**TITLE:**
X-Ray Crystallography to Study the Oligomeric State Transition of the *Thermotoga Maritima* M42 Aminopeptidase TmPep1050

**AUTHORS AND AFFILIATIONS:**
Raphaël Dutoit¹,², Nathalie Brandt², Dany Van Elder¹, Louis Droogmans¹

¹Laboratory of Microbiology, Department of Molecular Biology, Université Libre de Bruxelles, Brussels, Belgium
²Labiris Institut de Recherche, Brussels, Belgium

**Corresponding Author:**
Raphaël Dutoit (rdutoit@ulb.ac.be)

**Email Addresses of Co-Authors:**
Nathalie Brandt (nbrandt@spfb.brussels)
Dany Van Elder (dvelder@ulb.ac.be)
Louis Droogmans (ldroogma@ulb.ac.be)

**KEYWORDS:**
protein oligomerization, M42 aminopeptidase, oligomeric state transition, metalloenzyme, protein purification, protein crystallization, X-ray crystallography, data processing, molecular replacement

**SUMMARY:**
This protocol has been developed to study the dimer-dodecamer transition of TmPep1050, an M42 aminopeptidase, at the structural level. It is a straightforward pipeline starting from protein purification to X-ray data processing. Crystallogenesis, data set indexation, and molecular replacement have been emphasized through a case of study, TmPep1050_H60A_H307A variant.

**ABSTRACT:**
The M42 aminopeptidases form functionally active complexes made of 12 subunits. Their assembly process appears to be regulated by their metal ion cofactors triggering a dimer-dodecamer transition. Upon metal ion binding, several structural modifications occur in the active site and at the interaction interface, shaping dimers to promote the self-assembly. To observe such modifications, stable oligomers must be isolated prior to structural study. Reported here is a method that allows the purification of stable dodecamers and dimers of TmPep1050, an M42 aminopeptidase of *T. maritima*, and their structure determination by X-ray crystallography. Dimers were prepared from dodecamers by removing metal ions with a chelating agent. Without its cofactor, dodecamers became less stable and were fully dissociated upon heating. The oligomeric structures were solved by the straightforward molecular replacement approach. To illustrate the methodology, the structure of a TmPep1050 variant, totally impaired in metal ion binding, is presented showing no further breakdown of dimers to monomers.
INTRODUCTION:

Oligomerization is a predominant process that dictates the biological functions of many proteins. In *Escherichia coli*, it is estimated that only 35% of proteins are monomeric. Some proteins, called morpheeins, can even adopt several oligomeric states with subunits having distinct structure in each oligomeric state. The transition between their oligomeric states is often a means to regulate protein activity as each oligomeric state may have a different specific activity or function. Several examples of morpheeins have been well-documented in literature, notably the porphobilinogen synthase, HPr kinase/phosphatase, Lon protease, lactate dehydrogenase, glyceraldehyde-3-phosphate dehydrogenase, pyruvate kinase, citrate synthase, and ribonuclease A. Recently, we described the M42 aminopeptidase TmPep1050, another example of enzyme with morpheein-like behavior, whose activity depends on its oligomeric states. The transition between its oligomeric states is mediated by its metallic cofactors which induce several structural modifications of the subunits.

The M42 aminopeptidase family belongs to the MH clan, and is widely distributed among Bacteria and Archaea. The M42 aminopeptidases are genuine dinuclear enzymes degrading peptides up to 35 amino acid residues in length. They adopt a peculiar tetrahedron-shaped structure made of 12 subunits with their active sites are oriented towards an inner cavity. Such an arrangement is often described as a nano-compartmentalization of the activity to avoid uncontrolled proteolysis. The physiological function of the M42 aminopeptidases may be associated with the proteasome, hydrolyzing peptides generated from protein degradation. Indeed, *Pyrococcus horikoshii* possesses four M42 aminopeptidases, each presenting distinct but complementary specificities. Singularly, heterocomplexes made of two different types of subunits have been described in *P. horikoshii*, suggesting the existence of peptidasome complexes.

Several structures of M42 aminopeptidases have been described in the literature. The subunit is composed of two distinct domains, a catalytic domain and a dimerization domain. The catalytic domain adopts a common α/β domain conserved in the whole MH clan, the archetypal catalytic domain being the aminopeptidase Ap1 of *Vibrio proteolyticus*. The dimerization domain adopts a PDZ-like fold and may have, in addition to its role in the oligomerization, a role in controlling substrate access and binding in the inner cavity. As the basic building block is a dimer, the dodecamer is often described as the association of six dimers, each dimer being positioned at each edge of the tetrahedron. The oligomerization of the M42 aminopeptidases relies on the availability of its metal cofactors. Divalent metal ions, often Zn$^{2+}$ and Co$^{2+}$, are catalytically involved in the peptide binding and hydrolysis. They are found in two distinct binding sites, namely M1 and M2 sites. The two metal ions also drive and finely tune the oligomerization as demonstrated for PhTET2, PhTET3, PfTET3, and TmPep1050. When the metal cofactors are depleted, the dodecamer disassembles into dimers, like in PhTET2, PhTET3, and TmPep1050, or even monomers, like in PhTET2 and PfTET3.

