

Modeling of the mechanisms for post-translational D-amino acid incorporation in dermorphin and contryphan-R

Michael Hedrick R. Cabotage¹, Elsie C. Jimenez² and Neil Andrew D. Bascos^{1,3,*}

¹Protein Structure and Immunology Laboratory, National Institute of Molecular Biology and Biotechnology, University of the Philippines Diliman, Quezon City, 1101;

²Department of Physical Sciences, College of Science, University of the Philippines Baguio, Baguio City, 2600;

³Protein, Proteomics and Metabolomics Facility, Philippine Genome Center, University of the Philippines System, Diliman, Quezon City, 1101, Philippines.

ABSTRACT

Most peptides found in nature possess homochirality, that is, they contain almost exclusively L-amino acids. Nonetheless, D-amino acids have been discovered in peptides from various organisms. This study investigated the modes of post-translational D-amino acid incorporation in the ribosomal peptides, dermorphin and contryphan-R, through molecular modeling and molecular docking simulation techniques. The structures and interactions of the propeptides with appropriate isomerizing enzymes (racemases) were explored. Depending on the position and interaction with the bound prodermorphin, the dimeric alanine racemase adopted either a closed or an open conformation. On the other hand, regardless of the position of the procontryphan-R, a tryptophan-targeting racemase with broad substrate specificity assumed only the closed conformation. Analysis of the total energies from the different interaction types involved in the enzyme-substrate models revealed that the total energy present in the dimeric configuration of the racemase was reduced in the presence of the propeptide. A complex combination of these interactions occurred at specific racemase conformations, which suggest possible energy tradeoffs associated with the

association of the racemase and the propeptide. These conformational changes may represent different mechanisms through which different racemases may act on specific substrates. Analysis of the structural bases for the D-amino acid incorporation provides insights into the rare occurrence of these non-proteinogenic amino acids in peptides. The post-translational L- to D-amino acid isomerization process may be further explored for the functional benefits that it offers, including its application for the development of novel peptide-based therapies.

KEYWORDS: ribosomal peptide, D-amino acid, L- to D-amino acid isomerization, dermorphin, contryphan, racemase.

INTRODUCTION

Peptides synthesized *via* the ribosomal machinery (referred to as ribosomal peptides) are typically homochiral. Nevertheless, several review articles have documented the rare occurrence of D-amino acids (D-AAs) in peptides from widely diverse multicellular organisms [1-3]. In these rare cases, ribosomal peptides are synthesized using L-amino acids (L-AAs). The D-AAs are by some means incorporated at specific positions within the peptide sequence. Ribosomally translated peptide is initially in the form of propeptide that consists of the signal sequence, pro-region and mature peptide [4]. Proteolytic cleavage generates

*Corresponding author
neilandrew.bascos@mbb.upd.edu.ph

the propeptide (pro-region and mature peptide) that is the mature peptide precursor.

The cDNA cloning technique confirmed that a normal codon for L-AA is found at the position where D-AA occurs upon peptide maturation [5-8]. The incorporation of D-AAs as peptide components increases the variety of peptide sequences and their corresponding functions. The presence of D-AA has been documented to result in greater structural stability and/or enhanced biological activity [9-14].

A serine isomerase found in the venom of the spider, *Agelenopsis aperta*, provided the first evidence that defined the D-AA incorporation into the peptide of a multicellular organism [15]. The L- to D-AA isomerization occurred in the crustacean's neurosecretory organ to produce a mixture of L and D isoforms after propeptide cleavage. This process is a late step in the maturation of the crustacean hyperglycemic hormone (CHH) [16]. A few other studies have focused on the bases for the stereoselectivity of peptide generation.

This project investigated the potential bases for D-AA incorporation, in the light of this observed stereoselectivity for L-AAs. As study cases, we focused on the documented D-AA containing peptides, and the interactions of the propeptides with racemases. The racemases were chosen based on their ability to isomerize specific AA types to generate the experimentally validated D-AA-containing sequences for the peptides, dermorphin and contryphan-R.

Dermorphin is a heptapeptide characterized from the Argentinian frog, *Phyllomedusa sauvagei*. It possesses opiate activity as shown in bioassays using guinea pig ileum test, as well as rat tail-flick and hot-plate tests [17]. The mature peptide has the sequence of YAFGYPS, where A is D-alanine. The translated peptide sequence contains five propeptide sequences following a single signal sequence [5].

Contryphan-R is an octapeptide identified in the venom of the predatory cone snail, *Conus radiatus*. It causes "stiff tail syndrome" and other excitatory activities as demonstrated in fish bioassay. The peptide sequence is GCOWEPWC-NH₂, where O is 4-hydroxyproline and W is

D-tryptophan [18]. It is remarkable that while the tryptophan residues in contryphan-R are found in two locations (i.e., Trp4 and Trp7), only Trp4 was modified to the D isomer.

The codons for L-Ala and L-Trp are found at the positions that correspond to those occupied by D-Ala and D-Trp, respectively, in mature dermorphin and contryphan-R [5, 19]. These findings support a post-translational mode for the observed L- to D-AA isomerization. The present study investigated the potential mechanisms through which racemases might introduce D-AAs at specific locations. Specifically, molecular modeling and molecular docking were carried out to determine the structural bases for the post-translational D-AA incorporation into these peptides.

MATERIALS AND METHODS

Structure preparation and homology modeling

A crystal structure of a homodimeric alanine racemase (AlaR) bound to an alanine phosphonate from *Geobacillus stearothermophilus* was obtained from the Protein Data Bank (PDBID: 1BD0) [20]. All hetero atoms (e.g. H₂O, ions and phosphonic acid ligand) were removed and deleted from the PDB file, as part of the optimization process for the docking simulations.

Homology modeling allows the prediction of structures for the UniProt deposited sequences based on curated structures of related proteins. Hence, the protein sequences of the following molecules were submitted to the I-TASSER server (zhanglab.ccmb.med.umich.edu/I-TASSER) [21-23] for modeling: 1) the propeptide sequence of dermorphin (referred to as prodermorphin) from *Phyllomedusa sauvagei* (UniProt accession ID P05421), 2) the propeptide sequence of contryphan-R (referred to as procontryphan-R) from *Conus radiatus* (UniProt accession ID P58786), and 3) the broad substrate specificity Trp-targeting racemase from *Pyrococcus horikoshii*, subsequently referred to as tryptophan racemase (TrpR; UniProt accession ID O57878). The models with the highest C-scores were used for downstream *in silico* analyses. Swiss PDB Viewer [24] was used to generate extended linear structures of procontryphan-R, where the input was the propeptide sequence, and the output structure was saved as a PDB file.

Molecular docking and screening of multistep docked models

A multistep docking protocol was employed to produce static models of racemase and propeptide interactions. The AlaR targets Ala in dermorphin, while the TrpR putatively targets Trp in contryphan-R. The target propeptide sequence was initially docked into a racemase monomer using the ClusPro 2.0 server (cluspro.bu.edu/) [25-27] to determine possible modes of binding. These models were screened for docks that showed potential interactions between the catalytic Lys residue of the racemase and the target residues in prodermorphin and procontryphan-R. The racemase monomer was designated as the receptor, and the propeptide chain was used as the ligand. Among the returned models, the distances between the catalytic residues (i.e., AlaR Lys39, TrpR Lys295) and their target residues (i.e., dermorphin Ala2, and contryphan-R Trp4) were measured using an automated script. Screening for relevant docking conformations was based on the distances between these residues as calculated using their respective x, y and z coordinates and the distance formula,

$$D = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}.$$

The docked models were classified into three groups according to the position of the propeptide with respect to the racemase monomer (chain A): 1) propeptides that interacted with the non-catalytic surface, 2) propeptides whose pro-region interacted with the catalytic Lys residue, and 3) propeptides whose mature peptide region interacted with the catalytic Lys residue. These classifications were also related to the initial distances between the target D-AA and the catalytic residue. Structures where the propeptide chain interacted with the non-catalytic racemase surface were not included in the succeeding analyses. Models where the pro-region or mature peptide region interacted with the catalytic residue were submitted to the server for another round of docking to generate the dimeric forms (chains A and B), documented to be required for function [28].

