A highly stable Zn(II) dependent phospholipase C obtained by consensus design

Diego S. Val [1], Luisina Di Nardo [2], Fiorela Marchisio [1], Salvador Peiru [3], María Eugenia Castelli [1], Luciano Abriata [4], Hugo G. Menzella [1] and Rodolfo M. Rasia [2,4,5,*]

[1] Instituto de Procesos Biotecnológicos y Químicos (IPROBYQ), FbioyF-UNR-CONICET. Mitre 1998, 2000- Rosario, Argentina;[2] Instituto de Biología Celular y Molecular de Rosario (IBR), FbioyF-UNR-CONICET. Ocampo y Esmeralda, 2000-Rosario, Argentina; [3] Keclon S.A. Cerrito 847, 2000-Rosario, Argentina; [4] Ecole Polythechnique Federale de Lausanne, Rte Cantonale, 1015 Lausanne, Switzerland; [5] Plataforma Argentina de Biología Estructural y Metabolómica, Ocampo y Esmeralda, 2000-Rosario, Argentina

* Corresponding author: Rodolfo M. Rasia. Email: rasia@ibr-conicet.gov.ar

Abstract

Proteins' extraordinary performance in recognition and catalysis have led their use in a range of applications. But proteins obtained from natural sources are oftentimes not suitable for direct use in industrial or diagnostic setups. Natural proteins, evolved to optimally perform a task in physiological conditions, usually lack the stability required to be used in harsher conditions. Therefore, the alteration of the stability of proteins is commonly pursued in protein engineering studies.

Here we achieved a substantial thermal stabilization of a bacterial Zn(II) dependent phospholipase C by consensus sequence design. We retrieved and analyzed sequenced homologs from different sources, selecting a subset of examples for expression and characterization. A non-natural consensus sequence showed the highest stability and activity among those tested. Comparison of activity and stability parameters between this stabilized mutant with other natural variants bearing similar mutations allow us to pinpoint

the sites most likely to be responsible for the enhancement. We show that the stabilized version of the protein retains full activity even in the harsh oil degumming conditions.

Keywords: consensus sequence design, phospholipase C, thermostable enzymes, enzymatic degumming, circular dichroism

1 Introduction

Phospholipase C enzymes are currently used in the edible oil industry. Crude vegetable oil contains phospholipids that are removed in a process called degumming (Dijkstra and Van Opstal, 1989; Singh et al., 2021). The use of various enzymes in degumming increases the yield of refined oil and has been widely adopted (Cowan et al., 2008; Jiang et al., 2011). This process can benefit from the use of an enzyme variant with enhanced thermal stability.

Folding of a protein and the ensuing stability of its conformation is dictated by its amino acid sequence. There is a wealth of literature inquiring on the relationship between sequence, conformation and stability. In spite of this, the prediction of the stability of a protein from a protein sequence, or even from an experimentally determined structure, remains a largely unsolved problem. Therefore, engineering a protein to improve its thermodynamic stability is still an exceedingly difficult problem (Louis and Abriata, 2021).

Over the last decades, several strategies for the stabilization of protein folds were proposed, ranging from structure-based rational mutagenesis to random mutagenesis and selection and, more recently, machine learning approaches (Jones et al., 2017; Yang et al., 2019). Among these, the use of a consensus sequence was suggested initially by Steipe *et al*. (Steipe et al., 1994) and subsequently tested on several different proteins (Sternke et al., 2019). This strategy takes advantage of the information contained in multiple sequence alignments of homologous proteins. An artificial variant in which each position harbors the

most probable amino acid is likely to result in a protein more stable than each of the individual proteins present in the alignment.

Variants of the phosphatidylcholine preferring *Bacillus cereus* phospholipase C are currently used in edible oil refining (Jiang et al.,2011). We have previously worked in the engineering of this enzyme (Elena et al., 2017; Ravasi et al., 2015), and in the search for natural thermostable variants (Marchisio et al., 2022). In this work we employed the consensus sequence approach to produce a new variant of this enzyme.