Presented here is a protocol used for studying the structures of TmPep1050 oligomers. This protocol is a set of common methods including protein purification, proteolytic activity screening, crystallization, X-ray diffraction, and molecular replacement. Subtleties inherent to dealing with...
metalloenzymes, protein oligomerization, protein crystallization and molecular replacement are emphasized. A case of study is also presented to show whether TmPep1050 dodecamers may further dissociate into monomers or not. To address this question, a TmPep1050 variant, TmPep1050\textsubscript{H60A H307A}, has been studied whose metal binding sites are impaired by mutating His-60 (M2 site) and His-307 (M1 site) to Ala residues. This protocol may be accommodated to study other M42 aminopeptidases or any metalloenzymes with morpheein-like behavior.

**PROTOCOL:**

1. **Production and purification of recombinant TmPep1050**

   NOTE: Hereafter are described the cloning procedure and purification of wild-type TmPep1050 adapted from a previous study\textsuperscript{11}. Alternatively, the cloning can be done using a synthetic gene. To generate TmPep1050 variants, site directed mutagenesis can be performed following, for instance, the single-primer reactions in parallel protocol (SPRINP) method\textsuperscript{30}. The purification protocol can be used for TmPep1050 variants. The use of His-tag should be avoided as it interferes with metal ion binding.

   1.1. Expression vector design

   1.1.1. Acquire genomic DNA of *Thermotoga maritima* MSB8 (ATCC 43589) or TmCD00089984 (Joint Center for Structural Genomics).

   1.1.2. PCR amplify\textsubscript{[A1]}TM\textsubscript{1050} ORF\textsubscript{[A2]}\textsubscript{[A3]} using either genomic DNA or template plasmid, a high-fidelity DNA polymerase, and the following primers: ocej419 (5’-TTTAACTTTAAGAAGGAGATATACATACCCATGAAGGAACTGATCAGAAAGCTG) and ocej420 (5’-ATCCGCCAAAACAGCCAAGCTGGAGACCGTTTACGCCCCAGATACCTGATGAG). Run the PCR screening according to the following scheme: 5 min at 95 °C, 30 cycles of 3 steps (30 s at 95 °C, 30 s at 55 °C, 90 s at 72 °C), and 10 min at 72 °C as the final step.

   1.1.3. Clone the PCR fragment into a suitable expression vector (Table of Materials) by homologous recombination (Figure 1) in *E. coli* according to the SLiCE protocol\textsuperscript{31}. To 50 ng of linearized vector, add PCR fragment in a 10:1 molar ratio of fragment to vector, 1 µL of PPY strain extract, 50 mM Tris-HCl pH 7.5, 100 mM MgCl\textsubscript{2}, 10 mM ATP, and 10 mM dithiothreitol (DTT) for a reaction volume of 10 µL. Incubate for 1 h at 37 °C.

   1.1.4. Transform chemically competent *E. coli* XL1 blue strain (or any recA\textsuperscript{−} suitable strain) with 1 µL of recombination reaction. Plate the cells on LB medium containing 100 µg/mL ampicillin. Incubate the plates overnight at 37 °C.

   1.1.5. Pick colonies on fresh LB plates with 100 µg/mL ampicillin. Incubate the plates at 37 °C for a least 8 h.

   1.1.6. Screen for positive candidates by colony PCR using suitable primer pair (5’-
ATGCCATAGCATTTTTTATCC and 5′- ATTTAATCTGTATCAGGC if using the recommended vector listed in Table of Materials). With a microtip end, scratch a picked colony and transfer the cells to 20 µL of reaction mix containing 0.5 µM of each primer and 10 µL of a commercial Taq DNA polymerase mix.

1.1.7. Run the PCR screening according to the following scheme: 5 min at 95 °C as the denaturation step, 30 cycles of 3 steps (30 s at 95 °C, 30 s at 55 °C, 90 s at 72 °C), and 10 min at 72 °C as the final step.

NOTE: The PCR reactions can be stored overnight in the PCR machine at 12 °C.

1.1.8. Load 10 µL of each PCR reaction on a 0.8 % agarose gel prepared in Tris-acetate-EDTA (TAE) buffer. Run the electrophoresis for 25 min at 100 V.

NOTE: A 1.1 kbp amplicon is expected.

1.1.9. Extract plasmids from candidates using a commercial kit [A4][A5] and sequence them using the same primer pair used in step 1.1.6.

1.2. Cell culture

NOTE: When a suitable candidate has been identified by sequencing, the clone can be directly used as the expression if using the recommended vector (Table of Materials). In that case, the expression is controlled by the arabinose-inducible P_{BAD} promoter\(^{32}\).

