4-6-2018

Critical Amino Acids for the Folding of Carboxypeptidase O

Donn LaTour
Andrews University, latour@andrews.edu

J.N. Andrews Honors Program
Andrews University

HONS 497
Honors Thesis

Critical Amino Acids for the Folding of Carboxypeptidase O

Donn LaTour

April 6, 2018

Dr. Peter Lyons

Primary Advisor Signature:_____

Department of Biology

ABSTRACT

Enzymes must fold properly to function. For carboxypeptidases, a prodomain usually assists in this process. However, carboxypeptidase O (CPO) can fold without one. We hypothesized that this prodigious ability could arise from an amino acid whose fortunate positioning stabilizes its neighbors during folding. We identified four unique, conserved amino acids and mutated them to their equivalent counterpart in CPA, a cousin of CPO, which requires a prodomain. We expressed the mutated genes in mammalian cells and checked for expression by western blotting. Our results indicate that three of four mutations were not expressed, hinting at the possible importance of these sites in the folding of CPO.

INTRODUCTION

Carboxypeptidases carry out highly specific cleavage of C-terminal amino acids from a longer peptide chain. CPO is a metallocarboxypeptidase, and as explained by Gomis-Rüth (2008), its catalytic reaction is usually stabilized by one or two divalent metal ions which help to hold the substrate in place for the duration of the reaction. Catalysis proceeds through a general acid/base mechanism. Metallopeptidases are named after the kind of amino acid they selectively cut. For example, one enzyme cleaves terminal *aliphatic* amino acids, and so is named carboxypeptidase A, or CPA. Another enzyme cleaves terminal *basic* amino acids, and so is fittingly named CPB. Lyons and Fricker (2011) point out that until recently, it was not known how terminal acidic amino acids were cleaved. The discovery of CPO, a few years ago, through bioinformatics analysis, filled this gap in understanding.

Another unique feature of CPO is that it lacks an N-terminal prodomain, which is a strand of amino acids that acts as a cap over the active site, preventing the enzyme from pre-mature activation. For example, CPA and CPB are secreted by pancreatic acinar cells as inactive zymogens and are later activated by trypsin in the small intestines (Ventura *et al.* 1999). A number of studies have shown that prodomains also play an important role in protein folding. Many proteins, including CPA and CPB, require one to fold properly (Gomis Rüth 2008).

Protein folding is a complicated process. The exact pathways of protein folding are still quite theoretical, since some steps happen extremely quickly. One theory of protein folding posits that after an initial endothermic step, where the unfolded protein begins to approximate its folded native state, entropy drives the rest of the process toward the most stable conformation, which happens to be the native state. In this paradigm, the overall free-energy landscape during folding is believed to resemble a rugged funnel, where intermediates are funneled energetically

toward the native state (Onuchic *et al.* 1997). A protein's native state is more thermodynamically stable than the unfolded state, and this free energy difference creates a free-energy gradient which promotes folding. Folding can happen very quickly and it begins soon after translation, possibly even as a protein passes through the ribosomal exit tunnel (Sorokina & Mushegian 2016). The folding process may be affected by a variety of outside influences. It can be constrained by intramolecular bonds such as disulfide bridges that form during the creation of the native state (Bronsoms *et al.* 2003). It may also be sped up by assistant molecules such as chaperones.

Prodomains, specifically class 1 pro-peptides, assist the proteins to which they are attached by stabilizing their transitions states throughout the folding process *in vivo* (Bronsoms *et al.* 2003). Although not absolutely ubiquitous, these extra lengths of amino acid act as intramolecular chaperones that drive the correct folding of an attached protease domain (Ventura *et al.* 1999). This theory is easily demonstrated by a simple experiment. As a digestive enzyme, CPA is not useful within the cell. If the enzyme folds correctly, it is largely excreted into the extracellular region where digestion occurs. When Phillips and Rutter (1996) mutated certain critical amino acids in CPA1's prodomain involved in salt bridges with the active site , the enzyme was not secreted at all, indicating that these mutations to the prodomain crippled CPA1's folding. Improperly folded proteins are not usually secreted, but rather degraded and discarded within the cell, therefore a lack of secretion may indicate a protein's misfolding.

Because it does not code for a prodomain, the CPO gene was initially dismissed as a pseudo-gene, and considered to be nonfunctional. However, Lyons and Fricker show that CPO retains the ability to fold and function without a prodomain. Although this ability isn't exclusive to CPO--there are other enzymes that can also fold and function without a prodomain--it is

unique to CPO's immediate family of carboxypeptidases. Figure 4 shows the carboxypeptidase tree. Of the nine enzymes on the A/B carboxypeptidase evolutionary tree, CPO alone folds without a prodomain. What is the secret to CPO's folding success? It has been shown that the a protein's primary amino acid sequence largely determines folding kinetics by encouraging or discouraging the formation of a native-like state early in the folding pathway, which then quickly folds into the final native state (Rumbley *et al.* 2000). Given that the primary amino acid structure is so important to protein folding, we hypothesize that it is possible that one or more amino acids within the three-dimensional structure of CPO are positioned in such a uniquely favorable position that they stabilize their neighboring amino acids during folding. In doing so, they effectively substitute for the role of a prodomain by driving folding kinetics forward.