2 Methods

2.1 Consensus sequence design and variant selection

A set of 1059 natural homologs of *B. cereus* PLC was obtained using HHblits Alignment as provided by the Gremlin Server (Ovchinnikov et al., 2014) (http://openseq.org/submit.php) and filtered by keeping only those showing all active site Zn(II) ligands: W1; H14; D55; H69; H118; D122; H128; H142 and E146 (*B. cereus* sequence numbering). The consensus sequence (here named ChPLC) was then produced by selecting the most frequent residue present at each position (see supplementary material).

We selected four additional natural sequences present in the starting alignment as well, in order to express them and compare their performance with the consensus variant: *Bacillus cereus* PLC (CePLC), variants of which are used in industrial oil degumming; *Bacillus clarus* PLC (ClPLC), the natural sequence closest to the consensus; *Bacillus manliponensis* PLC (MaPLC), a sequence harboring a set of potentially stabilizing mutations present in the consensus but not in the closest homolog; *Bacillus pseudomycoides* PLC (PsPLC), a natural sequence closer to the *B. cereus* one (see supplementary material).

2.2 Strains, growth conditions and cloning

Escherichia coli Top10 was used for general cloning purposes, and *Corynebacterium glutamicum* ATCC 13869, for expression assays. Both of them were cultivated in Luria-Bertani (LB) medium (tryptone, 10 g/L; yeast extract, 5 g/L; and NaCl 5 g/L) supplemented with 50 mg/L kanamycin at 200 rpm. Growth temperature were 37 °C for *E. coli* and 30 °C for *C. glutamicum*.

We used the CePLC sequence previously constructed in the lab (Ravasi et al., 2015). The remaining sequences were codon-optimized and acquired from Twist Biosciences (CA, United States). The optimized sequences were designed with *Sac*II restriction site in the 5´ end and *Eco*RI in the 3´ end. plasmids were purified and digested restriction enzymes from New England Biolabs. The treated sequences were purified from 1.5 % agarose gel. The plasmid pTGR100 (Ravasi et al., 2015) containing CePLC was treated with *Sac*II and *Eco*RI, purified, and inserts were ligated with T4 DNA ligase from Thermofisher.

2.3 Protein production and purification

For the production of PLCs variants, *C. glutamicum* seeds carrying the expression vectors were grown overnight in LB medium. Then, a 100-fold dilution of the cultures was made and incubated with shaking at 30 °C until OD_{600} reached 0.5. At that moment, the cultures were induced with isopropyl-β-d-thiogalactopyranoside (IPTG, Genbiotech, Argentina) at a final concentration of 0.5 mM and left overnight with agitation. Next, the enzymes were recovered by precipitation with $(NH_4)_2SO_4$ as previously described (Ravasi et al., 2015).

The redissolved enzymes were incubated with Trypsin 0.5 ug/mL for 1h at 37 °C in order cleave the propeptide and then purified by HPLC (Agilent 1200, USA) using a hydrophobic interaction chromatography column Shodex PH-814. The recovered fractions were separated

by SDS-PAGE on 12% gels, stained with Coomassie Brilliant blue and quantified by absorbance measurement at 280 nm.

2.4 Thermal shift assay

For thermal shift assays 5 ul of 1xSYPRO orange were added to 45 ul of 5 uM protein in 5 mM Hepes pH 7. The samples were heated from 25 to 95 °C at a rate of 1 °C per minute in PCR tubes. The fluorescence intensity was measured using a Veriti real-time PCR (Applied Biosystems). Datasets were normalized and analyzed using in-house written Python scripts.

2.5 Activity measurements

Enzymatic activity in aqueous solution was measured using 4-Nitrophenyl phosphorylcholine (NPPC, Toronto Research Chemicals). For time course reactions, 75 ul of 200 mM NPPC were added to a cuvette with 75 ul with 10 uM PLC. For stability experiments (see 2.7), the activity remaining after incubation was determined using an Epoch microplate reader (Biotek) with 50 ul of 10 uM emzyme mixed with 50 ul of 2 mM NPPC.