Control docked models in the absence of the propeptides were similarly prepared by submitting

the racemase structures to the ClusPro 2.0 server. For these control docks, one monomeric subunit of racemase (Chain A) was considered as the receptor, and the other monomer subunit (chain B) was used as the ligand. Similarly, models returned by the server were initially screened based on the orientation of chain B with respect to chain A. Models consistent with the functional descriptions as described earlier [28] were selected for use in the succeeding analyses.

In addition to monomer subunit pairing, the orientations of the docked subunits were also screened for consistency with expected functional forms. This screening evaluated the distances between the sets of complementary catalytic residues and two arbitrary residues in the distal ends of the racemase, referred to as D_1 , D_2 and D_3 , respectively. These distances were measured using an in-house script written in Bash [29]. This was carried out for both control and multistep docked models. The selection of qualified models was done based on proposed mathematical assumptions, consistent with the known functional conformations of racemases [28]. These may be represented based on distance measurements between key positions in the interacting subunits and their ligand (Figure 1). D_1 and D_2 measure the distance from the catalytic residues of monomer 1 to monomer 2. D_2 measures the distance from monomer 2 to monomer 1. D_3 measures the separation of the two monomers at the distal ends of the enzyme, or the opposite position from the hinge region of AlaR. Values for D_1 , D_2 , and D_3 were taken either in the absence or presence of the target peptide. Patterns on the distances obtained from the control docked models were normalized and applied to the following formulas used for the selection of qualified models: $D_1 \vee D_2 \leq (\overline{D_{D1 \text{ or } D2 \text{ control}}} + 2 \overline{d_{\text{mature}}})$ and $D_1 \vee D_2 < D_3 < (\overline{d_{D3 \text{ control}}} + 4 \overline{d_{\text{mature}}})$. Docks that satisfied these conditions were selected for succeeding analyses.

Selected docks for AlaR had D_1 and D_2 values that were below 20.82Å, and D_3 values between 10.25Å and 48.52 Å. For TrpR, selected docks had D_1 and D_2 values that were below 34.59Å, and D_3 values between 18.06Å and 84.43 Å. Selected models for AlaR had $D_1 > D_2$. The inverse was observed for TrpR, where $D_1 < D_2$.

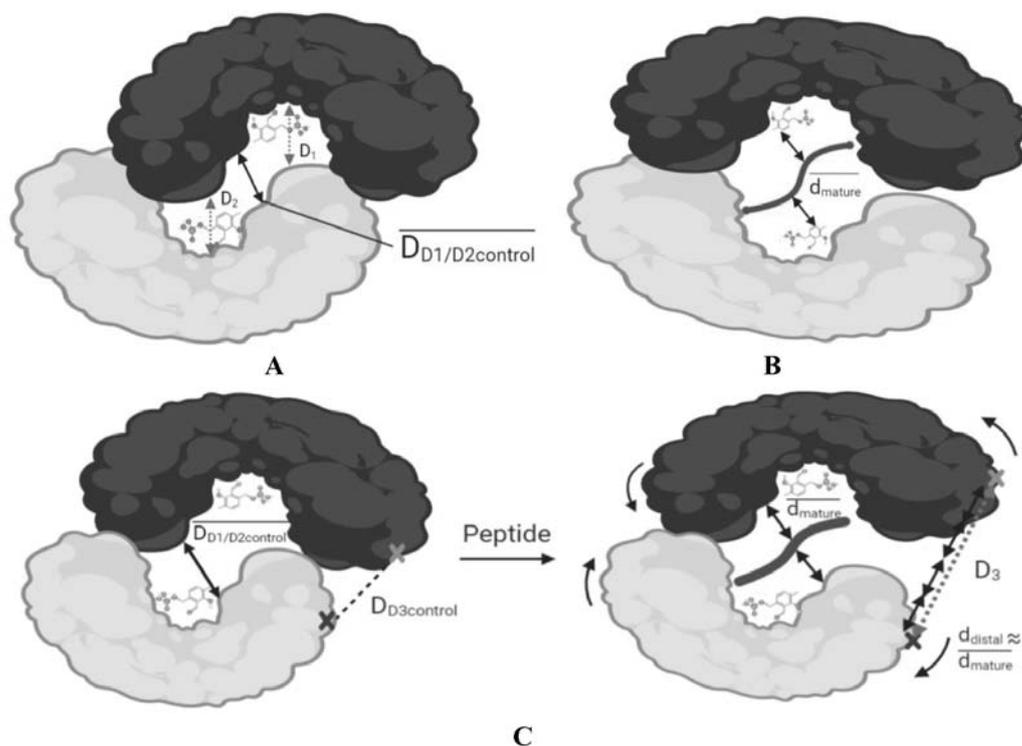


Figure 1. A model representation of the criteria for model selection. An arbitrary space between enzyme subunits was modelled to allow for propeptide interaction and movement. This space was based on the mean distances $\overline{D_{D1 \text{ or } D2 \text{ control}}}$ and $\overline{d_{\text{mature}}}$ (A, B) between the propeptide and the two racemase monomers. (C) shows the possible adjustments (*via* arrows) in the enzyme upon insertion of the propeptide. D_3 represents the distance between residues in the distal ends of each monomer subunit, marked with an “X”. This figure was created using BioRender (biorender.com) and is not drawn to scale.

The opposite trends for the $D_1:D_2$ ratios in the two racemases are consistent with a difference in the modes of subunit association for the two enzymes.

The range of the distances among these qualified models was checked if they followed a normal distribution. This ensured that the selected models not only followed the aforementioned assumptions but were also distributed around the data set’s mean and standard deviation. The models were then compared accordingly between setups (racemase – mature peptide region vs. racemase – pro-region interactions) using appropriate parametric or non-parametric statistical tests at $p < 0.05$ significance.

Ligand docking and quantification of chemical interactions

AlaR and TrpR are pyridoxal phosphate (PLP)-dependent enzymes, wherein PLP acts as a

cofactor that facilitates the racemization of the target amino acid. This is bound to the catalytic lysine residue in the enzyme’s active site. Models with the required PLP cofactor were generated to accurately predict the functional interactions of racemase and its target propeptide. A three-dimensional conformation structure of PLP was obtained from PubChem [30]. The ligand was prepared by assigning charges, defining rotatable bonds, renaming aromatic carbons, merging nonpolar hydrogens, and converting the PDB file to a PDBQT ligand file using AutoDock Tools [31]. This system was also used to prepare the racemase chain A receptor, deleting water molecules, adding polar hydrogens, as well as defining the expected receptor binding site. A configuration file was then prepared based on the coordinates of the expected site, and the docking was run using AutoDock Vina [32]. Docked models were

visualized and analyzed using the PyMOL molecular viewer [33].

The propeptide chain among the PDB structures of qualified models was renamed (e.g., Chain C) to distinguish it from the rest of the molecules in further analyses. The specific interchain interactions within these updated PDB structures were quantified using the RING 2.0 server (protein.bio.unipd.it/ring/) [34, 35]. Among control models, the interactions between racemase chains A and B were considered. Among experimental docked models, the interactions between chains A and C, and those between chains B and C were analyzed. The total energy of the system was calculated by adding the individual energies of the different interaction types (i.e., Hydrogen bonds, salt bridges or ionic bonds, van der Waals interactions, π - π stacking interactions, π -cation interactions, and disulfide linkages). Statistical significance was tested for differences observed between control and experimental models at $p < 0.05$.

RESULTS

Analysis of L- to D-amino acid incorporation in prodermorphin and procontryphan-R

Multiple sequence alignments of available prodermorphin and procontryphan sequences in the UniProt server were done using MultAlin [36] to check for conserved patterns that may guide the L- to D-AA isomerization. There are four available prodermorphin sequences from three species, two of which are from *Phyllomedusa sauvagei*. One sequence from *Phyllomedusa sauvagei* has five mature dermorphins, while the others have only four. Both signal sequences and pro-regions of procontryphans are generally

conserved among the 27 available sequences from various species. The conserved sequences in prodermorphins and procontryphans were used as target sequences for racemase binding.