METHODS

Gene Alignment and 3D Modeling. We looked for individual amino acids which were critical and therefore conserved in all orthologues of CPO, but conserved and different in CPA and CPB using an alignment in GeneDoc. We analyzed several CPO, CPA, and CPB variants from multiple species. We used Swiss-PDB Viewer to visualize these sites three-dimensionally, and using the tools for identifying neighboring amino acids, searched for potential side-chain or C-terminal interactions that could explain why these sites might be important to CPO's folding.

PCR. We designed forward and reverse primers for the parental pcDNA3.1-hCPO plasmid containing the gene for CPO using Agilent Technologies' *QuickChange Site-Directed Mutagenesis Kit.* Primers were designed to span between 25 and 45 nucleotides in length to promote stability, with a target melting point of 78 degrees Celsius or above, and a GC content of at least 40%. We mixed 5 µl of 10x reaction buffer with 2.5 µl of forward primer and 2.5 µl of

reverse primer, 50 ng/ µl, 1 µl dNTP, 37 µl H_2O, and 0.5 µl PfuULTRA II polymerase for each reaction. PCR occurred at an annealing temperature of 55° C for 16 reaction cycles. To verify the success of PCR, we ran the amplified DNA through a 1% agarose gel stained with ethidium bromide for UV visualization. Next, we precipitated the rest of the amplified DNA with cold 100% ethanol in preparation for bacterial transformation.

Transformation of E. coli, and Amplification and Purification of Plasmid. (We performed these procedures twice, first on a small scale to establish that we had created the correct plasmids, and second on a larger scale to amplify these plasmids. Large-scale volumes and instructions are in parenthesis.) We thawed 50 µl *E. coli* cells for each mutation and added 5 µl of mutant plasmid, then heat shocked the mixture at 42° C for 40 seconds. We incubated the transformed cells in 200 µl Luria Broth for 1 hour at 37° C before incubating on dry LB/amp plates for 16-20 hours. We transferred one colony from each plate to a solution of 3 mL (40 ml) LB and 3 µl (25 µl, then 20 µl about 10 hours later) Ampicillin, and incubated this suspension at 37° C for another 16-20 hours. Then we centrifuged the contents of each tube and re-suspended in 150 µl (4 ml) extraction buffer, which was 100 µg/ml lysozyme and 25 µg/ml RNAse A, dissolved in STET buffer (100mM Sodium Chloride, 10 mM Tris-HCL, 1mM EDTA, and 5% Triton X-100, at pH9.0). This lysed the cells. We incubated the bacterial suspension in extraction buffer at 65° C for 5 minutes (10 minutes), then centrifuged at maximum rpm (13,000 rpm) for 10 minutes (30 minutes in refrigerated centrifuge) and removed the pellet. Then we added 120 µl (2.8 ml) isopropanol to precipitate the plasmid, centrifuged at 7000 rpm (13,000 rpm) for 10 minutes (30 minutes), and discarded the supernatant. We added 500 µl (1.5 ml) of 70% ethanol, centrifuged at maximum rpm for 10 minutes (5 minutes), and again discarded the supernatant.

After air-drying the DNA pellet for 5-10 minutes, we dissolved it in 50 µl of 500 µl 10mM Tris at pH 8.5.

Sequencing. We linearized the plasmid with a restriction digest (we incubated 1 µl of plasmid, 1 µl 10X "Buffer E", 1 µl Bovine Serum Albumin, 0.5 µl Bam H1 restriction enzyme, and 6.5 µl of water at 37° C for 1 hour) and ran each mutant plasmid sample through a 1% agarose gel. Reference to a Tri-Dye 1 kb ladder made identification of the plasmid possible, and plasmid concentration was estimated from each band's brightness. We then sent the purified, mutant plasmids off for sequencing and compared the results to the genomic sequence of wild type CPO to verify the success of mutagenesis.

HEK293T Cell Transfection. We thawed mammalian HEK293T cells quickly from the deep-freezer by swirling in a 37° C water bath. We then plated these cells with 10 mL of Dulbecco's Modified Eagle Medium (DMEM), containing 10% Fetal Bovine Serum (FBS). We incubated the cells for 24-48 hours at 37° C. We then rinsed the cells with Dulbecco's Phosphate Buffered Saline solution (PBS), and added 1 mL Trypsin to release them from the dish. After a 4-5 minute incubation period in the hood, we added new media, and placed the cells back in the incubator. This process of splitting cells was repeated once more after a 48 hour period or more. The exact incubation times and splitting concentrations of cells varied. After replacing the old media, we counted the cells with a hemocytometer, and plated 1.5 million cells in each 25 cm^2 dish. We made a solution of 15 µg PEI, or Polyethylenimine 24 kD linear with 5 µg of mutant plasmid in serum-free DMEM and mixed this with the HEK293T cells after a 15 minute room temperature incubation to promote transfection. We harvested the transfected cells 48 hours post transfection.