Oil degumming experiments were conducted basically as described previously (Elena et al., 2017) using 5 mg of PLC per kg of oil. 3 g of crude soybean oil (1200 ppm phosphate) was preincubated at the given temperature in a VP710 stirrer (V&P Scientic Inc.). Subsequently, enzymes were added and the mixture was incubated for 2 hours with continuous stirring.

For phosphate determination, the polar heads were extracted according to Elena *et al.* (Elena et al., 2017). Following, 50 ul of aqueous solution were recovered and treated with 0.3 U of calf intestinal phosphatase (NEB) for 1 h at 37 °C. Finally, inorganic phosphate was determined relative to the Sumner method (Sumner, 1944).

2.6 CD spectroscopy

Circular dichroism spectra were acquired on a Jasco J-1500 spectropolarimeter equipped with a Peltier temperature controlling unit. Far UV spectra, from 210 to 240 nm, were obtained averaging four scans in a 0.1 cm path cuvette and converted to deg/cm^2.dmol. Near UV spectra, from 250 to 320 nm were obtained averaging eight scans in a 1 cm micro cuvette and converted to deg/mol.

2.7 Evaluation of the stability of the variants

Enzyme samples were incubated in buffer at the selected temperature for different time steps and the residual activity was measured using NPPC. The decrease in activity with time was then fit to an exponential function to obtain the denaturation rate at the selected temperature. Finally, an Arrhenius plot ($\ln(k_{obs})$ vs. 1/T) was produced for each variant to obtain estimates for the activation energies.

3 Results and discussion

3.1 Protein expression, spectroscopy and activity assays

We expressed all protein variants using *C. glutamicum* ATCC 13869 as expression host. All proteins were expressed at high levels and could be purified to homogeneity (see supplementary material). To assess the correct fold of the expressed proteins, far UV and near UV CD spectra were obtained (Figure 1). The far UV CD spectra of all variants were similar, showing that they acquire the expected helical fold. Near UV CD showed small differences between the proteins with a significant signal intensity indicating the presence of structured aromatic residues in the fold and a stable tertiary structure in all cases.

We then checked the variants for phospholipase C activity with the chromogenic substrate NPPC. Initial rate measurements of the reaction with CePLC resulted in a linear dependence of the reaction rate with substrate concentration (see supplementary material) showing that the K_m of BcPLC for this substrate is much larger than 10 mM. We therefore decided to use substrate concentrations much lower than the enzyme K_m to directly measure k_{cat}/K_m by fitting reaction time courses to exponential functions (see supplementary material). The obtained second order rate constants at 30 to 50 °C were similar for all variants, with the consensus protein showing the highest activity and the ClPLC and PsPLC variants showing the lowest activity.

3.2 Evaluation of the stability of the different variants

We obtained an initial estimate of the stability of the different variants by means of a differential scanning fluorimetry assay using SYPRO Orange as dye (Figure 1). The transition temperatures range from 50 to 95 °C, increasing in order PsPLC < ClPLC < MaPLC < CePLC < ChPLC (Table 1). The melting point for the consensus variant as determined by this method is almost 20 °C higher than that of the more stable natural variant tested, CePLC. Reaction time courses acquired between 55 and 80 °C show the same trend. In these experiments both substrate hydrolysis and enzyme inactivation occur simultaneously. ChPLC reaction rate increases continuously in the whole range, but the natural variants are inactivated as the temperature increases (see supplementary material).

Chemical and thermal unfolding of mature CePLC are irreversible, thus the stability cannot be measured by equilibrium measurements. For this reason, we resorted to the measurement of thermal unfolding kinetics to provide estimates for the activation energy (E^a) of the irreversible denaturation step.

When protein unfolding is irreversible, incubation of enzyme samples at high temperature will result in a decay of the activity with a characteristic first order rate constant. The amount of active enzyme remaining in solution can be estimated by measuring the residual activity at a lower temperature. The first order denaturation rate constants at different temperatures can be subsequently used to obtain an estimate for the activation energy of the irreversible denaturation step (Duy and Fitter, 2005). These energies are used in turn to compare quantitatively the relative stability of the different variants.