Although several sequences for the target propeptides are available in UniProt [37] this work focused on sequences based on their applicability for analyzing specific structural features. For prodermorphin [17], the sequence that contains five identical mature peptide sequences following the signal sequence [5] was selected for downstream structural analyses. Through this, the potential significance of tandem repetition of a single propeptide sequence was investigated.

Contryphan-R was the first contryphan to be characterized [18] and was found to share 100% sequence identity with the contryphans later identified in different *Conus* species [38-40]. Unlike prodermorphin, procontryphan-R sequence has a single mature peptide sequence following the pro-region [19]. Both prodermorphin and procontryphan-R sequences (Figure 2) were analyzed in terms of their structures and interactions with racemases to elucidate possible mechanisms of L- to D-AA isomerization.

Molecular docking experiments were done for the racemases that target the prodermorphin and procontryphan-R. Models were classified according to the propeptide's position with respect to the racemase catalytic residues.

The first round of molecular docking generated 17 models for prodermorphin and 20 models for procontryphan-R. All prodermorphin sequences docked in the pocket containing the AlaR catalytic residues, with five models interacting with the target mature peptide sequence and 12 interacting



Figure 2. Propeptide sequences: (A) prodermorphin [5]; (B) procontryphan-R [19]. Boxed sequences were modeled and used in the analyses.

with the pro-region. Fourteen out of 20 procontryphan-R models docked within the TrpR pocket. Eight models interacted with the mature peptide region and six models interacted with the pro-region. Six procontryphan sequences that docked to the TrpR racemase surface, away from the catalytic sites, were excluded in further analyses.

The second round of docking generated a maximum of 30 docked models per initial model from the first round, producing 930 dimeric models. Each of these docked models was then assessed based on the distances between the complementary catalytic residues, as well as two different arbitrary residues in the distal ends of each monomeric subunit (Figure 3). The ratios between the distances of complementary catalytic residues were measured for each of the models. These measurements gave a quantitative basis for the identification of the modes of association and for the screening of the multistep docked models for downstream analyses (Figure 1).

The 930 docked models presented numerous binding modes for the racemase chain B; 854/930 (91.83%) docks placed the chain B on the surface of the chain A in different orientations. A subset of the generated docks, 76/930 (8.17%), resulted in a clamping structure that placed the catalytic residue sites of chains A and B at complementary positions. These structures were seen for 20/510 (3.92%) models in prodermorphin, and 56/420 (13.33%) models in procontryphan-R. Based on the documented dimeric nature of functional

racemases, succeeding analyses were focused on these dimeric orientations that allowed both catalytic residue surfaces to function.

Control docks were carried out in the absence of propeptide, which were aligned and compared with the racemase crystal structures, if available. These control docks served as references for the selection of models that were similar to the experimentally observed crystal structures, which may represent functional forms. All 76 models of prodermorphin and procontryphan-R interacting with their respective racemases were processed using PyMOL. Figure 4 shows an example of the clamped structures achieved through the docking simulations. The prodermorphin-AlaR complex in Figure 4C (top) had a root mean square deviation (RMSD) of 1.978 when compared to the control docked model. This model obtained the lowest RMSD (1.493) from its reference crystal structure. Meanwhile, the procontryphan-R – TrpR complex in Figure 4C (bottom) had an RMSD of 3.389 from its control dock model, which was also the lowest value among docked structures.

Only the 76 docked models that satisfied the proposed assumptions were considered for further analysis. The quantification of the intermolecular distances in these docked models allowed a more systematic approach in narrowing down the generated docked models that would permit coordinated function based on the expected distances between key residues.

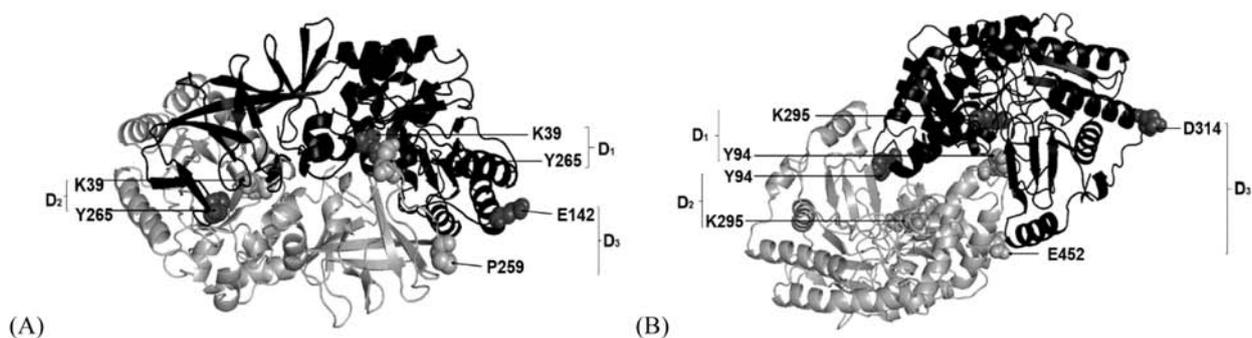


Figure 3. Positions of the residues to be measured in the screening process. The corresponding residues to be measured are in space filling model, while the remaining residues are in ribbon diagram. Chain A is colored black, and chain B is colored gray. The figure shows docked structures for (A) Ala R and (B) TrpR that satisfied the conditions for inclusion in the succeeding analyses.

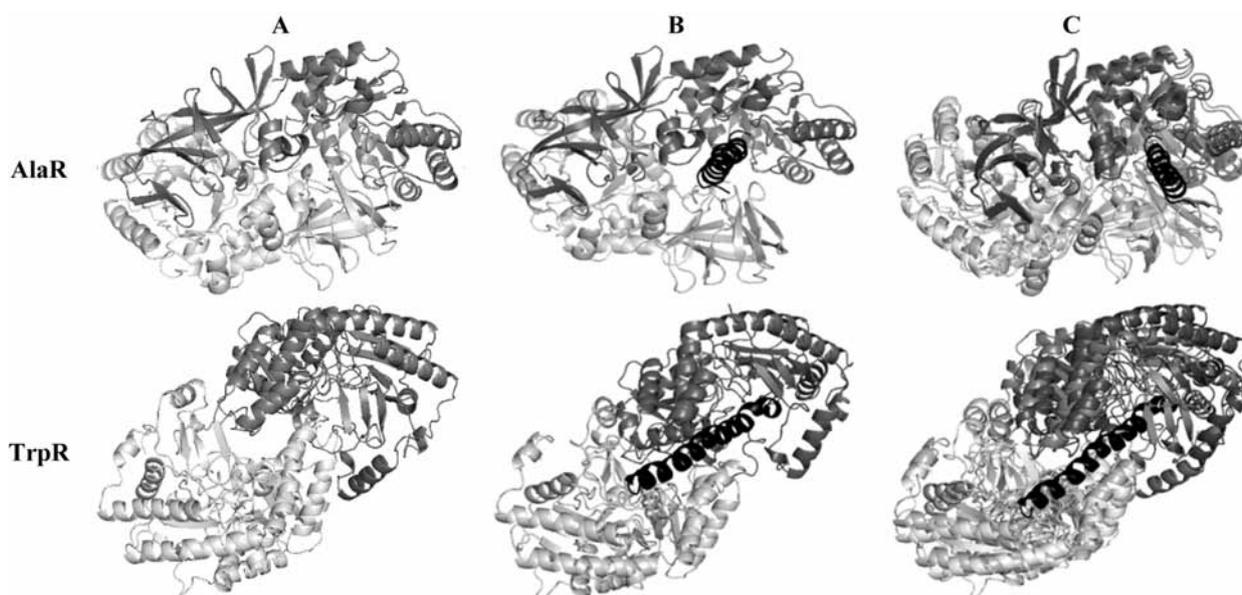


Figure 4. Modeled interactions based on the complementary orientations of the catalytic residues: (A) in the absence of peptide; (B) multiple docking of racemase subunits (chains A and B) and the peptide; (C) superimposed models of A and B. The two racemase monomers are colored in different shades of gray, while the propeptide is colored in black.