Preparing cells for freezing. We prepared 5 mL of lysis buffer that was 20 mM Tris, 150 mM NaCl, 1% NP-40 and also 1:100 RPI protease inhibitor. NP_4O is a detergent that breaks down cellular plasmid membranes, and RPI protease inhibitor was added to slow protein degradation in the freezer. We washed the cells in PBS, then centrifuged them at 4000 RPM for 1 min. We then re-suspended the cells in 1 mL of PBS and spun again, then re-suspended in 500 µl of NP_4O lysis buffer for temporary storage in the freezer.

SDS-PAGE page protein separation. Sodium dodecyl sulfate polyacrylamide gel electrophoresis, or simply SDS-PAGE unfolds proteins and effectively evens out their charge-to-mass ratio so that separation occurs by molecular weight and not other factors such as size or shape. 25 µl samples were loaded into the wells of a premade 4-20% gradient gel. We applied a voltage of 160 V for approximately 60 minutes to allow the proteins to separate.

Western blot analysis of CPO mutations. We assembled the transfer apparatus with the gel pressed against a sheet of nitrocellulose. We transferred the proteins from the gel to the nitrocellulose in of transfer buffer (30.2 gm Tris Base, 144 gm glycine in 1 L, 10% methanol) at a voltage of 100 V for 60 minutes. Next we washed the nitrocellulose on a rocker in 20 mL of 5% milk in TBST (25mM tris, 137mM sodium chloride, 2.7mM potassium chloride and 0.05% TWEEN-20) for 20 minutes. Then we washed for 1 hour in 10 ml of a 1:2000 dilution of rabbit primary antibody solution in 5% nonfat milk dissolved in TBST. We drained the primary antibody and washed with 10 mL of TBST once quickly, then on a rocker for 15 minutes, and twice for 5 minutes. We then washed for 1 hour in 10 ml of a 1-2000 concentration anti-rabbit secondary antibody solution, likewise in 5% nonfat milk dissolved in TBST. Again, the nitrocellulose was washed once quickly with TBST, then on a rocker for 15 minutes, and twice for 5 minutes. We mixed 0.5 ml of each of the two LumiGLO reagents together with 9 ml of

water, and washed the blot in this solution for 1 min. We drained the blot and dabbed it dry, then covered it with one layer of Saran wrap, and taped it to an X-ray film developing cassette. Then we placed a piece of X-ray film over the blot, and exposed for either 1 or 10 minutes.

RESULTS

In order to identify amino acids critical for the structure and function of CPO, we created an alignment of CPO and CPA/CPB from 12 vertebrate species. Our gene alignment, shown in Figure 1, revealed four amino acids, K99, L230, G239, and R299, in CPO that met our criteria of unique conservation in CPO as opposed to equivalent sites in CPA/CPB. In an effort to explain why these sites may be critical for stabilizing CPO during folding, we examined neighboring residues of K99, L230, and R299 to explore possible stabilizing interactions with these sites. Since CPO's crystal structure was not known at the time of this experiment, we used a model of CPO in SwissModeller, with CPA2 (1AYE) as a template. Figure 2a shows a potential interaction between R299 and a polar pocket of amino acid side chains. Figure 2b shows K99's side chain extending toward the carboxyl group of the C-terminus, helping to stabilize it. Finally, Figure 2c shows L230 interacting with a hydrophobic pocket of side chains. These interactions offer potential explanations as to why these amino acids are conserved in CPO. We were unable to suggest any logical interactions for G239, however, the alignment did identify this residue as a conserved amino acid, so we decided to proceed with it anyways.

We mutated these amino acids to their CPA/CPB counterparts as follows: K99P, L230K, G239S, and R299Y. For example, at the 99th amino acid in CPO, K was changed to a P. After transforming *E. coli* with these four mutated pcDNA 3.1-hCPO plasmids, we sequenced the

purified plasmid to confirm the identity of the mutations. Figure 3 shows that all of our mutations were successful.

