ protein itself is not highly soluble, about 20% of overexpressed protein could be observed in the soluble supernatant part. However in the case of MalZ-N^{Trunc} the whole protein could be observed in the insoluble pellet fraction. Thus, showing that the deletion of N-terminal has a profound effect on the solubility of MalZ protein. Chaperonin GroEL and GroES assist in the folding of MalZ full length protein in *E.coli* and this can be demonstrated from an SDS PAGE gel showing fractions of soluble protein in the supernatant and insoluble part in the pellet. However, to determine the effect of truncation a comparison of enzymatic activity of the protein was made since it is a more sensitive assay than comparison on gels to find any soluble and functional protein in the whole cell lysate directly or in the supernatant fraction. The values are average of four independent experiments and the cells were normalized according to the $O.D._{600}$

of the culture used for lysis. The lysis conditions were kept similar in all the cases. The value for BL-21 MalZ was taken as the reference and other values are calculated relative to that. The values of activity analysis in case of MalZ-N^{Trunc} were comparable to the basal level of MalZ natively expressed in *E.coli* without any transformation **(Figure 6d**). When enzymatic activity of MalZ was compared among *E.coli* cells having MalZ full, MalZ-N^{Trunc} and these co-expressed with GroEL and GroES **(Figure 6b)**, it was found that MalZ full shows a significant increase in the enzymatic activity when co-expressed with GroEL-ES **(Figure 6d)**. The relative increase in enzymatic activity of over-expressed MalZ with GroEL and GroES co-expression was about three times that of over-expressed MalZ alone. However, even when co-expressed with GroEL and GroES, MalZ-N^{Trunc} does not show any enzymatic activity in any of the fractions.