Analysis of the multistep docked structures from the prodermorphin-AlaR complex

When the target residue (Ala29 of prodermorphin, Ala2 of dermorphin) in the substrate interacted with the catalytic Lys39 of AlaR, the dimeric racemase complex adopted a closed conformation (Figure 5A). This closed orientation involved the β -sheet region of chain B orienting itself upward towards the TIM barrel region of chain A, whose base has AlaR's catalytic residues, such that it inserted the prodermorphin between chains A and B. This configuration was observed for all 10 qualified models that adopted a closed conformation. On the other hand, when the pro-region interacted with the same AlaR's catalytic Lys39, the dimeric racemase complex adopted an open conformation for all 10 qualified models. In the open conformation, the β -sheet region of chain B is shifted slightly away from the TIM barrel region of chain A, and the α -helices of chain B are shifted towards the β -sheet region of chain A (Figure 5B).

The AlaR closed and open conformations coincide with lower values of D_1 in models that interacted with the mature peptide (i.e., closed conformation) compared to higher D_1 values in

models that interacted with the pro-region (i.e., open conformation). The ranges of D_1 values in the closed and open conformations were 7.12Å–16.60Å and 14.61Å–18.80Å, respectively. While there is a small overlap in values for these sets (14.61Å–16.60Å), these only occurred in 1 of 10 models per set and are considered the extremes of the range. Both observed conformations are consistent with the dimeric conformational structures of AlaR, as described earlier [28], as well as with the observed intramolecular distances normalized from the crystal structure [20]. The difference in the range of measurements between the closed and open forms of D_1 are significant ($p = 0.0002$). In contrast, D_2 between the closed and open conformations do not present significant differences ($p = 0.9657$), suggesting that this set of catalytic residues may still act as an interdomain hinge region (IHR) that binds chains A and B together despite the occurrence of two distinct conformations. Additionally, a significant difference was observed between D_1 and D_2 ($p < 0.0001$). As such, overlapping the upper limit of the former and the lower limit of the latter gives an average of 15.61Å, the D_1 value at which the transition between a closed and an open

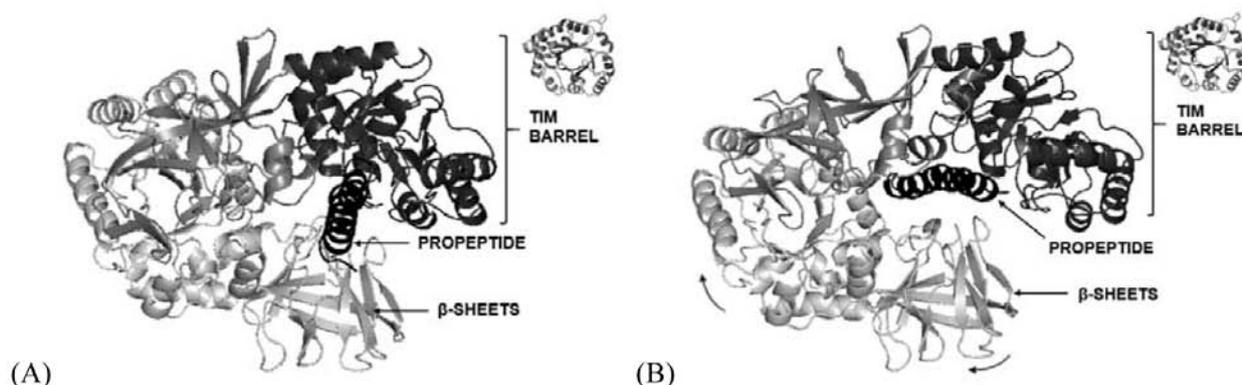


Figure 5. The conformation of the dimeric AlaR depending on the region of interaction between prodermorphin and AlaR. (A) shows the closed conformation when the mature peptide region interacts near the catalytic residue, while (B) shows the open conformation when it is the pro-region that interacts with the said residues. The individual monomers of AlaR are colored gray, whereas the TIM barrel is colored in dark gray. The propeptide is colored black.

conformation of AlaR presumably occurs. Furthermore, the D_3 values that represent the distal regions of AlaR provide a more stringent criterion for model screening, and these show a significant difference between the closed and open conformations ($p = 0.0023$).

The observation that two possible conformations can exist within AlaR provides a potential mechanism for how prodermorphin and AlaR interactions control the racemase activity. Aside from the largescale conformational changes, it was observed that when the complex was viewed from the perspective of the TIM barrel, the models whose mature peptide region directly interacted with the catalytic residues oriented the prodermorphin parallel to the barrel opening (Figure 6A). In contrast, in models where the catalytic residues interacted with the pro-region, the prodermorphin was misaligned from the barrel opening (Figure 6B). Such difference in orientation of prodermorphin could indicate its possible movement and guidance by AlaR. It is possible that the opening of AlaR facilitates prodermorphin misalignment from the barrel opening, decreasing its probability of interaction. The reverse is true for the mature peptide region, whose orientation with the closed conformation facilitates its targeting by the catalytic residue (Figure 6C).

In addition to single prodermorphin-AlaR interactions, the current study investigated the possible effect of tandem propeptides in the target sequence.

To minimize the computational requirements, a “multiple target analysis” was initially set with two-tandem propeptides, following the same multiple docking protocol. Three docked models that showed a properly oriented dimer conformation after the initial multistep docking were then subjected to another round of docking simulations, where the prodermorphin-AlaR dimer complex acted as the receptor, and another monomer of AlaR (referred to as monomer 3) acted as the ligand. These 3 models shared a conformation that is similar to those depicted in Figures 4B and 5A; hence, they were chosen for further analyses. From this round of simulation, instead of docking to the other target residue in the peptide, the AlaR’s monomer docked to the surface of the peptide-clamped dimer (Figure 7A) in 90 models. One of the possible reasons for this dock is the lack of space between AlaR monomers in the two-tandem propeptides. Hence, a three-tandem propeptide was docked with the AlaR’s chains A and B (Figure 7B), but this dock similarly did not target the other prodermorphin sequences among 30 models.

Analysis of the multistep docked structures from the procontryphan-R – TrpR complex

TrpR was observed to adopt a closed conformation for 56 docks, where the substrate interacted with the catalytic Lys295, irrespective of the procontryphan-R position (Figure 8). The closed