In order to show that these mutations are critical to stabilizing CPO during folding, we looked for reduced or no expression of mutant CPO following transfection in mammalian HEK293T cells. In experiment 1 (Figure 4a), we performed Western blotting for each mutation in duplicate and included a lane of previously prepared cells transfected with wild-type CPO. All of the mutations caused non-expression. As expected, the wild-type control indicated the presence of CPO. In experiment 2 (Figure 4c), we blotted all four mutations with two extra mutations, Y21H, and W297L, as well as a mock transfection, where the transfecting plasmid was absent, and an additional positive control prepared at the same time as the other samples with a wild-type CPO plasmid. As expected, the mock transfection and all four mutants showed no CPO; however W297L also was CPO-free (previous experiments have shown this mutation to be expressed at wild-type levels). The two wild-type controls showed CPO.

To control for total protein expression levels, we stained the Western blot from experiment 2 in Ponceau S (Figure 4d). In order from highest to lowest protein concentration we observed: R299Y, L230K and G239S, and finally K99P. During the transfection portion of experiment 2, several plates became infected, killing cells, and contributing to the lower levels of protein expression in some lanes of the Western blot. An overexposure of experiment 1's Western blot (Figure 4b) revealed similar levels of nonspecific bands in all lanes, suggesting equal protein amounts in each sample. Contrastingly, the Ponceau S stain in experiment 2 shows widely varying levels of total protein.

DISCUSSION

Our results demonstrate that after the mutations K99P, L230K, G239S, and R299Y, CPO was no longer produced by our HEK293T host cells. We attribute this to misfolding. As discussed by Phillips and Rutter (1996), misfolded proteins are recognized and degraded within the cell. Non-expression of CPO after mutation at any one of these four points supports the case that these amino acids are critical to stabilizing CPO's folding.

In experiment 2, some of the plates of HEK293T cells experienced an infection. The infection played the role of a confounding variable, and by killing cells reduced overall protein expression as well as the expression of CPO, which is the dependent variable in this experiment. The Ponceau S stain of experiment 1 in Figure 4d shows reduced expression in L230K and G239S, but especially in the K99P transfection. This makes it difficult to determine whether CPO's non-expression, especially for K99P, was caused by the mutation or the fact that the cells died and did not produce any protein in general. Fortunately, experiment 1, which did not experience an infection, corroborates the non-expression of CPO in experiment 2. We did not perform a Ponceau S stain in experiment 1. However, we did overexpose a 10-minute Western blot (Figure 4b), which showed an even, baseline response to the RP3-CPO antibody and non-expression of CPO for all four mutations. This is obviously a non-traditional way of controlling for protein expression. The infection also explains why W297L, which has been shown not to affect CPO's expression in previous experiments, was not expressed in experiment 2.

It is interesting that all four of the mutation sites, K99, L230, G239, and R299 lie close to the surface of the protein. If these amino acids were deep within CPO's hydrophobic interior, it would be easy to attribute the results of this experiment to their location, as amino acids deep within any protein play key roles in stabilization. Although amino acids close to a protein's

surface don't tend to have as significant an impact on protein stability as deeper amino acids, their ability to affect protein stability is not unheard of. Researchers observed that modification of the hydrophilic portions of a self-aggregating peptide's α-helix region affected the peptide's folding and catalytic ability (Blondelle *et al.* 2000). So, the position of our mutant amino acids near the surface, combined with their ability to cause CPO non-expression does hint that K99, L230, G239, and R299 may contribute to CPO's prodigious folding behavior.

Our work represents some of the first attempts to try to understand why CPO can fold without a prodomain. To strengthen this inquiry, it would be very interesting to reverse this experiment and mutate the analogues of K99, L230, G239, and R299 in CPA and CPB to their CPO counterparts. As it stands, all we can conclude is that mutation of these sites in CPO destabilizes the enzyme. If the reverse mutations caused CPA or CPB to acquire an ability to fold without their prodomains like CPO, this would greatly strengthen the claim that these amino acids are critical to the unique folding of CPO.

At the time of this experiment, the crystal structure of CPO was not known, so the possible side-chain interactions of the identified critical sites were based off of computer modelling. Our CPO model followed the pattern of the crystal structure of CPA2 (1AYE). Since the completion of our work, CPO's crystal structure has been determined, and the authors of this work are on the verge of publishing a paper on their results (Garcia-Guerrero *et al.* in press). This information will undoubtedly shed more light on the interactions of the amino acids discussed in this paper, and perhaps suggest new potentially critical amino acids for CPO's folding.