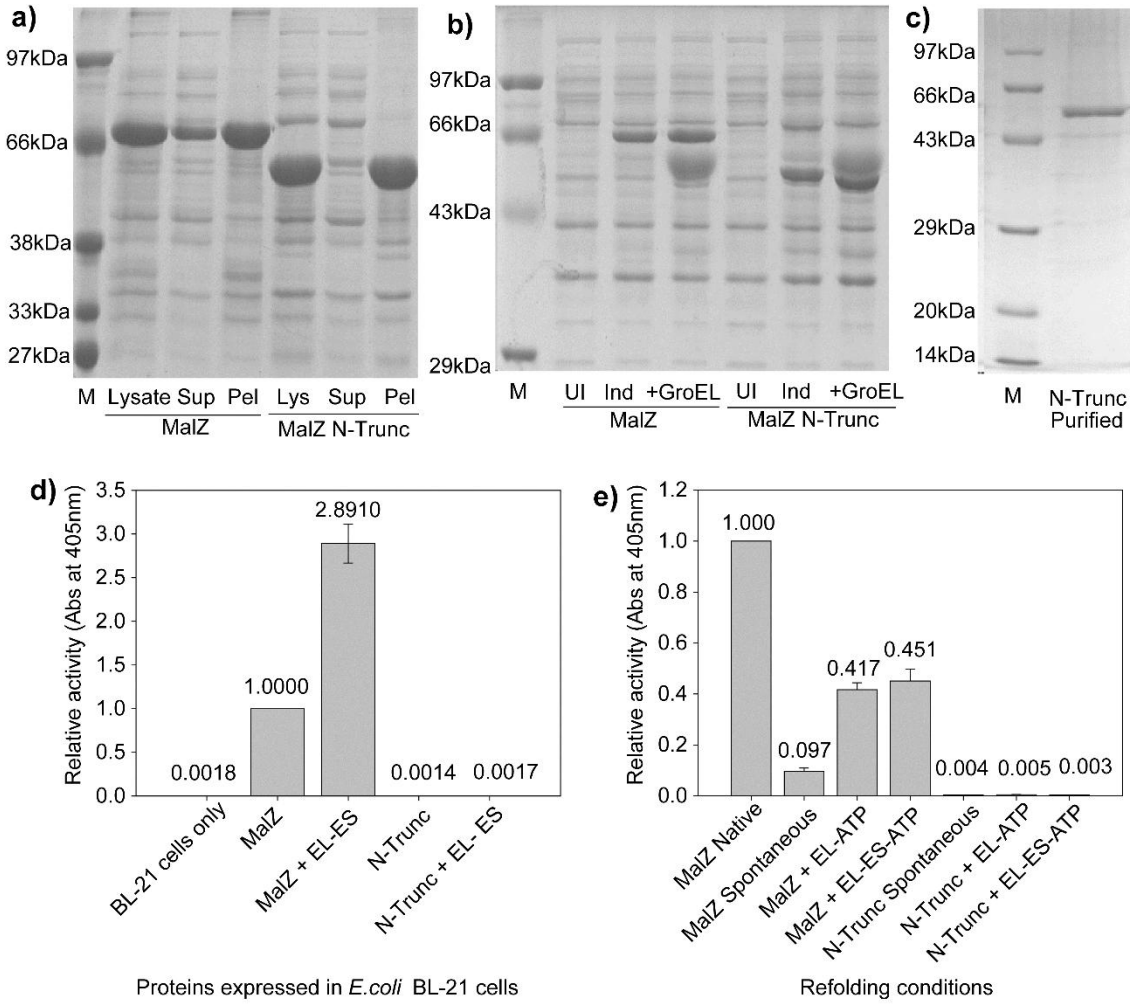


Figure 6: MalZ N^{Trunc} solubility analysis in *E.coli* cytosol and *in vitro*

a) Solubility analysis of MalZ and MalZ-N^{Trunc} through fractionation, the lysate, supernatant and the pellet fractions loaded on gel. It could be observed that while about 20% MalZ full length protein is soluble (Sup-MalZ), the truncated part is totally found in the insoluble fraction and no protein is visible in the soluble part (sup-MalZ N^{Trunc}). **b)** Co-expression of MalZ and N-Terminal truncated version with GroEL; and GroES. The truncated protein is 57 kDa same as GroEL and hence could not be resolved properly; **c)** Purification of MalZ-N^{Trunc}. SDS PAGE gel shows a band corresponding to 57 kDa for the purified protein; **d)** Relative enzymatic activities of MalZ and MalZ-N^{Trunc} when expressed in *E.coli.* Co-expression of GroEL and GroES helps in enhancing MalZ activity by increasing the overall soluble content of the protein. In case of MalZ-N^{Trunc} no enzymatic activity was found even when co-expressed with GroEL and GroES; **e)** Figure showing relative enzymatic activities of MalZ and MalZ-N^{Trunc} when refolded from denatured state *in vitro.* GroEL and GroES assistance in refolding helps in almost 45% recovery of MalZ activity as compared to less than 10% in case of spontaneous refolding. In case of MalZ-N^{Trunc} no enzymatic activity was recovered and the assistance of GroEL and GroES was also not sufficient for the recovery of activity.

3.9 MalZ-N^{Trunc} does not fold *in vitro* even with assistance of GroEL and GroES

MalZ full length protein refolds poorly from the chemically denatured state amounting to about 10% recovery in activity. *In vitro* refolding of MalZ however is enhanced and the activity of refolded protein was monitored to be about 45% of the native protein in the presence of GroEL and GroES in the refolding buffer, while compared to less than 10% activity recovered in the case of spontaneous refolding. This clearly shows the effect of chaperonin in the assistance of folding of MalZ. Purified MalZ-N^{Trunc} **(Figure 6c)** could not produce any activity in the refolding experiments and it was completely unable to fold spontaneously. GroEL and GroES assisted refolding was also attempted, but unlike full length MalZ protein, MalZ-N^{Trunc} did not show any activity even when refolded in presence of chaperonins GroEL and GroES. Thus it may be inferred that the N-terminal region of MalZ plays a role in the GroEL assisted folding of the protein. **(Figure 6e)**

3.10 The isolated MalZ N-terminal domain remains soluble at higher temperature

The previous experiments suggested that the MalZ-N^{Trunc} could not fold to an enzymatically active state in the presence of chaperonins GroEL and GroES. This proves that N-terminal is indispensable for the folding of MalZ to its native state. To further identify the role of N-terminal domain in the folding of MalZ protein, the N-terminal domain part which was truncated from MalZ was separately cloned in pACYC duet vector and expressed in *E.coli* BL-21 DE3 cells with the help of a T7 promoter present in pACYC duet **(Figure 7a)**.