1.2.1. Inoculate 10 mL of LB medium containing 100 µg/mL ampicillin with the candidate and incubate the preculture overnight at 37 °C under orbital shaking. Add 5 mL of the preculture to 1 L of LB medium with 100 µg/mL ampicillin. Mind to respect an air to liquid ratio of 3.

1.2.2. Let cells grow at 37 °C under orbital shaking. Monitor the optical density at 660 nm (OD\(_{660}\)).

1.2.3. When OD\(_{660}\) has reached 0.5–0.6, rapidly cool the culture for 5 min on ice and transfer it to an incubator set to 18 °C.

1.2.4. Add 0.2 g/L arabinose to induce gene expression and incubate for 12–18 h at 18 °C.

1.2.5. Harvest cells by centrifuging the culture at 6,000 x g for 30 min at 4 °C. Discard the supernatant and wash cells with 100 mL of 0.9% (w/v) NaCl.

1.2.6. Centrifuge again at 6,000 x g for 15 min at 4 °C and discard the supernatant.

NOTE: Cell pellets can be used directly for protein extraction or stored at -80 °C.

1.3. Protein purification
1.3.1. Resuspend the cell pellets in 40 mL of 50 mM MOPS, 1 mM CoCl₂, pH 7.2. Add 1 µL of 25 U/µL benzonase [A6] DNA/RNA endonuclease and one tablet of protease inhibitor cocktail that does not contain EDTA. Sonicate the suspension in pulse mode under cooling for 30 min.

1.3.2. Centrifuge the crude extract at 20,000 x g for 30 min at 4 °C. Collect the supernatant and heat it in a water bath at 70 °C for 10 min.

1.3.3. Centrifuge the denatured cell extract at 20,000 x g for 30 min at 4 °C and collect the supernatant for purification.

1.3.4. Use suitable anion exchange resin (Table of Materials) packed in a column of ~15 mL of volume. Refer to manufacturer’s recommendations for the working flow rate and column pressure limit. Equilibrate the resin with 50 mM MOPS, 1 mM CoCl₂, pH 7.2.

1.3.5. Load the supernatant collected from step 1.3.3 into the column. Monitor the absorbance of the eluate [A7] at 280 nm. When it has reached the baseline, proceed to the elution.

1.3.6. Apply a gradient from 0 to 0.5 M NaCl in 50 mM MOPS, 1 mM CoCl₂, pH 7.2 for 5 column volumes (CV). Wait till the conductivity is stabilized and the absorbance has reached the baseline.

1.3.7. Apply a final gradient from 0.5 to 1 M NaCl in 50 mM MOPS, 1 mM CoCl₂, pH 7.2 for 1 CV.

1.3.8. Analyze some fractions (see Figure 2A for guidance) by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE).

NOTE: TmPep1050 appears as a 36 kDa band after Coomassie staining. Alternatively, the presence of TmPep1050 can be confirmed by activity assay (see section 2.1). At this step, fractions can be stored at 4 °C overnight.

1.3.9. Pool the fractions containing TmPep1050 and add finely ground powder of (NH₄)₂SO₄ to obtain a concentration of 1.5 M (NH₄)₂SO₄. Mix gently by inverting the tube upside down till complete dissolution.

1.3.10. Use hydrophobic interaction resin (Table of Materials) packed in a column of ~30 mL of volume. Refer to manufacturer’s recommendations for the working flow rate and column pressure limit. Equilibrate the resin with 50 mM MOPS, 1.5 M (NH₄)₂SO₄, 1 mM CoCl₂, pH 7.2.

1.3.11. Load the sample into the column and monitor the absorbance of the eluate at 280 nm. When the absorbance has reached the baseline, elute bound proteins by applying a gradient from 1.5 M to 0 M (NH₄)₂SO₄ in 50 mM MOPS, 1 mM CoCl₂, pH 7.2 for 5 CV.

1.3.12. Analyze some fractions (see Figure 2B for guidance) by SDS-PAGE.
1.3.13. Pool the fractions containing TmPep1050 and concentrate to 2 mL using ultrafiltration units with 30 kDa cutoff (Table of Materials). Proceed to section 1.4 to determine the molecular weight.

1.4. **Size exclusion chromatography**

1.4.1. Use size exclusion resin (Table of Materials) packed in a column of ~120 mL of volume. Refer to manufacturer’s recommendations for the working flow rate and column pressure limit. Equilibrate the resin with 50 mM MOPS, 0.5 M (NH₄)₂SO₄, 1 mM CoCl₂, pH 7.2.

1.4.2. Load the sample into the column and monitor the absorbance of the eluate at 280 nm. Fractionate from the column dead volume (~0.33 CV) till the end of the elution (1 CV).

1.4.3. Measure the elution volume for each observed peak.

**NOTE:** For guidance, dodecameric TmPep1050 elutes at ~82 mL (Figure 3A) under current experimental conditions, while dimeric TmPep1050, such as the TmPep1050<sub>H60A H307A</sub> variant, elutes at ~95 mL (Figure 3B). Some TmPep1050 may adopt both oligomeric forms, such as TmPep1050<sub>H60A</sub> (Figure 3C).