Literature Cited

Blondelle, S. E., Esteve, V., Celda, B., Pastor, M. T., Perez-Paya, E. (2000). influence of the hydrophilic face on the folding ability and stability of α-helix bundles: relevance to the peptide catalytic activity. J Peptide Res., 56, 121-131

Bronsoms, S., Villanueva, J., Canals, F., Enrique, Q., Aviles, X. (2003). Analysis of the effect of potato carboxypeptidase inhibitor pro-sequence on the folding of the mature protein. European Journal of Biochemistry, 270, 3641-3650. DOI:10.1046/j.1432-1033.2003.03754.x

Garcia-Guerrero, Maria., Garcia-Pardo, Javier., Berenguer, Esther., Fernandez-Alvarez, Roberto., Barfi, Gifty., Lyons, Peter., Aviles, Francesc., Huber, Robert., Lorenzo, Julia., Reverter, David. (In press). *Crystal structure and mechanism of human carboxypeptidase O, the missing player in dietary protein digestion.*

Gomis-Rüth, F.X. (2008). Structure and Mechanism of Metallocarboxypeptidases. Critical Reviews in Biochemistry and Molecular Biology, 43, 319-345. DOI:10.1080/10409230802376375

Lyons, P. J., & Fricker, L. D. (2011). Carboxypeptidase O Is a Glycosylphosphatidylinositol-anchored Intestinal Peptidase with Acidic Amino Acid Specificity. Journal of Biological Chemistry, 286(45), 39023-39032. DOI:10.1074/jbc.M111.265819

Onuchic, Jose N., Luthey-Schulten, Zaida., Wolynes, Peter G. (1997). Theory of Protein Folding: The Energy Landscape Perspective. Annual Review of Physical Chemistry, 48, 545-600.

Phillips, M. A., & Rutter, W. J. (1996). Role of the Prodomain in Folding and Secretion of Rat Pancreatic Carboxypeptidase A1. Biochemistry, 35(21), 6771-6776

Rumbley, Jon., Huang, Linh., Mayne, Leland., Englander, S. W. (2000). An amino acid code for protein folding. Proceedings of the National Academy of Sciences of the United States of America, 98(1), 105-112, https://doi.org/10.1073/pnas.98.1.105

Sorokina, Irina., & Mushegian, Arcady. (2016). The role of the backbone torsion in protein folding. Biology Direct, 11(64), DOI:10.1186/s13062-016-0166-x.

Ventura, S., Villegas, V., Sterner, J., Larson, J., Vendrell, J., Hershberger, C.L., Avilés, F.X. (1999). Mapping the Pro-region of Carboxypeptidase B by Protein Engineering: Cloning, Overexpression, and mutagenesis of the porcine proenzyme. Journal of Biological Chemistry, 274(28), 19925-19933. DOI:10.1074/jbc.274.28.19925

Figures and Data

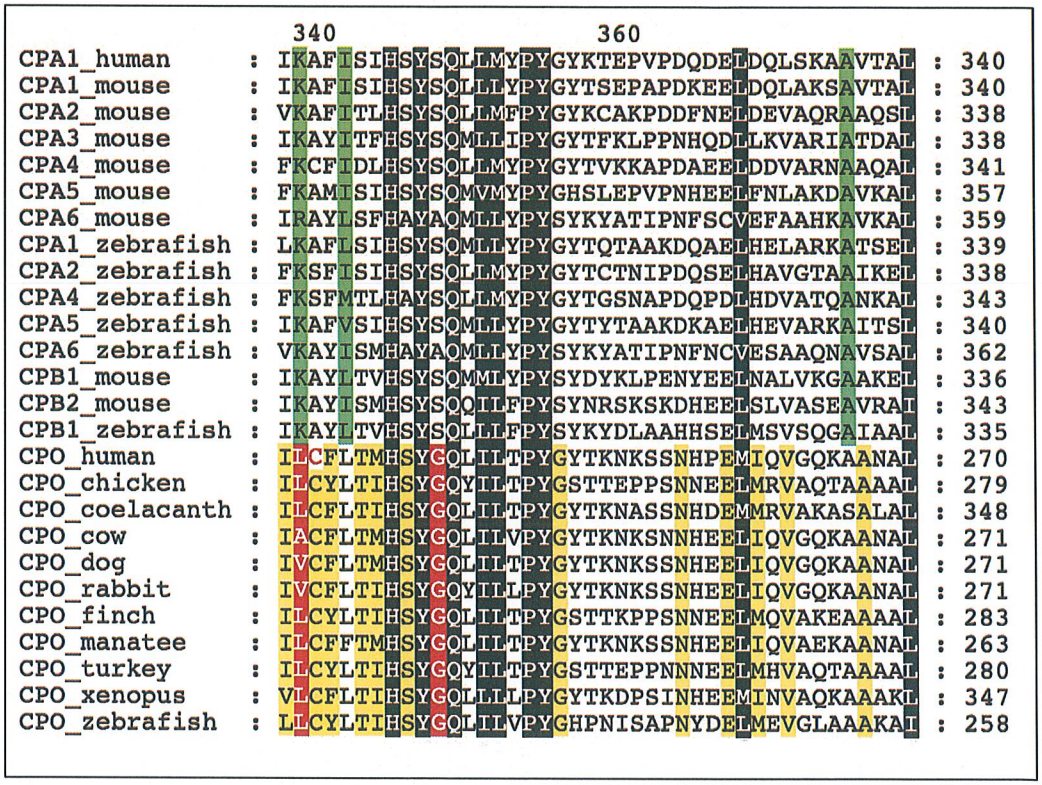

Figure 1. CPO and CPA/CPB gene alignment. We aligned CPO with CPA and CPB to identify unique, conserved amino acids in CPO. Green highlighting indicates conservation within the CPA/CPB group, while yellow highlighting indicates conservation within the CPO group. The red highlighting in this figure occurs at amino acids 230 and 239 in human CPO, and represents two of the four sites for mutation. All four mutation sites were selected because of contrasting conservation in CPO and CPA/CPB