We obtained Arrhenius plots for all five variants tested (Figure 1). The estimated ΔG^{++}_{denat} decrease in order ChPLC >> CePLC > MaPLC ≈ ClPLC > PsPLC (Table 1). Unfolding of ChPLC requires 17 kcal/mol more than that of the more stable natural variant tested.

Zn(II) binding in the active site contributes significantly to the overall stability of the folded protein (Little and Johansen, 1979). Therefore, stability could be linked to Zn(II) affinity. We determined the relative Zn(II) off-rate of the variants by measuring the activity loss in the presence of the metal chelator EDTA (see supplementary material), but found no significant differences between them, ruling out this hypothesis.

Finally, the relative stabilities of the apoenzymes was evaluated by measuring the fraction of folded protein remaining as a function of temperature (see supplementary material). We found that the more stable variants in the apo-form are ChPLC and MaPLC. This finding goes in line with a particular region in the consensus protein contributing significantly to the overall increase in stability observed (see 4.4).

3.3 Activity in crude oil

Phospholipases C are employed in enzymatic degumming of edible oils. The reaction conditions in this process are radically different from the reaction in solution potentially changing the stability trends observed in aqueous solution. We therefore performed degumming reactions at temperatures ranging from 50 to 90 °C (Figure 1 and Table 1). All natural variants show a decrease in activity with increasing temperature, down to ca. 1/3 of the maximum at 90 °C. In contrast, the consensus variant shows a slight increase in activity up to 80 °C, lowering the efficiency at 90 °C but still releasing twice as much phosphate as any of the other natural variants at this temperature.

3.4 A structural basis for the stability of ChPLC

We modeled the structures of the apo forms of all variants produced in order to understand the higher stability of ChPLC using AlphaFold2 (Jumper et al., 2021) (Figure 2). The interatomic contacts of the structures (Arpeggio web server (Jubb et al., 2017), (see supplementary material) show a higher number of aromatic and hydrophobic contacts for ChPLC and MaPLC than for the other variants. These two variants display the highest melting temperature in the apo form. A detailed analysis of the environment of aromatic residues in the modeled structures show a stacking interaction between F120 and Y234 in both ChPLC and MaPLC. A phenylalanine in position 106, present only in ChPLC, can also participate in an aromatic cluster opposite the active site.

The larger increase in E^a and melting temperature found for holo ChPLC indicates that the Zn(II) site contributes significantly to the stability of this variant. When analyzing the second shell ligand residues in the active site, we found that ChPLC introduces a threonine in position 121 where all other variants tested have a glycine. The T121 sidechain can H-

bond with the sidechain of H14, one of the Zn(II) ligands, potentially increasing the stability of the Zn(II) site and thus of the protein as a whole. Other noteworthy mutations differentiating the consensus sequence from the natural variants tested are the N85 and T88 residues, that introduce hydrogen bonds bridging a loop behind the active site to helix 4 in the structure.

4 Conclusions

A stability versus activity tradeoff is oftentimes reported in mutagenesis studies. This effect is usually ascribed to the need of flexibility for catalysis to be efficient. This is not the case here, where the consensus variant is not only the most stable, but also appears to be as active as the more efficient variant even at lower temperatures.

The consensus variant also outperformed all natural variants in degumming assays, showing the highest activity at 80 °C, thus makings it extraordinarily valuable in this process refinement. We expect that this improved enzyme will find its way in the edible oil industry.

5 Acknowledgements

Funding from projects PICT-2016-4124 and PICT 2019-01528. LDN, FM and DSV received fellowships from CONICET.

6 References

Cowan, D., Oxenbøll, K.M., Holm, H.C., 2008. Enzymatic bioprocessing of oils and fats. Inf. - Int. News Fats Oils Relat. Mater. 19, 210–212.

Dijkstra, A.J., Van Opstal, M., 1989. The Total Degumming Process.