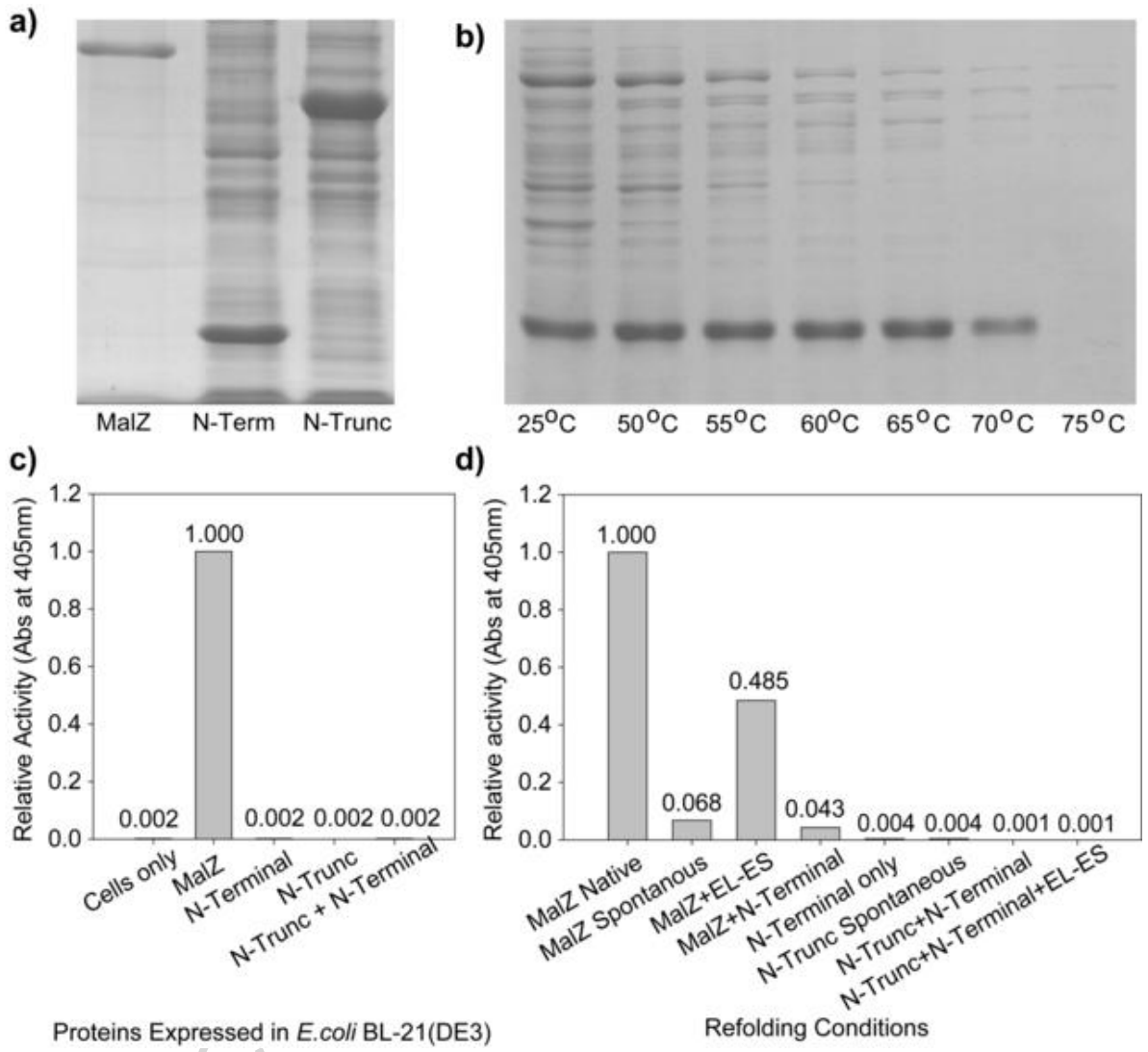


Figure 7. Folding of MalZ N^{Trunc} with separately cloned N-terminal. a) Expression of MalZ-N^{Trunc} and N-terminal compared to pure MalZ on SDS PAGE gel; **b)** Stability of MalZ N-terminal at higher temperatures, from room temperature to 65°C the N-terminal was found to be stable. The lanes show the soluble fractions of protein at the mentioned temperatures; **c)** *in vivo* refolding attempt of MalZ-N^{Trunc} in the presence of N-terminal region, the truncated version could not fold in any condition; **d)** *in vitro* refolding of MalZ-N^{Trunc} in the presence of N-terminal region and GroEL and GroES in 1:50 ratio, similar to the previous experiments, the activity of the truncated protein could not be restored by providing the N-terminal *in trans*.

While the MalZ-N^{Trunc} is insoluble and aggregates during overexpression as it is found in inclusion bodies, the separate N-terminal region was soluble and was found in supernatant after a

fractionation experiment. Purification of the N-terminal region was attempted through gel filtration but the purity levels were not satisfactory. The eluted fraction from gel filtration was subjected to heat treatment at various temperatures. The N-terminal region was found to be stable at temperatures of up to 65^{o}C for 5 minutes and was found in supernatant when the heat treated samples were fractionated between soluble and insoluble fractions through high speed centrifugation. This method was also utilized for purification of MalZ N-terminal as the impurities left in the sample during purification aggregated at higher temperatures and were found in the pellet fraction while the N-terminal domain was present in the supernatant as purified form. **(Figure 7b).**

3.11 Enzymatic activity of the MalZ-N^{Trunc} could not be recovered with addition of N-terminal domain *in trans*.

Folding of MalZ-N^{Trunc} was attempted through co-expression of MalZ N-terminal domain with the MalZ-N^{Trunc} in BL-21 DE3. The cells were grown in LB media with ampicillin and chloramphenicol for selection and IPTG was used for over-expression. In order to identify the role of N-terminal domain in the solubility of MalZ, the solubility of MalZ-N^{Trunc} was analysed using an enzymatic activity assay. The assay was performed using cell lysate containing co-expressed N-terminal and MalZ-N^{Trunc}. Enzymatic activity in MalZ-N^{Trunc} could not be recovered in the presence of a co-expressed N-terminal region which suggests that the N-terminal region is an essential part of the MalZ protein and the protein is not able to fold in the presence of an external N-terminal region when co-expressed *in trans* **(Figure 7c).**

Further, an *in vitro* experiment was performed where folding of MalZ-N^{Trunc} was attempted in the presence of different stoichiometric ratios of the purified N-terminal part of the protein.

MalZ-N^{Trunc} was purified in denaturation conditions and its *in vitro* refolding was attempted in refolding buffers containing N-terminal MalZ protein only, and also in the presence of N-terminal MalZ and GroEL & GroES together. None of the two refolding conditions provided any traces of refolded MalZ. **(Figure 7d)**

3.13 The N-terminal region has direct role in stability of the protein.

The limited proteolysis of native MalZ by trypsin yields a major band at 54 kDa as already mentioned in figure 5. The enzymatic activity of the digested products were monitored after the proteolysis. The intensities of the major bands were compared to the enzymatic activity of MalZ (**Figure 8**). The band intensities and relative enzymatic activity were both compared to that of native MalZ on a relative scale. The relative band intensities were plotted on a stacked bar graph, with the lower bars showing the sum of intensities of all bands above 54kDa, and the top bars showing the intensities of the 54kDa bands which were identified to be the MalZ-N^{Trunc} part by mass spectrometric analysis. The line plot shows relative enzymatic activities after proteolytic digestion of MalZ at different times. The gel image of the proteolytic digestion products shows clear band at 54kDa **(Figure 8b)**. The intensity of 54 kDa band, the N-truncated MalZ generated from proteolytic digestion, however does not increase with time, showing that the MalZ N^{Trunc} is relatively less stable than native MalZ protein and might be getting further digested soon after the N-terminal is truncated. The values of enzymatic activity recorded after various time intervals are more than the undigested MalZ (quantitated from the bands), but lower than the expected value from the sum of uncleaved MalZ and its truncated version (Had it been stable). It further supports the assumption that the N truncated protein is getting degraded and hence losing its function in a time dependent manner. The trypsin digestion products were separated from undigested MalZ protein by employing hydrophobic interaction chromatography (Figure 8c). The protein

concentration was 0.08mM as determined by Bradford assay. The separated fractions containing the 54kDa truncated protein also displayed enzymatic activity. These observations consolidate the argument that the truncated MalZ is having enzymatic activity but it is relatively less stable than the native protein. Thus, it can be proposed that the N-region of MalZ is not directly involved in the enzymatic property of the protein.