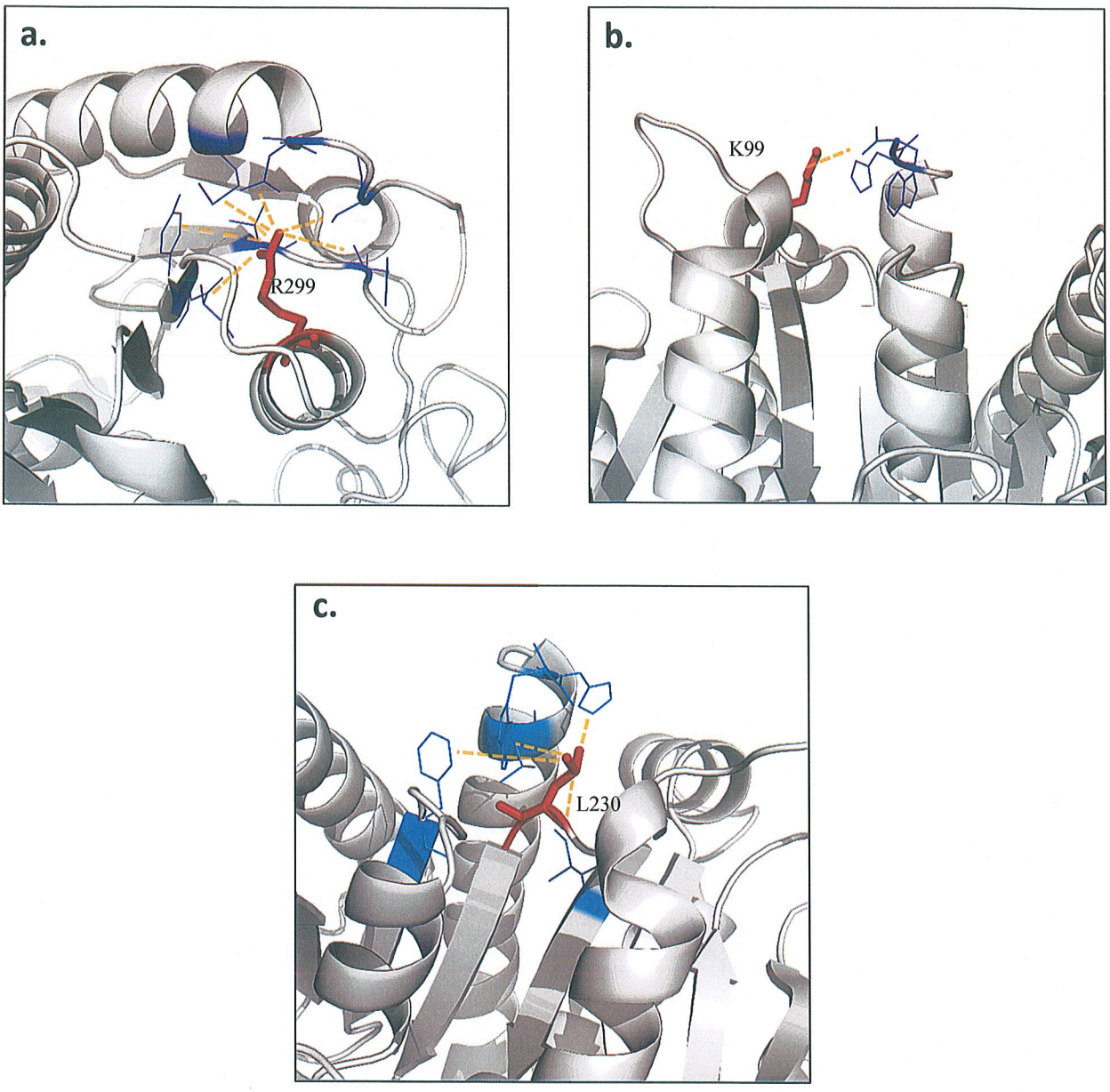

Figure 2. Three dimensional modeling in PyMol suggests possible interactions between L230, G239, R299 and neighboring amino acids. These figures represent proposed stabilizing interactions between these three amino acid side chains and other residues. A. shows R299 interacting with a polar pocket of amino acid side chains 244T, 246Y, 253S, 255N, 259M, 306S, 307Y). B. shows K99 stabilizing the carboxyl group of the C-terminus. C. shows L230 stabilizing a hydrophobic pocket of amino acid side chains (349W, 348H, 344V, 305F, 101I).