Duy, C., Fitter, J., 2005. Thermostability of irreversible unfolding α-amylases analyzed by unfolding kinetics. J. Biol. Chem. 280, 37360–37365. https://doi.org/10.1074/jbc.M507530200

Elena, C., Cerminati, S., Ravasi, P., Rasia, R., Peiru, S., Menzella, H.G., Castelli, M.E., 2017. B. cereus phospholipase C engineering for efficient degumming of vegetable oil. Process Biochem. 54, 67–72. https://doi.org/10.1016/j.procbio.2017.01.011

Jiang, F., Huang, S., Imadad, K., Li, C., 2012. Cloning and expression of a gene with phospholipase B activity from Pseudomonas fluorescens in Escherichia coli. Bioresour. Technol. 104, 518–522. https://doi.org/10.1016/j.biortech.2011.09.112

Jiang, F., Wang, J., Kaleem, I., Dai, D., Zhou, X., Li, C., 2011. Degumming of vegetable oils by a novel phospholipase B from Pseudomonas fluorescens BIT-18. Bioresour. Technol. 102, 8052–8056. https://doi.org/10.1016/j.biortech.2011.05.050

Jones, B.J., Lim, H.Y., Huang, J., Kazlauskas, R.J., 2017. Comparison of Five Protein Engineering Strategies for Stabilizing an α/β-Hydrolase. Biochemistry 56, 6521–6532. https://doi.org/10.1021/acs.biochem.7b00571

Jubb, H.C., Higueruelo, A.P., Ochoa-Montaño, B., Pitt, W.R., Ascher, D.B., Blundell, T.L., 2017. Arpeggio: A Web Server for Calculating and Visualising Interatomic Interactions in Protein Structures. J. Mol. Biol. 429, 365–371. https://doi.org/10.1016/j.jmb.2016.12.004

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S.A.A., Ballard, A.J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer,

T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A.W., Kavukcuoglu, K., Kohli, P., Hassabis, D., 2021. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589. https://doi.org/10.1038/s41586-021-03819-2

Little, C., Johansen, S., 1979. Unfolding and refolding of phospholipase C from Bacillus cereus in solutions of guanidinium chloride. Biochem. J. 179, 509–514. https://doi.org/10.1042/bj1790509

Louis, B.B.V., Abriata, L.A., 2021. Reviewing Challenges of Predicting Protein Melting Temperature Change Upon Mutation Through the Full Analysis of a Highly Detailed Dataset with High-Resolution Structures. Mol. Biotechnol. 63, 863–884. https://doi.org/10.1007/s12033-021-00349-0

Marchisio, F., Di, L., Diego, N., Val, S., Cerminati, S., Espariz, M., Rasia, R.M., Gabriel, H., María, M., Castelli, E., 2022. Characterization of a novel thermostable phospholipase C from T . kodakarensis suitable for oil degumming. Appl. Microbiol. Biotechnol. https://doi.org/10.1007/s00253-022-12081-z

Ovchinnikov, S., Kamisetty, H., Baker, D., 2014. Robust and accurate prediction of residue-residue interactions across protein interfaces using evolutionary information. eLife 2014, 1–21. https://doi.org/10.7554/eLife.02030

Ravasi, P., Braia, M., Eberhardt, F., Elena, C., Cerminati, S., Peirú, S., Castelli, M.E., Menzella, H.G., 2015. High-level production of Bacillus cereus phospholipase C in Corynebacterium glutamicum. J. Biotechnol. 216, 142–148. https://doi.org/10.1016/j.jbiotec.2015.10.018

Singh, R., Arora, A., Singh, V., 2021. Biodiesel from oil produced in vegetative tissues of biomass – A review. Bioresour. Technol. 326, 124772. https://doi.org/10.1016/j.biortech.2021.124772

Steipe, B., Schiller, B., Pluäckthun, A., Steinbacher, S., 1994. Sequence statistics reliably predict stabilizing mutations in a protein domain. J. Mol. Biol. https://doi.org/10.1006/jmbi.1994.1434

Sternke, M., Tripp, K.W., Barrick, D., 2019. Consensus sequence design as a general strategy to create hyperstable, biologically active proteins. Proc. Natl. Acad. Sci. U. S. A. 166, 11275–11284. https://doi.org/10.1073/pnas.1816707116