1.4.4. Analyze fractions corresponding to the maxima and tails of observed peaks using SDS-PAGE.

**NOTE:** TmPep1050 appears as a 36 kDa band after Coomassie staining.

1.4.5. Pool the fractions of each peak and concentrate using ultrafiltration units with 30 kDa cutoff (Table of Materials) to obtain a concentration of ~300 µM.

1.4.6. Measure the absorbance at 280 nm on a nano-volume spectrophotometer and calculate the concentration using the molecular extinction coefficient of 18,910 M⁻¹ cm⁻¹.

1.4.7. Store the purified protein at -18 °C.

1.4.8. To determine the molecular weight, calibrate the size exclusion chromatography (SEC) column using molecular weight standards (Table of Materials). Analyze the standards using 50 mM MOPS, 0.5 M (NH₄)₂SO₄, 1 mM CoCl₂ pH 7.2 as the running buffer.[A9][A10]

2. **Activity assay and apo-enzyme preparation**

**NOTE:** Originally, the apo-enzyme was prepared by diluting 1 volume of TmPep1050 in 10
volumes of 2.1 M malic acid pH 7.0 and concentrated back to 1 volume prior to dialysis\textsuperscript{11}. Below is presented an alternative procedure using 1,10-phenanthroline, a metal ion chelator. This procedure reduces protein loss and gives the same results than the previously published method.

2.1. Activity assay

2.1.1. Prepare a stock solution of 100 mM L-Leucine-p-nitroanilide (Table of Materials) in methanol.

2.1.2. Add 25 µL of 100 mM L-Leucine-p-nitroanilide in 965 µL of 50 mM MOPS, 250 µM CoCl\textsubscript{2}, pH 7.2, 10% methanol. Preincubate the reaction mix at 75 °C in a dry bath.

2.1.3. Dilute the enzyme in 50 mM MOPS pH 7.2 to a concentration of 1 µM. Add 10 µL to the reaction mix, vortex, and incubate at 75 °C either until it has turned yellowish or for 1 h.

2.1.4. Stop the reaction by adding 1 mL of 20% acetic acid. Vortex well and let it cool down to room temperature.

2.1.5. Transfer the reaction mix in a spectrophotometer cell. Read the absorbance at 410 nm against a negative control (incubated reaction mix without enzyme).

2.2. Apo-enzyme preparation

2.2.1. Prepare a stock solution of 1 M 1,10-phenanthroline in ethanol. Add 10 µL of 1,10-phenanthroline stock solution to 890 µL of 50 mM MOPS, 0.5 M (NH\textsubscript{4})\textsubscript{2}SO\textsubscript{4}, pH 7.2. Add 100 µL of purified TmPep1050 (300 µM–1 mM concentration).

2.2.2. Check the activity loss using the activity assay described in section 2.1 without adding CoCl\textsubscript{2} in the reaction mix.

2.2.3. Transfer the sample in a dialysis tube. Dialyze against 200 mL of 50 mM MOPS, 0.5 M (NH\textsubscript{4})\textsubscript{2}SO\textsubscript{4}, pH 7.2 at 4 °C. Exchange thrice the dialysate with fresh buffer during the 48 h dialysis.

2.2.4. Collect the sample from the dialysis tube and concentrate back to 100 µL using ultrafiltration units with 30 kDa cutoff (Table of Materials). Check the concentration by reading the absorbance at 280 nm using a nano-volume spectrophotometer.

2.3. Dimer preparation

2.3.1. Dilute the apo-enzyme to a concentration of 1 µM in 50 mM MOPS, 0.5 M (NH\textsubscript{4})\textsubscript{2}SO\textsubscript{4}, pH 7.2. Incubate for 2 h at 75 °C in a dry bath, then let the sample cool down to room temperature.

2.3.2. Concentrate the sample to an enzyme concentration of at least 50 µM. Check the molecular weight by SEC (see section 1.4). The elution peak must shift from ~82 mL to ~95 mL.
(under current experimental conditions).

3. TmPep1050 crystallization

NOTE: Protein crystallization remains an empirical science as it is a multifactorial phenomenon. While some parameters can be identified and controlled (such as temperature, pH, precipitation agent concentration), others may influence elusively the crystallization (such as protein and chemical purity, proteolysis, sample history). Nowadays protein crystallization is tackled in a rational and systematic manner thanks to a bunch of commercial crystallization screening conditions and automation. The optimization of a crystallization condition, however, relies mostly on a trial-and-error approach. Hereafter are described a blueprint for crystallizing proteins and several tips for optimizing the crystallization conditions.

3.1. Crystallization screening

NOTE: Using commercial crystallization kits, crystals of dodecameric TmPep1050 have been obtained in 2.2 M DL-malic acid pH 7.0, 0.1 M Bis-Tris propane pH 7.0 and 0.18 M tri-ammonium citrate, 20% polyethylene glycol (PEG) 3350. Crystals of dimeric TmPep1050 have been obtained in 0.1 M sodium citrate pH 5.6, 0.2 M ammonium acetate, 30% PEG4000. Crystals of dodecamers appear within a week and reach their full size in a month. Crystals of dimers usually appear within 24 h and grow to full size in a week.