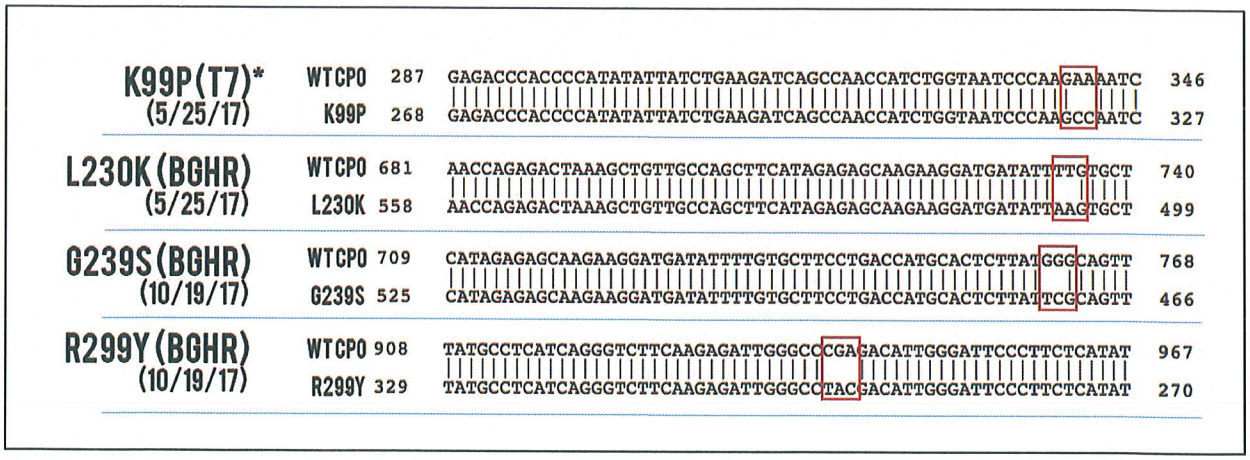

Figure 3. Confirming the K99P, L230K, G239S, and R299Y mutations. We sent in our plasmids for sequencing to verify their identity, and compared this data with the data of wild-type CPO. The targeted amino acid codons are boxed above. T7 and BGHR describes the forward and reverse primers used to initiate sequencing. The BGHR primed sequences may be read forward, but K99P, which used a T7 primer here, must be read in reverse. We observed that all mutations were successful, and we did not observe any other undesired mutations.