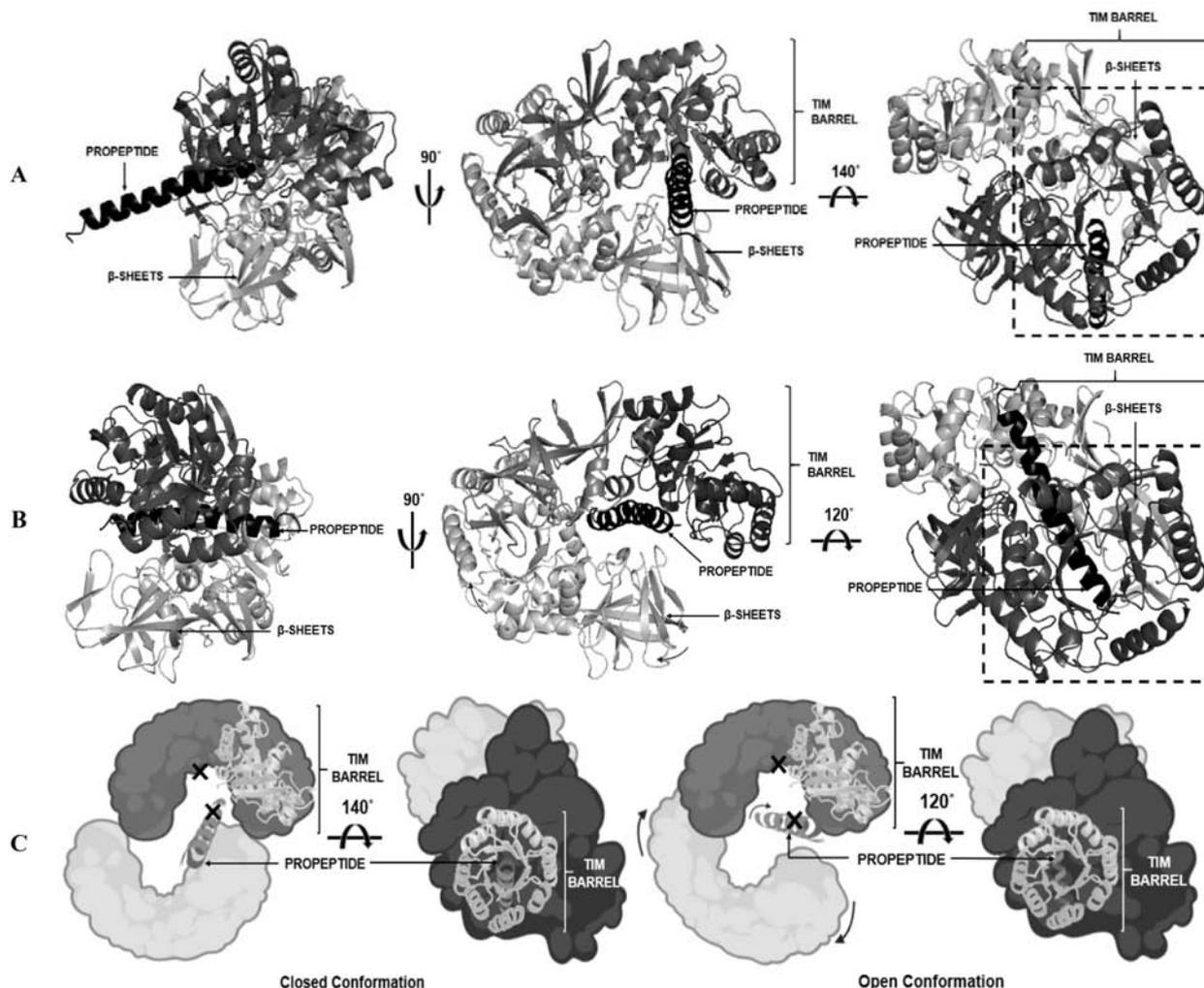


Figure 6. The different orientations of the multistep docked models. (A) shows the closed conformation viewed from the TIM barrel (rotated 140° with respect to [w.r.t.] the x-axis) and the side (rotated 90° w.r.t. the y-axis). (B) shows the open conformation viewed from the TIM barrel (rotated 120° w.r.t. the x-axis) and the side (rotated 90° w.r.t. the y-axis). The two enzyme subunits are colored in different shades of gray, and the prodermorphin is colored black. (C) shows a cartoon representation of the frontal and rotated views of the propeptide misalignment from the TIM barrel region when racemase changes its conformation. Marked in 'X' are the catalytic and target residues. Figure 6C was created using Bio Render (biorender.com).

conformation was described as having the enzymatic small domain containing 3-4 α -helices inserted into the pocket of the complementary subunit.

The observation of the closed conformations in TrpR coincides with the observed values of D_1 and D_2 . In three distance parameters, there was no significant difference between the measurements using a paired t -test. The differences in the ranges of D_1 and D_2 distances in either procontryphan-R

positions are negligible. Although D_3 was greater than both D_1 and D_2 , all models adopted a closed conformation. The paired t -test of the values of D_1 , D_2 and D_3 between the pro-region and mature peptide interactions in procontryphan-R-TrpR complex showed that at $p < 0.05$, D_1 , D_2 and D_3 values are not significantly different with p -values of 0.158337, 0.910107 and 0.21058, respectively, where $n = 35$ for docks with mature peptide region and $n = 21$ for docks with pro-region.

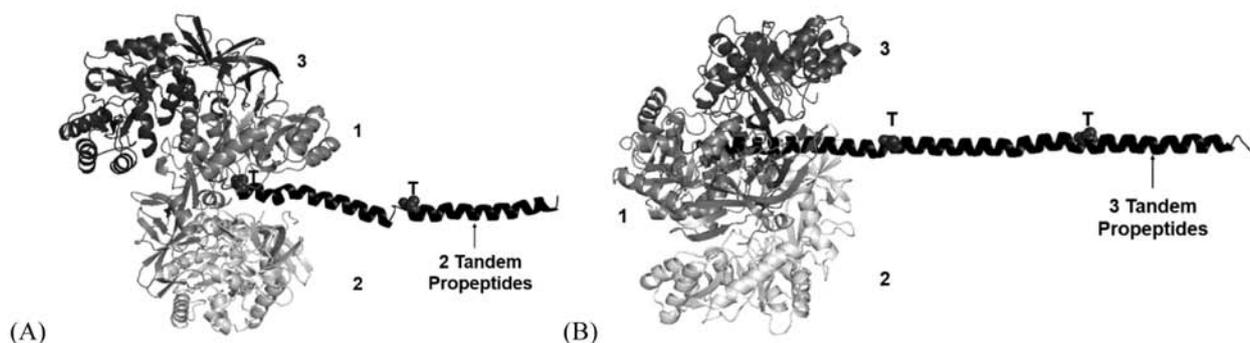


Figure 7. Multiple prodermorphin target analysis of the AlaR conformation. The third monomer (3) docks on the surface of chain A or the first monomer (1) instead of the other target residue in the peptide (T, space filling model). Number 2 is chain B or the second monomer that forms a dimeric conformation with (1). (A) docks in two-tandem propeptides; (B) docks in three-tandem propeptides.

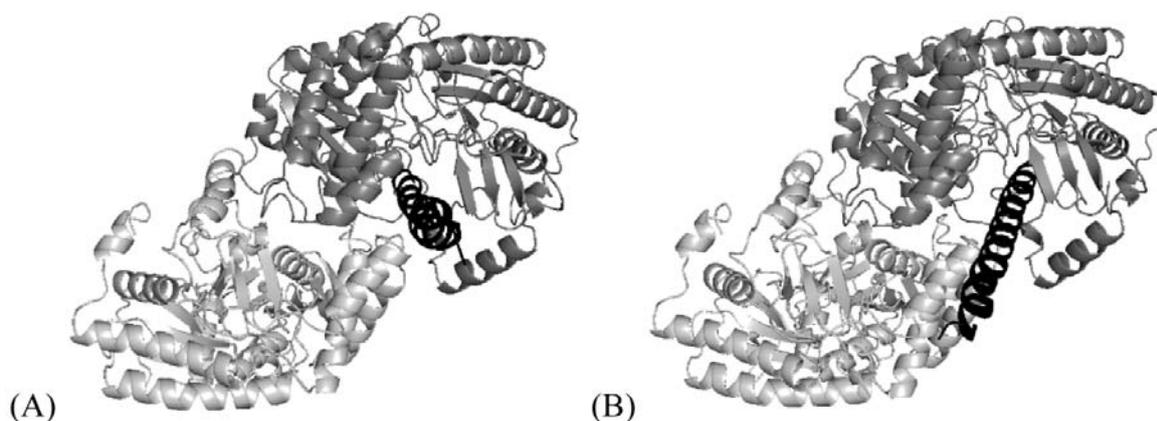


Figure 8. Multiple procontryphan-R target analysis of the TrpR conformation. A closed conformation was observed in the docked models when procontryphan-R was interacting with the catalytic Lys295 residue. (A) mature peptide region interacting with Lys295; (B) pro-region interacting with Lys295.

Unlike in AlaR where largescale closed-open conformational transitions were observed, TrpR adopted only a closed conformation. The initial distances between the pro-region and the catalytic Lys residue in TrpR may possibly explain the difference in the adoption of conformations between AlaR and TrpR. In AlaR, the mean initial distance between the target residue and the catalytic lysine residue was 5.4411Å, while in TrpR, the mean initial distance was 9.3998Å. The distance between the Lys residue and the catalytic site N2 in the PLP co-factor (obtained through AutoDock) is 3.60Å for both AlaR and TrpR. Subtracting the distance between the Lys residue and the PLP from the initial distance gives catalytic distances of 1.8411Å and 5.7998Å,

respectively, for the residues in AlaR and TrpR. Accounting for the other subunit of the dimeric racemase yields catalytic distances of 3.6822Å and 11.5996Å. Both of these distances are adequate for catalytic interaction between PLP and peptide substrate to proceed [39]. However, since the distance in TrpR is significantly larger, this space probably allows procontryphan-R interaction and movement between the TrpR chains A and B without having to adopt an open conformation.