3.1.1. Acquire several commercial crystallization kits (see Table of Materials for examples).

3.1.2. Set up crystallization plates (Table of Materials) for the hanging drop method. Fill the wells with 500 µL of each solution of a crystallization screening kit.

3.1.3. For each well, set up a crystallization support. On the support, deposit a 1 µL drop of purified protein (usually ~10 mg/mL).

3.1.4. Immediately pipet 1 µL of crystallization solution from the well. Add it carefully to the protein drop and mix gently by pipetting upside down thrice. The drop must remain semispherical without any bubbles.

3.1.5. Screw the support on top of the corresponding well. Repeat the operation for the whole kit.

3.1.6. After setting up the plates, observe each drop with a binocular. Refer to the crystallization kit user guide for interpretation (clear drop, phase separation, precipitate, needles, etc.).

3.1.7. Incubate the plates at 20 °C. Check the plates once per day during the first week and once per week afterwards.

3.1.8. Score each well using the score sheet and the user guide provided with the crystallization
3.2. Crystallization optimization

NOTE: The initial crystallization conditions of dodecameric TmPep1050 have been optimized to 2.1 M DL-malic acid pH 6.75 and 0.18 M tri-ammonium citrate pH 7.5, 40% (w/v) PEG3350 while the crystallization condition of dimeric TmPep1050 has been shifted to 0.1 M sodium citrate pH 6.0, 10% (w/v) PEG3350. One cycle of seeding has been necessary to improve crystallinity. Hereafter is described how the crystallization of TmPep1050_{H60A,H307A} variant has been optimized.

3.2.1. Prepare stock solutions of 0.5 M sodium citrate buffer at different pH (4.5, 5.2, and 6.0), and 50% (w/v) PEG3350 solution.

3.2.2. Set up a crystallization plate as a matrix of pH vs. precipitation agent (see Figure 4A).

3.2.3. Incubate the plate at 20 °C. Observe each well with a binocular once per day during a week.

3.2.4. Score each well according to crystal size and shape. Select a condition giving crystals of at least 50 µm. Proceed to microseeding.

3.3. Microseeding

NOTE: Microseeding is a powerful method to improve the shape, size and crystallinity of protein crystals. A faster seeding approach is streak seeding using a cat whisker. See Figure 4 as an example how crystallization optimization and microseeding have improved the crystal shape and size for TmPep1050_{H60A,H307A}.

3.3.1. Prepare the seeds using the selected well in step 3.2.4.

3.3.2. Increase the volume of a drop containing crystals to 10 µL by adding crystallization solution from the well. Pipet the drop and add 90 µL of crystallization solution from the well. Vortex thoroughly and keep the seeds on ice.

3.3.3. Prepare several dilutions of the seeds: 1x, 10x, 25x, and 100x. Vortex well the seeds before pipetting. Keep the dilutions on ice.

3.3.4. For each seed dilution, set up a crystallization plate as a matrix of pH vs. precipitation agent (see Figure 4B). Use the stock solutions prepared in step 3.2.1.

3.3.5. When making the drop, add 0.2 µL of seeds for a 2 µL drop.

3.3.6. Incubate the plate at 20 °C. Observe each well with a binocular once per day during a week.

NOTE: Crystal size distribution and shape must be improved, see Figure 4C as an example for
TmPep1050H60A H307A. For further uses, seeds can be stored at -20 °C.

4. X-ray diffraction

4.1. Crystal picking

NOTE: Sample preparation depends on the X-ray source facility (home facility vs. synchrotron). Use storage devices (vials and vial-holder basket) accordingly. The addition of cryoprotectant (such as glycerol) may be required depending on salt/precipitation agent concentration. For TmPep1050 crystals, a cryoprotectant is not necessary as PEG or buffer concentration is high enough to avoid water crystals.

4.1.1. Prepare a bath filled with liquid nitrogen, plunge any vials or basket used for sample handling.

4.1.2. Set up sample picking loops of different sizes: 100, 150, and 200 µm (Table of Materials). Choose the sizes according to the crystal size.

4.1.3. Using a binocular, check the drop containing crystals and spot isolated crystals (the easiest to pick). With a loop, gently pick a crystal from the bottom.

4.1.4. Immediately plunge the loop in liquid nitrogen and place the loop in a suitable vial.

4.2. Data collection

NOTE: Data collection may vary greatly depending on the X-ray source (home facility vs. synchrotron) and detector sensitivity. The collection strategy may also differ greatly from a sample to another, depending on the resolution, spot intensity, space group, etc. The topic has been extensively reviewed by Dauter.

4.2.1. Mount the loop carrying the crystal on the goniometer head of the diffractometer.

4.2.2. Tune the goniometer head along XYZ axes to align the crystal with the X-ray beam path.

4.2.3. Set the wavelength to 0.98 Å and move the detector to get 2 Å of resolution.

4.2.4. Start a short data collection by acquiring images in at least two different crystal orientations. Take 10 images (1 image per 0.1°) at 0° and 90°.