Sumner, J. B., 1944. A method for the colorimetric determination of phosphorus, Science 100, 413–414. DOI: 10.1126/science.100.2601.413

Yang, K.K., Wu, Z., Arnold, F.H., 2019. Machine-learning-guided directed evolution for protein engineering. Nat. Methods 16, 687–694. https://doi.org/10.1038/s41592-019-0496-6

6.1 Figure captions

6.2 Figure 1.

Biophysical and kinetic characterization of the enzyme variants tested. Top left, far UV CD spectra; top center, near UV CD spectra. Top right, thermal shift assay with Sypro Orange. Bottom left, Arrhenius plots showing the inactivation rate of the enzymes at different temperatures. Bottom right, activities in crude oil, expressed as amount of phosphate released, at different temperatures.

6.3 Figure 2.

Highlights of structural features at the base of the stability increase of ChPLC. The sequences of the regions displayed are shown at the top left. Regions of the structure of CePLC (PDB 1AH7) and the AlphaFold2 modeled structures of ChPLC and MaPLC are overlaid

with the sidechain carbon atoms color coded as follows: CePLC, cyan; MaPLC, light green; ChPLC, dark blue.

7 Tables

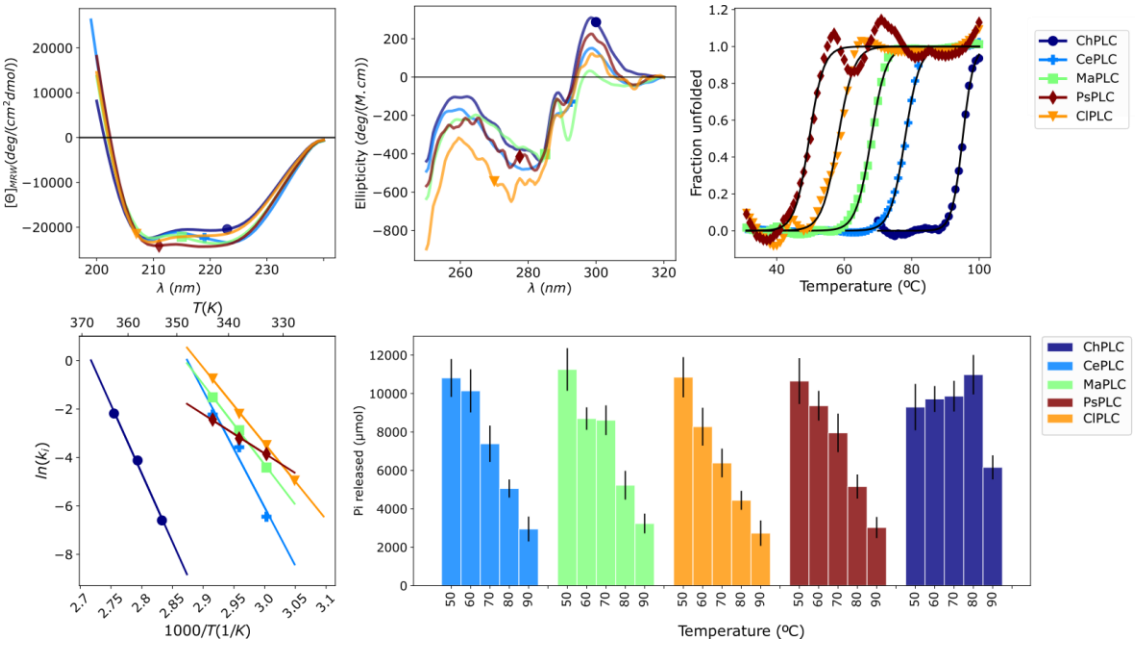
7.1 Table 1

Enzyme	kcat/Km (1/(M.s))	Tm (TSA, °C)	E^a (kcal/mol)	Tm apo (°C)	Koff Zn(II)(1/min)	Pi release 80/50 °C
ChPLC	34.9	95.1	112.6	58.8	3.6	1.18
CePLC	30.8	78.0	96.0	40.7	3.1	0.47
MaPLC	14.5	68.0	65.9	58.4	3.1	0.46
PsPLC	15.9	49.6	57.6	49.9	4.0	0.48
ClPLC	23.3	58.1	62.4	38.4	8.3	0.41