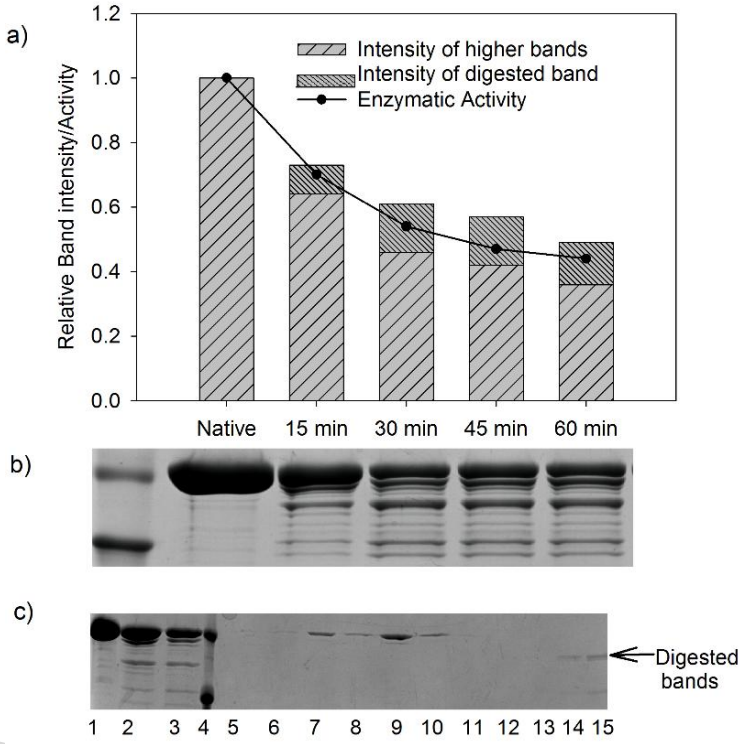


Figure 8. The N-terminal does not have direct role in activity

a) Stacked bar chart showing the band intensities after MalZ trypsin digestion, the lower stacks are the combined intensities of bands above the 54kDa band obtained in digestion, and the top stacked bars show intensities of the 54kDa(MalZ-N^{Trunc}) band obtained after trypsin digestion. The line plot shows the relative activity of digested product as compared to the native MalZ. The enzymatic activity is higher than the undigested native band intensities, but lower than the total of native and MalZ-N^{Trunc} suggesting that the truncated band may have transient activity but it is not stable and loses some activity after misfolding due to truncation. b) The proteolytic digestion profile on SDS PAGE showing the bands compared in panel (a), digested for time points of 15, 30, 45 and 60 mins. c) SDS PAGE showing fractions of digested band separated from MalZ full length protein. The trypsin digestion reaction mixture was loaded on a source15ISO column and eluted using a linear gradient from 1.2M ammonium sulphate to 50mM ammonium sulphate.

Lane1: native MalZ, lane2: digestion reaction mix, lane3: flow through, lane 4: Molecular weight marker, lane5-15 eluted fractions. Fractions 14 and 15 showing isolated digested products.

3.14 The N-terminal region consists of β-sheet and random coil structures

The circular dichroism spectra of the N-terminal region was analysed to get insights to the structural properties of the truncated part alone **(Figure 9a)**. The spectra was analysed by secondary structure prediction algorithms on the Jasco spectra manager software. The Yang's [38] and Reed's [39] references were used for the secondary structure analysis. The Yang's reference analysis shows that while 20% and 25% structure is α-helical and β-sheet respectively, the major part shows turns and random coils which consists of 21% and 33%. In case of prediction by Reed's reference the α-helical and β-sheet were 8% and 51% respectively while remaining 41% was random coil structure. The CD spectra of N-terminal hence shows that although β-strands are predominant, the region is somewhat unstructured with almost half of the region showing random coil structures.

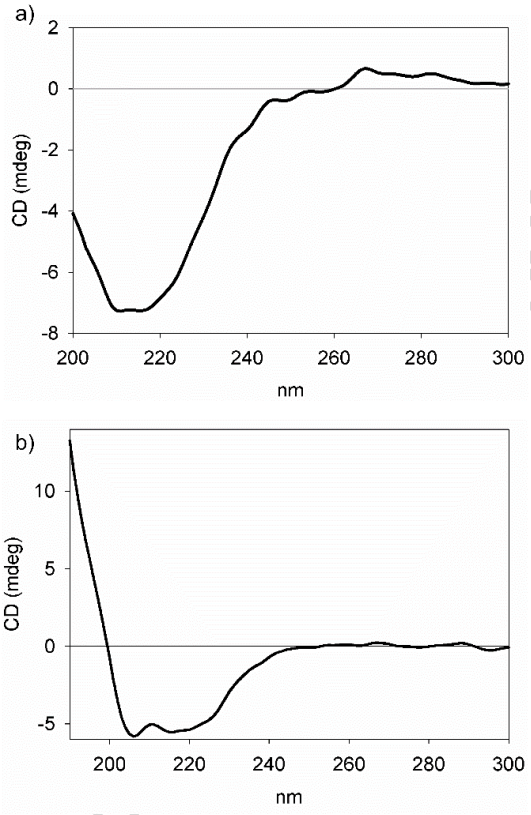


Figure 9. The CD spectra of isolated N-terminal domain of MalZ
a) The CD spectra of MalZ N terminal domain, the secondary structure prediction suggests that almost half of the N-terminal region consists of random coil and loop regions. b) The CD spectra of full length MalZ protein shows almost equal distribution of alpha helices and beta sheets.