4.2.5. Check the collected images with suitable software (e.g., ADXV, XDS-Viewer or Albula Viewer). Determine the spot intensity and the highest resolution where spots are seen. Check also the monocrySTALLinity and spot separation.

4.2.6. Eventually, repeat steps 4.2.3–4.2.5 by changing the detector position for higher or lower
resolution and the exposure time in accordance to the observation.

4.2.7. Start data collection around 360° with 1 image taken per 0.1°. Remember to set the detector position and exposure time optimally.

5. Indexation, molecular replacement and model building

5.1. Indexation

NOTE: Indexation is a method for measuring diffraction spots intensity, giving the amplitudes of structure factors. Four software packages are commonly used for processing collected images: Mosflm, HKL2000, DIALS, and XDS. The latter has been used for indexing the data sets obtained from TmPep1050 crystal diffraction.

5.1.1. Install XDS package and XDSME. If processing HDF5 files, install XDS Neggia plugin (available on Dectris website). For more information, visit XDS wiki web page [link] and XDSME web page [link].

5.1.2. Before processing data, create a folder from where XDS will be run. Locate the path to the images.

5.1.3. To run XDSME, type in a terminal window: `xdsme /path_to_images/image.extension`

5.1.4. After XDS has ended the job, check the CORRECT.LP file. Note the probability of the space group determination, data completeness, the highest resolution, crystal mosaicity, and data quality. Check also XDS_pointless.log to obtain the likelihood of space groups.

NOTE: See Figure 5 as an example of output.

5.1.5. Rerun XDSME with different space group solutions proposed by XDS in separate folder to avoid overwriting the previous process. Type `xdsme -s space_group_name -c "unit_cell_parameters" /path_to_images/image.extension` (e.g., `xdsme -s P21 -c "43.295 137.812 61.118 90.000 110.716 90.000"`).

5.1.6. Check the CORRECT.LP files and choose the best solution based on data statistics.

5.1.7. Run XSCALE by typing `xscale.py XDS_ASCII.HKL`. Run XDSCONV by typing `xdsconv.py XSCALE.HKL ccp4`.

NOTE: In some cases, XDSME fails to identify the space group or fails to cut the resolution range properly or generates weird data statistics. If such a problem is encountered, it is worth to run XDS natively. Several parameters must be introduced in the XDS.INP initiation file (see XDS wiki [link]).
When using XDS, the likelihood of possible space groups can be checked by using Pointless, part of CCP4 package. To cut the data set resolution, $R_{\text{meas}} < 60\%$ and $I/\sigma \sim 2$ are commonly accepted to determine the highest resolution. The molecular replacement and model refinement, however, can be improved by extending the resolution to $I/\sigma \sim 0.5$–1.5 and $CC_{1/2}$ down to 0.2–0.4.

5.2. Molecular replacement

**NOTE:** Experimental data give access to the amplitude of structural factors but, without knowing the phase, they are useless. The phase can be determined experimentally by different methods relying on an anomalous signal (from a heavy atom, for instance). Molecular replacement is another method for determining the phase without an anomalous scattering atom. This method uses the coordinates of a related molecule to find and improve the phase iteratively. We use Phaser in Phenix GUI for molecular replacement.

5.2.1. Prepare the starting model for molecular replacement using 4P6Y coordinates. From the pdb file, extract the monomer A and truncate its aminoacids in alanine using the PDB file editor in Phenix (under the **Model tools** tab).

5.2.2. Run Xtriage in Phenix (under the **Data analysis** tab) with the reflection file generated by XDSCONV (5.1.9) and the sequence as inputs.

5.2.3. Check the log file from Xtriage. Note the completeness, the number of subunits in the asymmetric unit, the anisotropy, the presence of ice rings, and twinning occurrence.

5.2.4. Run Phaser-MR in Phenix (under the **Molecular replacement** tab) for molecular replacement using the reflection file, the sequence and the starting 4P6Y model truncated in polyalanine (step 5.2.1).

5.2.5. Upon completion, check if a model has been found and the score of the molecular replacement.

**NOTE:** A translation factor Z-score (TFZ) of at least 8 indicates that the solution is definitively correct.

5.3. Model building

**NOTE:** After determining the phase by molecular replacement, the model must be built and refined. This protocol uses Phenix GUI for automatic building and iterative refinements, and Coot for manual structure building and refinement.

5.3.1. After molecular replacement using Phase-MR in Phenix, select **Run Autobuild**. All the required files will be automatically added. Simply press **Run** to start autobuild.
5.3.2. Upon completion, check the model in Coot. Build and refine the model manually according to the electron density map in Coot.

5.3.3. Refine the manually curated model in Phenix (in the Refinement tab) using the model, the sequence, and the diffraction data as inputs. Refer to Phenix help to choose the right strategy.