Energies of interactions

The total energies of all interaction types among the qualified models in AlaR and TrpR structures in the presence or absence of the target prodermorphin

and procontryphan-R were obtained. The data were first checked for normality using the Shapiro-Wilk test before applying the corresponding statistical analyses at $p < 0.05$ significance. Parametric t -tests were used to determine significance for the differences in total energies between normally distributed measurements of mature peptide and pro-regions of interactions. Non-parametric t -tests were used for those that do not follow normal distribution. One-sample t -tests were used to compare the crystal structure of AlaR with the control models. Lastly, one-way analysis of variance (ANOVA) was carried out to compare between control and experimental docks.

In both racemases, there are significant differences in the total energies between control (i.e., no peptide) and experimental (i.e., peptide bound) models. The presence of the target propeptide significantly reduced the total interactions and energy present between chains A and B of the racemases ($p < 0.0001$). Between the crystal structure of AlaR and the control models, a significant difference was observed ($p = 0.0206$). The difference is due to hydrogen bonding

($p = 0.0302$) and van der Waals interactions ($p = 0.0076$). Between mature peptide region and pro-region interaction with the AlaR, there are no significant differences in total energy ($p = 0.1237$). The inverse is true between the same regions among TrpRs ($p = 0.0466$).

Ala-R-prodermorphin interaction energies

For AlaR, one aspect of the observed decrease in total interchain energy may be correlated with a reduced inter-subunit binding at one of the IHRs in the presence of the target prodermorphin. While the IHR is not directly involved in substrate binding [41], its dissociation to accommodate the propeptide binding may alter the interactions between the racemase subunits that it links. The insertion of the target prodermorphin between the dissociated chains A and B near the IHR replaced some of the original inter-subunit interactions that were altered with IHR dissociation. Specifically, the prodermorphin bound to residues of the IHR that were previously bound to the other subunit (Table 1). The total energy associated per interaction type is dependent on the number of interactions per model.

Table 1. Chemical interactions between residue pairs within the IHR in the presence or absence of the target prodermorphin.

| Type of interaction (Energy value) | Absence of prodermorphin (Chain A – B IHR) $n = 5$ | Presence of prodermorphin (Chain C – Chain A/B IHR) | |
|---|---|--|--------------------------------------|
| | | Mature Peptide; $n = 10$ (Closed conf.) | Pro-region; $n = 10$ (Open conf.) |
| Van der Waals Interactions (6.0 kJ/mol) | A168 – A267 | E24 – A267 | K6 – G266 |
| | T169 – A267 | E24 – T268 | K6 – T268 |
| | - | - | S9 – G266 |
| | - | - | S9 – A267 |
| | - | - | N13 – T268 |
| Hydrogen bond (17.0 kJ/mol) | A168 – G266 | E24 – G266 | E12 – G266 |
| | A168 – A267 | E24 – T268 | E12 – A267 |
| | E172 – G266 | R27 – G266 | - |
| | E172 – A267 | - | - |
| Total bonds | 6 | 5 | 7 |
| Total energy | 80.0 kJ/mol | 63.0 kJ/mol | 64.0 kJ/mol |

Note: Energy values per interaction type are based on Martin *et al.* (2003) and Piovesan *et al.* (2016) [34, 35].

The change in interaction type combinations may explain the non-significant difference ($p = 0.1237$) in the total energy observed between the mature peptide region and pro-region interactions with AlaR. Since AlaR putatively operates with the observed largescale transitions between closed and open states, a compensatory interplay of the associated interaction energies in these conformations may occur in order to promote enzymatic activity. Specifically, when AlaR assumes an open conformation, the pro-region interacts with the catalytic residues. This higher energy state attained when the pro-region binds to AlaR drives the shift to a lower energy state where the mature peptide region binds to AlaR in its closed conformation. However, the corresponding rearrangement of intramolecular contacts in AlaR with the shift from the open conformation results in a similar net value for total energies in the closed and open states.

Further investigation of the individual interaction types was done to characterize this shift in associations. Among the six interaction types examined, four types had a significant difference in the total energies between control and experimental docked models. More salt bridges were observed in AlaR-pro-region interactions, with $p = 0.0234$, while more π - π interactions were found in AlaR-mature peptide interactions, with $p = 0.0498$. However, no significant differences in hydrogen bonding and van der Waals interactions were observed between the two experimental setups ($p = 0.01519$ and $p = 0.6459$, respectively).

Since pro-region binding is assumed to occur with the open conformation of AlaR, the strong and long-range electrostatic interactions brought about by the salt bridges may be necessary to stabilize the overall system for AlaR function [42]. The π - π interactions were observed to be higher in regions of AlaR interaction with the mature peptide region than in the pro-region. This increased stability provides a means to orient the target residues for racemization.

No significant difference was observed for the total hydrogen bond and van der Waals energies of interaction between the mature peptide region vs. pro-region and the AlaR's catalytic residues. Both these energy values remain to be low as the

corresponding control models associated in each interaction type. In the absence of the target prodermorphin, a significant difference was observed between the crystal structure and the control ($p = 0.0302$ for hydrogen bonds and $p = 0.0076$ for van der Waals interactions). Majority of the contacts with the pro-region and mature peptide region with AlaR are hydrogen bonds and van der Waals interactions, with some salt bridges. This persisted despite the shift in bond frequencies between pro-region and mature peptide region of interactions with AlaR.

TrpR-procontryphan-R interaction energies

Unlike in AlaR, the total interaction energies observed when the mature peptide region binds to TrpR are significantly different from those when the pro-region binds to TrpR ($p = 0.0466$). This is despite the absence of a transition between open and closed conformations.

A higher amount of energy is needed to bind and stabilize the mature peptide region with the TrpR's catalytic residues for catalysis to proceed. The specific interaction types and their associated energies are provided in Table 2 which shows one representative model for each set-up. The lower energy associated with pro-region binding suggests a potential sequence for the binding of these two targets, where initial binding occurs with the pro-region. Then, propeptide movement allows eventual access to the mature peptide region for catalysis.

In the TrpR-procontryphan-R system, significant differences between control and experimental groups were observed for salt bridges, $p = 0.0196$; van der Waals interactions, $p < 0.0001$; π - π stacking interactions, $p = 0.0004$; and π -cation interactions, $p = 0.0271$. Among these interactions, only π - π interactions were not significant between the mature peptide region and pro-region interactions with TrpR ($p = 0.0919$).