The CD spectra of full length MalZ **(Figure 9b)** shows 30% and 24% structure as α-helical and β-strands respectively, while 45% of the structure is expected to be random. The refined homology model of full length MalZ consists of 23% α-helical and 19% β-sheet content, which is somewhat similar to the CD spectra. The crystal structure of MalZ shows about 11% α-helical and 3% β-sheet content. The electron density at N-terminal is missing which might be leading to a lower β-sheet content in the crystal structure, and moreover the low resolution of the x-ray diffraction data could be a reason that the crystal structure is showing low percentages of these secondary structures.

DISCUSSION

The Maltodextrin glucosidase enzyme is an aggregation prone protein and folds poorly when allowed to fold spontaneously from the chemically denatured state. The chaperonins GroEL and GroES provide assistance to the protein to attain its native form. There are many reports in literature correlating the role of hydrophobicity [40] and secondary structures [41-43] to the requirement of chaperones for folding to native state. To understand the role of structural features of MalZ in its aggregation prone behaviour and requirement of chaperone assistance, elucidation of the structure was necessary. In the absence of crystal structure, sequence analysis and homology model prediction were done, which suggested that the N-terminal of this protein is a distinct region. Further, disorder prediction algorithms identified that a major part of N-terminal has a disorder propensity. In order to obtain more precise information, we decided to solve the crystal structure of the MalZ protein. The protein was crystallized, and the crystals, despite having a very sharp and clear morphology, diffracted rather poorly and the structure was solved and refined to 3.7Å. Despite the low resolution, few details corresponding to the nature of the protein were readily obtained. The structure clearly showed the presence of a Tim-barrel domain, which is found in many GroEL substrates [34, 35 and 41]. Interestingly, the N-terminal region (Met1 to Phe131) was not observed in the structure, which could be attributed to flexibility of the region or a propensity to be statistically disordered. The role of crystallization condition in causing the disorder or flexibility was ruled out experimentally. Limited proteolysis of the native MalZ followed by analysing the peptide fragments also suggested the presence of unstructured N-terminal region. In order to have a better understanding of the role of N-terminal region of the protein in its stability and function, a truncated version of MalZ (Ala111-His616) was prepared. The decision on the selection of truncated region has been taken on the basis of the facts that (a) removal of first 130 amino acid residues may disrupt the important part of the

protein, (b) even though Met1 to Phe110 segment was slightly smaller than the missing zone Met1 - Phe131 in the MalZ crystal structure, it constitutes the major part of the N-terminal region as predicted by homology modelling. It was found that the truncated version is completely insoluble in *E.coli* cytosol as observed from the fractionation experiment. Based on the background information on the GroEL-ES assisted folding of full length MalZ, chaperonin mediated folding of the MalZ-N^{Trunc} was attempted in *E.coli* and *in vitro.* Further, assisted folding experiments were done with the separately cloned N-terminal region of the protein used *in trans*. The results from the above experiments suggested that the isolated N-terminal part of MalZ could not help in the folding of MalZ-N^{Trunc} protein. Thus, the N-terminal region of MalZ, as an integral part, is essential for the folding of the nascent protein. The overall analysis of secondary structures of MalZ and its isolated N-terminal region show that N-terminal primarily consists of β-sheet structures, and significant percentage of random structure. The major reason for the absence of N-terminal region in the crystal structure of MalZ could still be the flexibility in the linker region as evident in the limited proteolysis experiment. These findings suggest that while the N-terminal region of MalZ might not be having a rigid structure in its crystallized form, it may probably be indispensable for the folding, stability and solubility. These results may also suggest that since the N-terminal region of MalZ is unstructured, co-expressed GroEL-ES might be protecting this region from misfolding during the over-expression of the protein in *E.coli* and also in the *in vitro* refolding. MalZ protein could be characterized as a hybrid protein having both structured and disordered region [44]. The unstructured N-terminal region could be stabilizing the whole protein by interacting with other domains as observed in certain cases [45]. It is especially important as the C-terminal region of the MalZ protein consists of highly hydrophobic contiguous amino acids and the N-terminal might be playing a key role in

stabilizing this region. N-terminal region in some similar proteins are known to form dimer interfaces [46], however the MalZ protein is known to be present as a functional monomer [2]. The requirement of N-terminal for stability and solubility can also be indicative of an internal chaperone like activity [47]. Based on various reports indicating isolated protein domains having internal chaperonin activities [48, 49], attempts to fold the truncated MalZ in presence of the isolated N-terminal region were made in similar manner. However, co-expression of isolated N-terminal fragment with the MalZ N^{Trunc} could not provide any soluble and functional form of the truncated protein. The *in vitro* refolding experiments of denatured MalZ N^{Trunc} also could not yield any refolding when the isolated N-terminal part was added *in trans*. Multiple reports cite that the N-terminal region can play a very important role in the folding and function of the protein and while the truncations of N-terminal region have in many cases resulted in loss of function [50], it can also result in enhancement of function for the protein [51, 52]. The N-terminal may play an active role in the catalytic activity as it has been reported in case for some amylases that N-terminal binds to starch or pullalan and may help in catalytic activity [53]. These cases however discuss the possible role of N-terminal in the extra catalytic functions like hydrolysis of α-1,6 glycosidic linkages in pullalan or transglycosylation activity. Fusion proteins attached to the N-terminal have also been known to enhance the solubility and function of proteins in multiple studies [54, 55] and the N-terminal may also play a similar role. Our results from the limited proteolysis followed by enzymatic activity measurements suggest that while the N-terminal may not be directly playing a role in the enzymatic activity, its deletion is affecting the solubility and stability of the protein.