5.3.4. After refinement, check the results: $R_{\text{free}}$ and $R_{\text{work}}$ must decrease, Molprobity indicators must be respected, and outliers with low real-space correlation must be limited.

5.3.5. Repeat steps 5.3.2 to 5.3.4 until the best refined model is generated.

5.3.6. Run Molprobity on the server: http://molprobity.biochem.duke.edu/. Check any outliers identified by Molprobity.

5.3.7. Eventually repeat steps 5.3.2 to 5.3.6 until the best refined model is obtained.

**REPRESENTATIVE RESULTS:**

To study a possible dodecamer dissociation into monomers in TmPep1050, the His-60 and His-307 codons were replaced by alanine codon using a synthetic gene. This gene was then cloned in pBAD vector for expression and purification of the corresponding TmPep1050 variant subsequently named TmPep1050-H60A H307A. Size exclusion chromatography (Figure 3B) showed that the purified protein had an apparent molecular weight of 56 kDa (molecular weight of the monomer being 36.0 kDa). A similar apparent molecular weight, 52 kDa, has been reported for TmPep1050 dimer. Hence, the oligomeric state of TmPep1050-H60A H307A could be inferred as dimeric. Regarding its specific activity, TmPep1050-H60A H307A was completely inactive on L-Leu-pNA as substrate, even in the presence of cobalt ions. This result strongly suggests that the variants cannot bind any metal ions.

The crystallization condition of TmPep1050-H60A H307A was optimized by varying pH vs. PEG concentration (Figure 4) around the condition of the dimer (i.e., 0.1 M sodium citrate pH 6.0 10% PEG3350). The best crystals of TmPep1050-H60A H307A were obtained in 0.1 M sodium citrate pH 5.2 20% PEG3350 with one cycle of microseeding for improving monocrystallinity. A complete data set was collected at Proxima 2 beamline (SOLEIL synchrotron) at a resolution 2.36 Å (Table 1). Data indexation showed that the space group of the TmPep1050-H60A H307A crystal is C222₁ but XDS proposed another solution, the $mP$ space group (see Figure 5). According to Pointless, the likelihood of C222₁ and P2₁ space groups were 0.711 and 0.149, respectively. According to the data quality analysis, two monomers are found in the asymmetric unit. The analysis by Xtriage revealed that the data set is probably twinned but twinning in C222₁ space group is unlikely. Twinning results from crystal growth anomaly where several definite domains have some of their lattice directions parallel to each other. Twinning may also result from a higher crystal symmetry, indicating an erroneous data indexation. Hence, a pseudo-merohedral twin may exist so that a P2₁ crystal lattice looks like a C222₁. The data set was subsequently indexed in space group P2₁ and tested in molecular replacement. Xtriage analysis of the data set indexed in P2₁ revealed a pseudo-merohedral twin following a twin law $h, -k, -l$. 
Using the coordinates of a monomer from dodecameric TmPep1050 (PDB code 4P6Y), a molecular replacement solution was found for the data set indexed in P2₁ only, with a TFZ score of 28.9. Therefore, the diffraction data were treated as a twinned data set for model building. To minimize the bias of molecular replacement, a first model was built by using phenix.autobuild [54,55]. The structure of TmPep1050_{H60A \ H307A} was completed after several cycles of automated and manual refinement in Phenix and Coot (Table 1 and Figure 6A). The structure confirmed the oligomeric state with an interface surface of 1,710 Å² between both monomers and a $\Delta G$ of -16.2 kcal mol⁻¹ as calculated by PDBe Pisa [56]. In comparison, the interface surface and $\Delta G$ of dimeric TmPep1050₂-mer is 1,673 Å² and -16.7 kcal mol⁻¹, respectively.

The structure of TmPep1050_{H60A \ H307A} is highly similar to the wild-type dimer structure with RMS of 0.774 Å upon alignment. Importantly, the same structural modifications are observed in both structure: high flexibility of the α8 and α10 helices, disordered active site Gln-196–Val-202, and the displacement of Lys-229–Ala-235 and Lys-247–Ser 254. These modifications were correlated previously with the hindrance of dodecamer formation in absence of its metal cofactor [11]. The two mutations of His-60 and His-307, however, had a slight effect on the side chains of Asp-168 and Asp-62. They appeared to be locked in a conformation different from the wild-type dimer (Figure 6B). The Asp-168 carboxylate is rotated by 40° due to the absence of His-60 and His-307. Hence both histidine residues are important for positioning the Asp-168 carboxylate correctly for bridging the two metal ions. The Asp-62 side chain is oriented towards Glu-18 carboxylate, outside the catalytic site. Asp-62 may have an important role in catalysis as it is assumed to modulate the $pK_a$ of His-60 and, thus, influence metal ion binding in the M2 site. In addition, it could be implicated structurally in the stabilization of the catalytic site upon metal ion binding, favoring the formation of the dodecamer.