Regardless of procontryphan-R position, lower π - π interactions were observed, as compared to the control. Sequence analysis revealed minimal distribution of AAs that form π bonds in procontryphan-R's pro-region and mature peptide region. There are two tryptophan residues in the mature peptide region and one phenylalanine in the pro-region. The occurrence of π - π stacks in

Table 2. A comparison of the interchain chemical interactions between residue pairs on one set of docked models in the presence or absence of procontryphan-R.

| Type of interaction (Energy value) | Absence of procontryphan-R (Chain A – B) | | Presence of procontryphan-R (Chain C – Chain A/B; Closed conformation) | | |
|--|--|--|--|--|---|
| | | | Mature Peptide | | Pro-region |
| Hydrogen Bond (17.0 kJ/mol) | F127-H325 Y94-I45 Y31-E109 R47-E86 | K39-Y100 K39-Y98 E46-K90 S128-T324 | G40-K295 G32-Y98 C33-Y98 E36-Y98 | R30-I45 R30-E46 R29-Y100 | P37-Y40 R30-E46 E36-S48 |
| Van der Waals Interactions (6.0 kJ/mol) | I32-V101 I45-K90 I24-E102 L27-R106 Y31-V105 Y31-A108 Y31-E109 I32-Y100 I32-V105 S36-Y100 F38-Y98 K39-L99 I45-F91 I45-Y94 F71-Y94 T76-K89 V81-K89 V84-I88 I88-I88 F92-M301 | F92-P302 F92-N331 F127-F127 F127-H325 F127-H328 S128-T324 S128-H325 L177-P319 P302-H325 P302-H328 F323-F323 I32-Y100 P33-Y100 P33-V318 S36-T98 K39-Y98 I45-Y94 F92-330 F92-441 N93-P302 | W35-F38 P37-A68 W38-Y40 C39-F416 G40-D238 G32-L99 R30-E46 F24-I24 V27-I24 L28-I24 L28-L27 C39-F38 F24-L27 M25-Y31 | V27-V44 L28-I32 W35-K39 W38-L65 W38-I412 C39-Y40 N26-K90 N26-F91 R29-F91 R29-Y94 R29-L99 C33-Y94 E36-Y98 | C33-I45 W35-S67 S21-I88 S21-F92 F24-F92 M25-V84 M25-I88 M25-F92 M25-M301 M25-V330 M25-N331 R29-M301 C33-F71 W35-F38 E36-I45 E36-I53 E36-S67 E36-H75 P37-M43 P37-I45 W38-K39 |
| Ionic bonds (20.0 kJ/mol) | E46-K90 | R47-E86 | E36-K173 | | - |
| π-π stacking (9.4 kJ/mol) | F38-Y98 F127-F127 F127-H325 F127-H328 | F323-F323 F127-F323 F71-Y94 | - | | F24-F92 |
| π-cation (9.6 kJ/mol) | Y98-K39 | - | - | | - |
| Total bonds | 58 | | 35 | | 25 |
| Total energy | 491.4 kJ/mol | | 301.0 kJ/mol | | 186.4 kJ/mol |

Note: Energy values per interaction type are based on Martin *et al.* (2003) and Piovesan *et al.* (2016) [34, 35].

procontryphan-R is localized in these aromatic residues.

Three interaction types showed significant differences in the total energies between the two test groups. Specifically, these are the salt bridges ($p = 0.0190$), van der Waals interactions ($p = 0.0001$), and π -cation interactions ($p = 0.0163$). More salt bridges were found among models of TrpR-pro-region interactions compared to TrpR-mature peptide region docks. Van der Waals interactions were greater in models of TrpR-mature peptide region contacts. Very few π -cation interactions were observed in TrpR, regardless of control or experimental docks. Furthermore, most of the bond types present in the models were van der Waals interactions, as well as some hydrogen bonds and π - π stacking interactions. It was observed that the concentration of the interactions encompasses the C-terminal end of the pro-region and the mature peptide region. Additionally, when the mature peptide region interacts with TrpR, there are less hydrogen bonds and van der Waals interactions observed in Trp4, as compared to Trp7. This observation may suggest why only Trp4 is isomerized to the D-isomer, and not Trp7. The decreased stabilization of Trp4, with its decreased interactions suggests greater freedom for mobility, possibly aiding the isomerization process.

Interactions of putative recognition sequences

From the multiple sequence alignment of available prodermorphin and procontryphan-R sequences,

the tripeptides SEM (for prodermorphin) and AVP (for procontryphan-R) were found to be fully conserved in the pro-regions of the two peptides. Due to their full conservation, it has been hypothesized that these tripeptide sequences are possible signaling sequences for the isomerization of the target residues, Ala2 and Trp4, of dermorphin and contryphan-R, respectively. Upon identification of the interactions observed for the multistep docked models, 2 of 10 (20%) pro-region-interacting models showed the interaction between the catalytic Lys of AlaR and the tripeptide SEM (residues 23-25 of prodermorphin) through hydrogen bonding (Figure 9). However, the interaction between the catalytic Lys of TrpR and the tripeptide AVP was not observed for any model.

Because the interaction of tripeptide SEM with the catalytic residues was only seen in 20% of the multistep docked models of AlaR that represent functional interactions, this may represent one of many possible conformations during the racemization process. Thus, while our current data does not confirm the hypothesis that the SEM tripeptide acts as the signal for propeptide recognition, it does not preclude its potential importance in orienting the amino acid for racemization. Analysis of the individual interchain interactions when the mature peptide region interacts with the AlaR's catalytic residues suggest that the SEM tripeptide interacts more frequently with certain residues in AlaR in docks that qualify as functional interactions based on the previously stated criteria

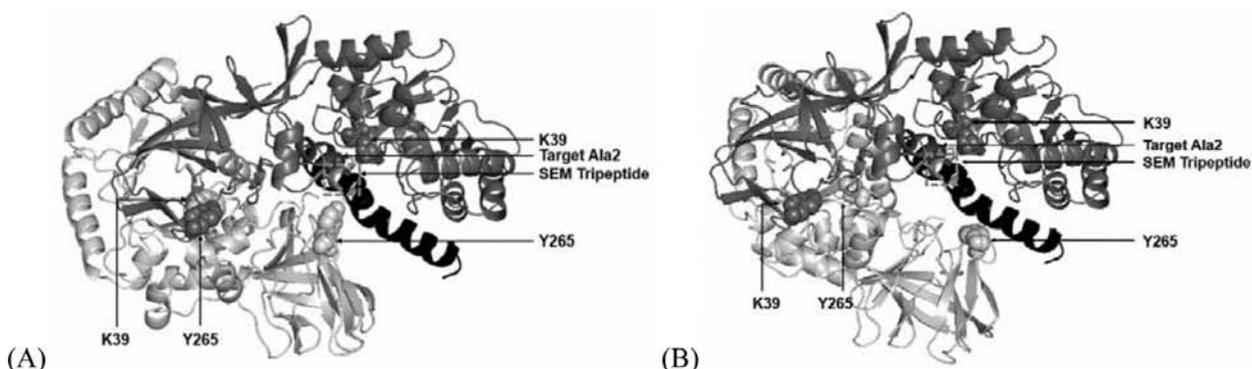


Figure 9. Interactions of the SEM tripeptide (colored light gray, stick model) with AlaR, particularly with the K39 catalytic residue (space filling model) of chain A (dark gray), (A) in two qualified multistep docked models; (B) each shows a different model). Nearby residues from chain B (light gray) are also able to interact with this region.

(Figure 1). Inter-chain van der Waals interactions were found in E350 and I352 (each with frequencies of 6/10 from screened models; 60%) of chain A, and hydrogen bonding was found in R290 (6/10 screened models; 60%) and R309 (7/10 screened models; 70%) of chain B. Meanwhile, the interactions of the AVP tripeptide with TrpR revealed duplicate van der Waals interactions in E85 and L178 of chain B, each with a frequency of 3/35 screened models (8.57%). While the observed interaction frequencies remain to be adequate among multistep docked models that follow the functional dimeric orientation of the racemase, the current data may suggest the idea that the full conservation of the tripeptide sequences may play a role in orienting the target residue location.

DISCUSSION

The present work investigated the post-translational mechanisms of D-AA incorporation into peptides, with dermorphin and contryphan-R as models. The interactions between the propeptides and their respective racemases were examined using molecular modeling and molecular docking simulations. Specific chemical interactions were analyzed to rationalize the possible modes of racemase dimer formation and to correlate them with the functional activity of racemase. The analyses to determine key residues in both the racemase and target propeptide sequences provided further elucidation of the L- to D-AA isomerization process.

Our findings suggest that racemases may utilize at least two modes of action for the isomerization process. It was observed that AlaR assumed closed and open conformations depending on the position and interaction with the bound prodermorphin. In contrast, irrespective of the position of procontryphan-R, TrpR adopted only a closed conformation. The dissimilarity in the sequences for prodermorphin and procontryphan-R, in addition to their respective racemases may provide a basis for predicting the appropriate mode of action expected for other target peptides based on their relatedness to these two peptides.