In conclusion, the N-terminal region of MalZ appears to remain as a flexible entity in the native protein and may not be directly involved in the enzymatic property of the protein. However, the

N-terminal plays a crucial role in the folding of the newly synthesized MalZ protein and in keeping the protein in soluble form in *E.coli* cytosol.

ACKNOWLEDGEMENTS

TKC acknowledges the financial support from Department of Science and Technology (DST), Govt. of India, and Council of Scientific and Industrial Research (CSIR) Govt. of India. AP acknowledges Council of Scientific and Industrial Research (CSIR) Govt. of India for Doctoral fellowship. TPS acknowledges the financial supports from Indian National Science Academy, New Delhi and DST. The authors acknowledge Indian Institute of Technology Delhi for infrastructural support. The authors acknowledge Indian Institute of Technology Delhi for infrastructural support.

AUTHOR CONTRIBUTIONS

AP, AKS, TPS and TKC planned and designed the experiments, AP, AKS, purified and crystallized MalZ, cloned N-truncated and performed all the computational, activity and solubility based experiments. PKS and TPS refined the crystal structure, MJE cloned and purified isolated N-terminal. STM performed and analysed mass spec. experiment. AP, TPS and TKC have prepared the manuscript.

SUPPORTING INFORMATION

The refined crystal structure of MalZ protein has been submitted to the PDB, with ID: 5BN7. The sequence comparison with templates used for homology model and crystal structure refinement using molecular replacements is provided in **Supplementary Table 1.** The proteolytic sites for trypsin and proteinase K have been mentioned in **Supplementary Table 2**.

REFERENCES:

[1] R. Dippel, W. Boos, The Maltodextrin System of *Escherichia coli* : Metabolism and Transport, J. Bacteriol, 187 (2005) 8322–8331. doi:10.1128/JB.187.24.8322.

[2] S. Tapio, F. Yeh, H. A. Shuman, W. Boos, The malZ gene of *Escherichia coli*, a member of the maltose regulon, encodes a maltodextrin glucosidase, J. Biol. Chem. 266 (1991) 19450–19458.

[3] R. Peist, C. Schneider-Fresenius, W. Boos, The MalT-dependent and malZ-encoded maltodextrin glucosidase of *Escherichia coli* can be converted into a dextrinyltransferase by a single mutation, J. Biol. Chem. 271 (1996) 10681–10689. doi:10.1074/jbc.271.18.10681.

[4] K.M. Song, J.H. Shim, J.T. Park, S.H. Kim, Y.W. Kim, W. Boos, et al., Transglycosylation properties of maltodextrin glucosidase (MalZ) from *Escherichia coli* and its application for synthesis of a nigerose-containing oligosaccharide, Biochem. Biophys. Res. Commun. 397 (2010) 87–92. doi:10.1016/j.bbrc.2010.05.073.

[5] J.T. Park, J.H. Shim, P.L. Tran, I.H. Hong, H.U. Yong, E.F. Oktavina, et al., Role of maltose enzymes in glycogen synthesis by Escherichia coli, J. Bacteriol. 193 (2011) 2517–2526. doi:10.1128/JB.01238-10.

[6] R. Dippel, T. Bergmiller, A. Bohm, W. Boos, The Maltodextrin System of *Escherichia coli*: Glycogen-Derived Endogenous Induction and Osmoregulation, J. Bacteriol. 187 (2005) 8332–8339.

[7] S. Sivaramakrishnan, D. Gangadharan, K. Madhavan, C.R. Soccol, A. Pandey, K.M. Nampoothiri, et al., alpha-amylases from microbial sources - An overview on recent developments, Food Technol. Biotechnol. 44 (2006) 173–184.

[8] K. Ichikawa, T. Tonozuka, R. Uotsu-Tomita, H. Akeboshi, A. Nishikawa, Y. Sakano, Purification, Characterization, and Subsite Affinities of Thermoactinomyces vulgaris R-47 Maltooligosaccharide-metabolizing Enzyme Homologous to Glucoamylases, Biosci. Biotechnol. Biochem. 68 (2004) 413–420. doi:10.1271/bbb.68.413.

[9] J. Kim, S. Cha, H. Kim, T. Kim, N. Ha, S. Oh, et al., Crystal Structure of a Maltogenic Amylase Provides Insights into a Catalytic Versatility, J. Biol. Chem. 274 (1999) 26279–26286. doi:10.1074/jbc.274.37.26279.

[10] R. Gupta, P. Gigras, H. Mohapatra, V.K. Goswami, B. Chauhan, Microbial α-amylases: A biotechnological perspective, Process Biochem. 38 (2003) 1599–1616. doi:10.1016/S0032-9592(03)00053-0.

[11] S. Kamitori, A. Abe, A. Ohtaki, A. Kaji, T. Tonozuka, Y. Sakano, Crystal structures and structural comparison of *Thermoactinomyces vulgaris* R-47 α-amylase 1 (TVAI) at 1.6 Å resolution and α-amylase 2 (TVAII) at 2.3 Å resolution, J. Mol. Biol. 318 (2002) 443–453. doi:10.1016/S0022-2836(02)00111-0.

[12] S. Paul, S. Punam, T.K. Chaudhuri, Chaperone-assisted refolding of *Escherichia coli* maltodextrin glucosidase, FEBS J. 274 (2007) 6000–6010. doi:10.1111/j.1742-4658.2007.06122.x.

[13] K. Braig, M. Simon, F. Furuya, J.F. Hainfeld, A.L. Horwich, A polypeptide bound by the chaperonin groEL is localized within a central cavity., Proc. Natl. Acad. Sci. U. S. A. 90

(1993) 3978–3982. doi:10.1073/pnas.90.9.3978.

[14] T.K. Chaudhuri, G.W. Farr, W.A Fenton, S. Rospert, A.L. Horwich, GroEL/GroES-mediated folding of a protein too large to be encapsulated., Cell. 107 (2001) 235–46. http://www.ncbi.nlm.nih.gov/pubmed/11672530.

[15] S. Paul, C. Singh, S. Mishra, T.K. Chaudhuri, The 69 kDa *Escherichia coli* maltodextrin glucosidase does not get encapsulated underneath GroES and folds through trans mechanism during GroEL/GroES-assisted folding., FASEB J. 21 (2007) 2874–85. doi:10.1096/fj.06-7958com.

[16] S. Altschul, T. Madden, A. Schaffer, J. Zhang, Z. Zhang, W. Miller, et al., Gapped BLAST and PSI- BLAST: a new generation of protein database search programs, Nucleic Acids Res. 25 (1997) 3389–3402. http://nar.oxfordjournals.org/content/25/17/3389.short.

[17] A. Roy, A. Kucukural, Y. Zhang, I-TASSER: a unified platform for automated protein structure and function prediction., Nat. Protoc. 5 (2010) 725–738. doi:10.1038/nprot.2010.5.

[18] E. Potenza, T.D. Domenico, I. Walsh, S.C.E. Tosatto, MobiDB 2.0: an improved database of intrinsically disordered and mobile proteins, Nucleic Acids Res. 43 (2015) D315–D320. doi:10.1093/nar/gku982.

[19] J. Navaza, AMoRe: an automated package for molecular replacement, Acta Crystallogr. Sect. A. 50 (1994) 157–163. doi:10.1107/S0108767393007597.

[20] G.N. Murshudov, A.A. Vagin, E.J. Dodson, Refinement of Macromolecular Structures by the Maximum-Likelihood Method, Acta Crystallogr. Sect. D. 53 (1997) 240–255.

doi:10.1107/S0907444996012255.

[21] T.A. Jones, J.-Y. Zou, S.W. Cowan, M. Kjeldgaard, Improved methods for building protein models in electron density maps and the location of errors in these models, Acta Crystallogr. Sect. A. 47 (1991) 110–119. doi:10.1107/S0108767390010224.

[22] E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M.R. Wilkins, R.D. Appel, et al., Protein Identification and Analysis Tools on the ExPASy Server, Proteomics Protoc. Handb. (2005) 571–607. doi:10.1385/1-59259-890-0:571.

[23] A. Shevchenko, H. Tomas, J. Havlis, J. V Olsen, M. Mann, In-gel digestion for mass spectrometric characterization of proteins and proteomes, Nat. Protoc. 1 (2007) 2856–2860. doi:10.1038/nprot.2006.468.

[24] M. Reyes, N.A. Treptow, H.A. Shuman, Transport of p-nitrophenyl-alpha-maltoside by the maltose transport system of *Escherichia coli* and its subsequent hydrolysis by a cytoplasmic alpha-maltosidase., J. Bacteriol. 165 (1986) 918–922.

[25] A. Ohtaki, A. Iguchi, M. Mizuno, T. Tonozuka, Y. Sakano, Mutual conversion of substrate specificities of *Thermoactinomyces vulgaris* R-47 α-amylases TVAI and TVAII by site-directed mutagenesis, Carbohydr. Res. 338 (2003) 1553–1558. doi:10.1016/S0008-6215(03)00219-2.

[26] H. Hondoh, T. Kuriki, Y. Matsuura, Three-dimensional Structure and Substrate Binding of *Bacillus stearothermophilus* Neopullulanase, J.Mol.Biol. 326 (2003) 177–188. doi:12547200.

[27] R. Linding, L.J. Jensen, F. Diella, P. Bork, T.J. Gibson, R.B. Russell, Protein Disorder

Prediction, Structure. 11 (2003) 1453–1459. doi:10.1016/j.str.2003.10.002.

[28] I. Walsh, A.J.M. Martin, T. Di Domenico, S.C.E. Tosatto, ESpritz: accurate and fast prediction of protein disorder, Bioinformatics. 28 (2012) 503–509. doi:10.1093/bioinformatics/btr682.

[29] Z. Dosztanyi, V. Csizmok, P. Tompa, I. Simon, IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content, Bioinformatics. 21 (2005) 3433–3434. doi:10.1093/bioinformatics/bti541.

[30] Z.R. Yang, R. Thomson, P. McNeil, R.M. Esnouf, RONN: The bio-basis function neural network technique applied to the detection of natively disordered regions in proteins, Bioinformatics. 21 (2005) 3369–3376. doi:10.1093/bioinformatics/bti534.

[31] Z. Otwinowski, W. Minor, Macromolecular Crystallography Part A, Methods Enzymol. 276 (1997) 307–326. doi:10.1016/S0076-6879(97)76066-X.

[32] T. Le Gall, P.R. Romero, M.S. Cortese, V.N. Uversky, A.K. Dunker, Intrinsic disorder in the Protein Data Bank., J. Biomol. Struct. Dyn. 24 (2007) 325–342. doi:10.1080/07391102.2007.10507123.

[33] G. Pujadas, J. Palau, Evolution of alpha-amylases: architectural features and key residues in the stabilization of the (beta/alpha)(8) scaffold., Mol. Biol. Evol. 18 (2001) 38–54.

[34] A. Azia, R. Unger, A. Horovitz, What distinguishes GroEL substrates from other Escherichia coli proteins?, FEBS J. 279 (2012) 543–550. doi:10.1111/j.1742-4658.2011.08458.x.

[35] F. Georgescauld, K. Popova, A.J. Gupta, A. Bracher, J.R. Engen, M. Hayer-Hartl, et al.,

GroEL / ES Chaperonin Modulates the Mechanism and Accelerates the Rate of TIM-Barrel Domain Folding, Cell. 157 (2014) 922–934. doi:10.1016/j.cell.2014.03.038.

[36] A. Fontana, P.P. De Laureto, B. Spolaore, E. Frare, P. Picotti, M. Zambonin, Probing protein structure by limited proteolysis, Acta Biochim. Pol. 51 (2004) 299–321. doi:035001299.

[37] D. Ellison, J. Hinton, S.J. Hubbard, R.J. Beynon, Limited proteolysis of native proteins: the interaction between avidin and proteinase K., Protein Sci. 4 (1995) 1337–1345. doi:10.1002/pro.5560040709.

[38] Y.H. Chen, J.T. Yang, A new approach to the calculation of secondary structures of globular proteins by optical rotatory dispersion and circular dichroism., Biochem. Biophys. Res. Commun. 44 (1971) 1285—1291. doi:10.1016/S0006-291X(71)80225-5.

[39] J. Reed, T.A. Reed, A set of constructed type spectra for the practical estimation of peptide secondary structure from circular dichroism., Anal. Biochem. 254 (1997) 36–40. doi:10.1006/abio.1997.2355.

[40] T.K. Chaudhuri, P. Gupta, Factors governing the substrate recognition by GroEL chaperone : a sequence correlation approach, Cell stress and chaperones. 10 (2005) 24–36.

[41] W.A. Houry, D. Frishman, C. Eckerskorn, F. Lottspeich, F.U. Hartl, Identifcation of in vivo substrates of the chaperonin GroEL, Nature 402 (1999) 147–154.

[42] M. Schmidt, J. Buchner, Interaction of GroE with an all-β-protein, J. Biol. Chem. 267 (1992) 16829–16833.

[43] S.J. Landry, L.M. Gierasch, The chaperonin GroEL binds a polypeptide in an alpha.-

helical conformation, Biochemistry. 30 (1991) 7359–7362. doi:10.1021/bi00244a001.

[44] R. Van Der Lee, M. Buljan, B. Lang, R.J. Weatheritt, G.W. Daughdrill, a K. Dunker, et al., Classification of Intrinsically Disordered Regions and Proteins, Chem. Rev. 114, (2014) 6589−6631. dx.doi.org/10.1021/cr400525m

[45] D.C.F. Lanza, J.C. Silva, E.M. Assmann, A.J.C. Quaresma, G.C. Bressan, I.L. Torriani, et al., Human FEZ1 has characteristics of a natively unfolded protein and dimerizes in solution., Proteins. 74 (2009) 104–21. doi:10.1002/prot.22135.

[46] J.H. Fong, B.A. Shoemaker, S.O. Garbuzynskiy, M.Y. Lobanov, O.V. Galzitskaya, A.R. Panchenko, Intrinsic disorder in protein interactions: Insights from a comprehensive structural analysis, PLoS Comput. Biol. 5 (2009) 13–17. doi:10.1371/journal.pcbi.1000316.

[47] P. Tompa, P. Csermely, The role of structural disorder in the function of RNA and protein chaperones., FASEB J. 18 (2004) 1169–1175. doi:10.1096/fj.04-1584rev.

[48] A. Chassaing, S. Pichard, A. Araye-Guet, J. Barbier, V. Forge, D. Gillet, Solution and membrane-bound chaperone activity of the diphtheria toxin translocation domain towards the catalytic domain, FEBS J. 278 (2011) 4516–4525. doi:10.1111/j.1742-4658.2011.08053.x.

[49] R. Tomar, D.K. Garg, R. Mishra, A.K. Thakur, B. Kundu, N-terminal domain of Pyrococcus furiosus L-asparaginase functions as a non-specific, stable, molecular chaperone, FEBS J. 280 (2013) 2688–2699. doi:10.1111/febs.12271.

[50] G. Anderluh, J. Pungercar, I. Krizaj, B. Strukelj, F. Gubensek, P. Macek, N-terminal

truncation mutagenesis of equinatoxin II, a pore-forming protein from the sea anemone Actinia equina., Protein Eng. 10 (1997) 751–755.

[51] H.J. Xu, K. Xu, Y. Zhou, J. Li, W.F. Benedict, S.X. Hu, Enhanced tumor cell growth suppression by an N-terminal truncated retinoblastoma protein, Proc. Natl. Acad. Sci. U. S. A. 91 (1994) 9837–9841.

[52] J.H. Dunham, R.C. Meyer, E.L. Garcia, R.A. Hall, GPR37 surface expression enhancement via N-terminal truncation or protein-protein interactions., Biochemistry. 48 (2009) 10286–97. doi:10.1021/bi9013775.

[53] A. Abe, T. Tonozuka, Y. Sakano, S. Kamitori, Complex structures of Thermoactinomyces vulgaris R-47 α-amylase 1 with malto-oligosaccharides demonstrate the role of domain N acting as a starch-binding domain, J. Mol. Biol. 335 (2004) 811–822. doi:10.1016/j.jmb.2003.10.078.

[54] A. Gaudry, B. Lorber, A. Neuenfeldt, C. Sauter, C. Florentz, M. Sissler, Re-designed N-terminus enhances expression, solubility and crystallizability of mitochondrial protein, Protein Eng. Des. Sel. 25 (2012) 473–481. doi:10.1093/protein/gzs046.

[55] M. Hammarström, E. A. Woestenenk, N. Hellgren, T. Härd, H. Berglund, Effect of N-terminal solubility enhancing fusion proteins on yield of purified target protein, J. Struct. Funct. Genomics. 7 (2006) 1–14. doi:10.1007/s10969-005-9003-7.