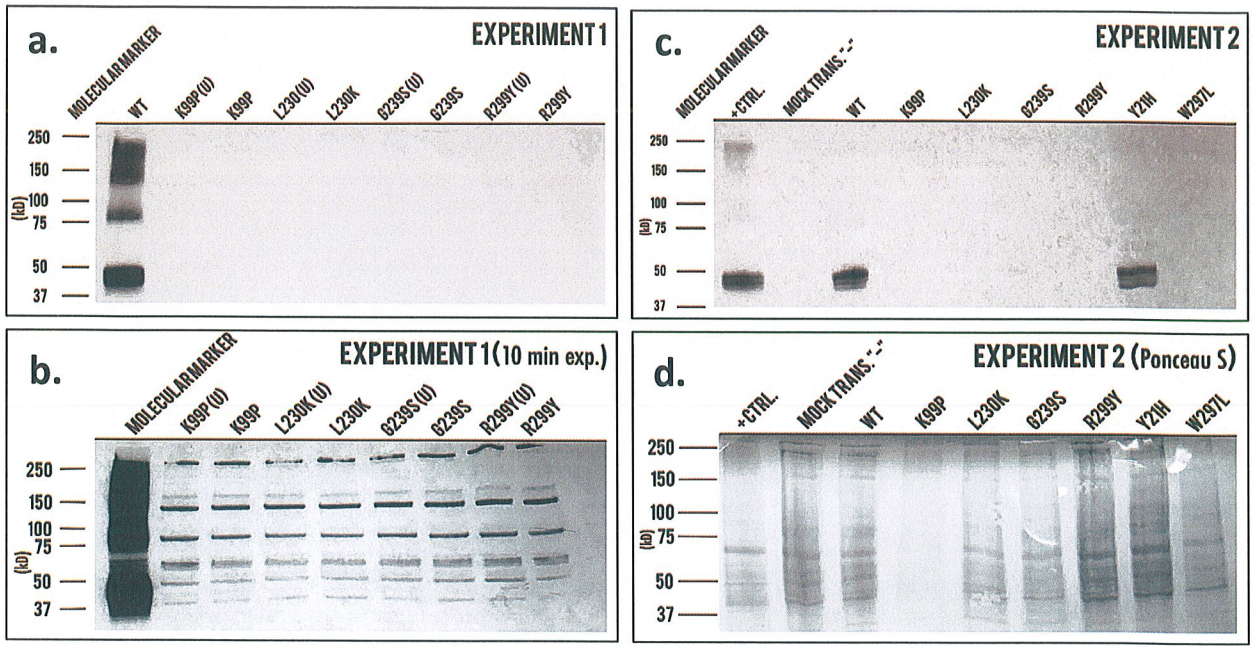

Figure 4.* Mutating K99, L230, G239, and R299 to their CPA/CPB analogues prevents expression of CPO, hinting at the significance of these sites in folding stabilization. HEK293T cells were transfected with plasmids expressing wild type CPO (WT and +Ctrl), various CPO mutations, and no CPO (Mock Trans. "-"). Two additional mutations (Y21H, W297L) were included in experiment 2 as controls. Extracts from the above cells were analyzed by western blotting. a) In experiment 1 there was no expression of CPO from any of the four mutations. b) However a long exposure showed the presence of protein in all lanes. c) Experiment 2 also showed no expression of CPO for all four mutations. d) A Ponceau S stain revealed a decreased amount of protein extracted from cells carrying the K99P plasmid. The same was true to a lesser degree for the L230K and G239S mutations.

* The U's after some mutations in Experiment 1 signify "unmixed," and refer to a failure to mix these cells thoroughly before plating during transfection.

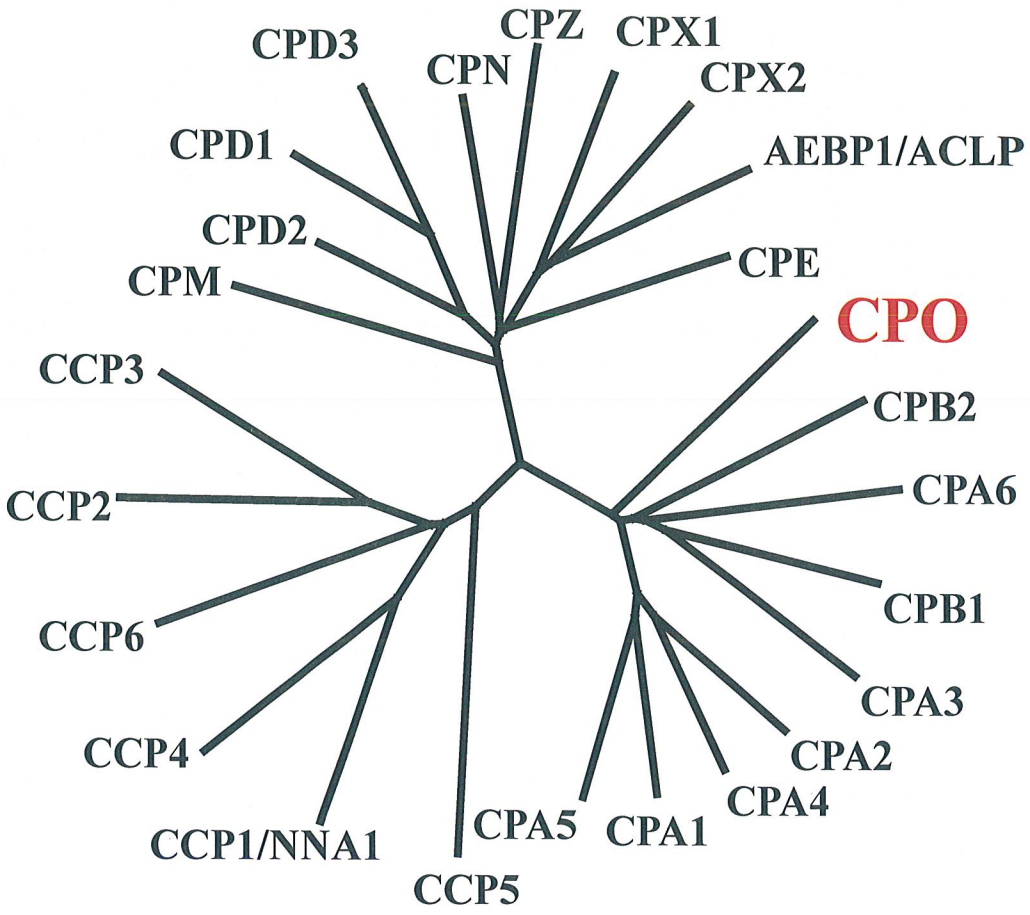

Figure 5. The Carboxypeptidase Tree. CPO belongs to the A/B family of carboxypeptidases. There are eight other enzymes very closely related to CPO's. All of them are CPA or CPB, and all require a prodomain to fold properly. The other more distantly related carboxypeptidases shown here can also fold without a prodomain, but they are not digestive enzymes.