**FIGURE AND TABLE LEGENDS:**

**Figure 1:** Schematic representation of TM_1050 ORF cloning into pBAD vector by homologous recombination. The ORF is flanked by two 30 bp sequences homologous to the promoter BAD end and the sequence upstream Pmel restriction site.

**Figure 2:** Chromatograms of TmPep1050 purification. (A) Anion exchange chromatography. (B) Hydrophobic interaction chromatography. The absorbance (Abs), expressed in milliunits of absorbance (mUA), is shown in plain line. The conductivity, expressed in mS cm⁻¹, is shown in dashed line. The grey box indicates where TmPep1050 eluates on the chromatograms.

**Figure 3:** Size exclusion chromatography of (A) TmPep1050 dodecamer, (B) TmPep1050_{H60A \ H307A}, and (C) TmPep1050_{H60A \ H307A}. Samples were analyzed using Superdex 200a size exclusion chromatography [57] resin packed in a XK-16/70120-ml column. Absorbance (Abs) is expressed in milliunits of absorbance (mUA). (D) Calibration of the SEC column using thyroglobulin (T), ferritin (F), aldolase (Ald), conalbumin (C), and albumin (Alb) as standards. The correlation between the logarithm of the relative mass and the elution volume is linear, with a $R^2$ of 0.91. The 95% confidence intervals are represented as dots.
Figure 4: Optimization of TmPep1050\textsubscript{H60A H307A} crystallization. (A) The first optimization strategy consists of varying pH (between 4.5 and 6.0) vs. PEG3350 concentration (between 5% and 25%). The crystallization plate is schematized, and the wells are color coded: red for precipitate, yellow for polycrystals, and green for monocrystals. (B) The second optimization strategy includes the use of seeds diluted 25x with a narrower variation of pH vs. PEG3350. (C) Crystal shape and size before (upper image) and after (lower image) crystallization optimization and microseeding.

Figure 5: Excerpts from the log output CORRECT.LP of TmPep1050\textsubscript{H60A H307A} data indexation by XDS. Upper panel, the possible Bravais lattices, the most likely being \textit{mC}, \textit{mP}, and \textit{oC}. Middle panel, overall statistics of data indexed in C222\textsubscript{1} space group. Lower panel, overall statistics of data indexed in P2\textsubscript{1} space group.

Figure 6: Structure of TmPep1050\textsubscript{H60A H307A}. (A) Structural alignment of a TmPep1050\textsubscript{H60A H307A} subunit (red, PDB code 5NE9) vs. a dodecamer subunit (white, PDB code 6NW5) and a dimer subunit (blue, PDB code 5NE6). Arrows indicate the structural dissimilarities between dodecamers and dimers. (B) Close-up of the TmPep1050\textsubscript{H60A H307A} active site (red) compared to the active site of TmPep1050 dimer (blue) and dodecamer (white).

Table 1: Data collection and refinement statistics. Values in parentheses are for the highest-resolution shell.

DISCUSSION:

The protocol described herein allows understanding the dimer-dodecamer transition of TmPep1050 at the structural level. The methodology was experienced previously for determining the structure of both TmPep1050 oligomers\textsuperscript{11}. The most challenging step was to find conditions promoting the dissociation of dodecamers into stable dimers. Such conditions had to be mild enough to permit the reassociation of dimers into dodecamers when the metal ion cofactor was added. The separation of oligomers was also a critical step as it conditions the structural studies and further biochemical characterization (e.g., studying the dodecamer reassociation in various doses of Co\textsuperscript{2+}). The molecular replacement, a proven method for phase determination, was used to solve the structures of TmPep1050 oligomers and its variants. The proposed protocol may be adapted to study other metallo-enzymes whose oligomerization states depend on the availability of their metal cofactors.

To illustrate the protocol, a case of study was presented, TmPep1050\textsubscript{H60A H307A} whose metal binding sites were impaired by mutating His-60 and His-307 to alanine. These residues bind Co\textsuperscript{2+} at the M2 and M1 sites, respectively. Interfering in metal binding could have perturbed the oligomerization state and led to a complete dissociation into monomers. Evidences of such a phenomenon have been reported for PhTET2 and PfTET3, two M42 aminopeptidases from \textit{P. horikoshii} and \textit{P. furiosus}, respectively\textsuperscript{24,29}. TmPep1050\textsubscript{H60A H307A} did not behave as expected as this variant formed dimers only. Its structure showed the same modifications as the wild-type dimer but with two small exceptions. Indeed, the side chains of Asp-168 and Asp-62 appeared to be locked in an unconventional orientation preventing the stabilization of the active site. Their orientation seemed to be imposed by His-60 and His-307 as such modifications were not
observed in the single point mutation variants.

ACKNOWLEDGMENTS:
We thank Martine Roovers for proofreading this paper and giving constructive comments. Access to Proxima 2 beamline (SOLEIL synchrotron) was within Block Allocation Groups 20151139.

DISCLOSURES:
The authors have nothing to disclose.

REFERENCES:


43. Wlodawer, A., Minor, W., Dauter, Z., Jaskolski, M. Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular