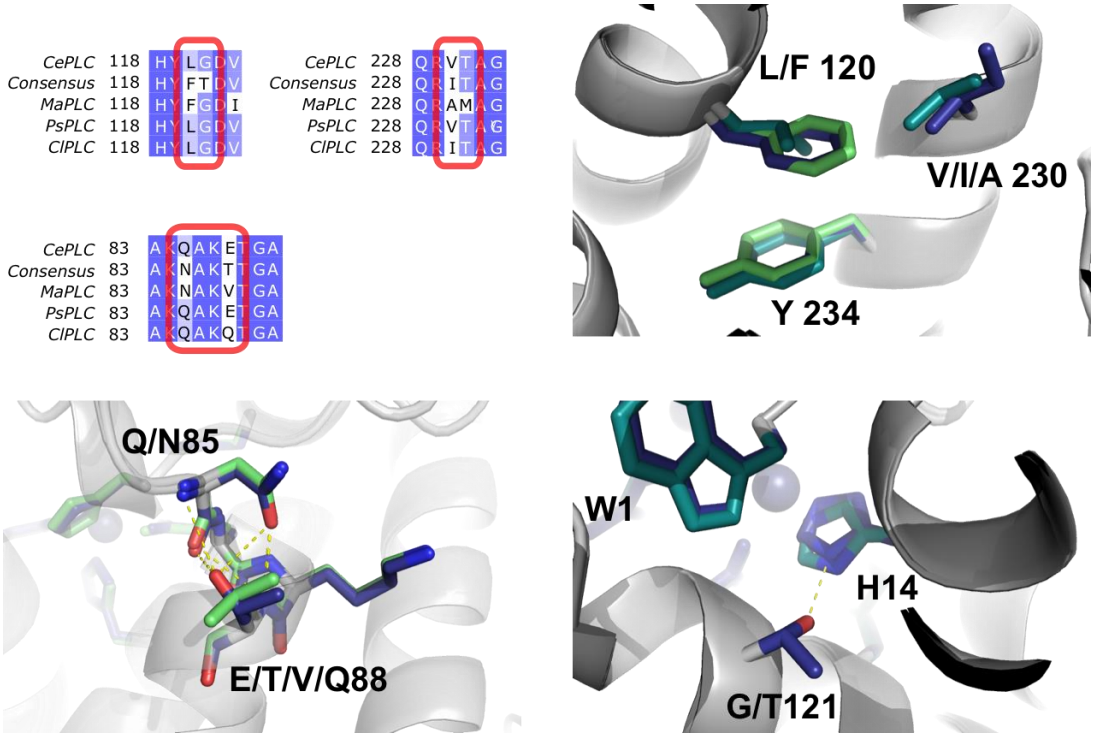
Table 1. k_{cat}/K_m, catalytic efficiency obtained at 40 °C using NPPC as substrate. Tm, denaturation temperatures measured by thermal shift assay. E^a, activation energy for thermal denaturing measured by activity loss. Tm apoenzyme, melting temperature of the apoenzymes measured by far UV CD. Pi release, relative degumming activity at 80 vs 50 °C.

8 Figures

8.1 Figure 1



8.2 Figure 2


CePLC 118 HYLGDV
Consensus 118 HYFTDV
MaPLC 118 HYFGDI
PsPLC 118 HYLGDV
ClPLC 118 HYLGDV
CePLC 228 QRVTAG
Consensus 228 QRITAG
MaPLC 228 QRAMAG
PsPLC 228 QRVTAG
ClPLC 228 QRITAG
CePLC 83 AKQAKETGA
Consensus 83 AKNAKTTGA
MaPLC 83 AKNAKVTGA
PsPLC 83 AKQAKETGA
ClPLC 83 AKQAKQTGA
L/F 120
V/I/A 230
Y 234
Q/N85
E/T/V/Q88
W1
H14
G/T121