Despite the difference in modes, both systems involved different binding interactions for the pro-region and mature peptide region. For AlaR, this

provided a possible mechanism for their respective capture and alignment with the catalytic residues of the racemase TIM barrel domain.

The observed changes in binding may be related to the energies of the predominant interactions with either the pro-region or mature peptide region. In AlaR, the importance of long-range ionic interactions is seen for the initial binding events with the pro-region. Shifts to more hydrophobic interactions (e.g. VDW and π -stacking) with mature peptide region binding aid peptide orientation for catalysis. In contrast, for TrpR, hydrophobic interactions within the closed conformation orient the bound pro-region away from the catalytic residues until a shift in VDW interactions allows the binding of the mature peptide target.

In addition to shifts in the predominant interactions, the total energy for the dimeric racemase was reduced in the presence of the propeptide. The different combinations of interaction types involved with the binding of either the pro-region or the mature peptide region coincide with energy tradeoffs that may drive the binding and isomerization within the target AA residues.

Testing the possibility of post-translational L- to D-amino acid isomerization before folding and propeptide cleavage

Isomerization from L- to D-AA is known to be possible through post-translational modifications. Yet, the structural bases for this process have not been fully elucidated for many D-AA-containing peptides. The availability of several propeptide sequences and structures for the short D-AA-containing-peptides, dermorphin and contryphan-R, and the racemases that can isomerize their target AA residues allowed us to investigate these potential structural bases *in silico*.

The discovery of D-Trp, and the identification of an encoding cDNA for contryphan-R prepropeptide (referred to as preprocontryphan-R, which consists of signal sequence, pro-region and mature peptide) [4] indicates that contryphan-R is possibly generated in *Conus radiatus* venom ducts through the pathway shown in Figure 10. It is presumed in this pathway that the L- to D-AA isomerization occurs prior to the propeptide

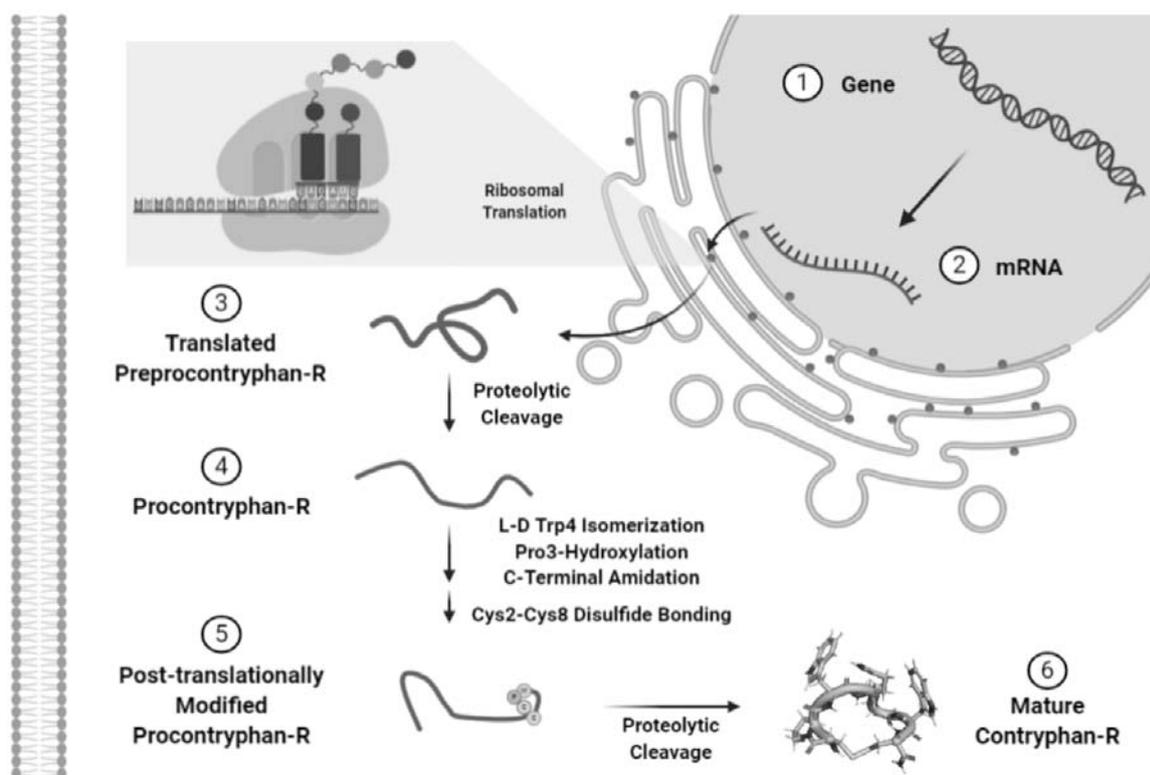


Figure 10. Proposed pathway for the generation of mature contryphan-R. The cyclic structure of mature contryphan-R was adapted from Pallaghy *et al.* (2000; PDB 1DG0) [43]. This figure was created using BioRender (biorender.com).

cleavage [19]. Interestingly, our docking studies documented binding interactions between the TrpR racemase and the linear form, but not between the racemase and the cyclic form of procontryphan-R. This indicates that folding may occur after the isomerization. The observed interaction of AlaR and the linear prodermorphin in either open or closed conformation suggests that prodermorphin isomerization and maturation presumably occur through a similar pathway.

L- to D- amino acid isomerization in multicellular organisms

The occurrence of post-translational isomerization is documented in several venom peptides. The L- to D-Ser46 isomerization in the 48-AA peptide, Aga-IVB, isolated from the venom of the spider, *Agelenopsis aperta* was found to be catalyzed by an enzyme with aminoacyl-L/D isomerase activity also found in the venom. This enzyme can isomerize some AA residues, including Ala, within peptide chains, in either way from the L- or

the D-isomer. The process involved a two-base mechanism, in which the abstraction of a proton by a base from one side occurred simultaneously with the delivery of a proton by the conjugate acid of the second base from the other side [44]. Similar L- to D-AA isomerization process was observed with the isomerase isolated from the venom of the platypus, *Ornithorhynchus anatinus* [45]. The occurrence of the isomerizing enzyme in the venom, in the absence of the translational machinery, supports the hypothesized post-translational nature of this process.

Our current results with dermorphin and contryphan-R support their isomerization in a similar post-translational process. In addition, our results suggest that the isomerization occurs prior to propeptide cleavage, as indicated by the cooperative binding of the propeptides with their respective racemases.

In another study, the L- to D-AA isomerization of the crustacean hyperglycemic hormone (CHH) was

observed to occur after propeptide cleavage. This process took place in the perikarya of fully specialized neurosecretory cells [16]. The relatively large size of CHH (72-AA) compared to dermorphin (7AA) and contryphan-R (8 AA) probably entails that isomerization takes place after the proteolytic cleavage and release of this peptide.

Significance of D-amino acid incorporation

It must be noted that the data analyzed for this study included a relatively small number of docks that satisfied presumed conditions for functional racemase interactions. The low number of selected docks may reflect the general rarity of the occurrence of L- to D-AA isomerization in the context of the pervasive stereo-selective bias for L-AA-containing peptides.

Elucidating the mechanisms for D-AA incorporation in peptide targets provide means to design novel peptide-based therapies. Dermorphins, D-AA-containing opioid peptides derived from frog skin are of great interest for the development of novel analgesics with distinct mechanisms of action. Dermorphin originally derived from *Phyllomedusa sauvagei* was found to be more potent and selective for the μ -opioid receptor, with long-lasting analgesic effect compared to morphine upon administration in animal models [46]. The initial clinical trial of randomized, placebo-controlled study in the relief of post-operative pain showed that dermorphin administered intrathecally was much better than morphine and placebo [47].

Contryphans comprise a large family of D-AA-containing peptides that have potential pharmacological applications. Several contryphans were found to have neuroexcitatory activities that may be further explored as potential therapies for neurological conditions [3].

The presence of D-AAs in peptides often results in the formation of unique tertiary structure, increased stability and/or enhanced potency. The D-Ala2 in dermorphin is vital for its binding to the μ -opioid receptor and for its pharmacological activity. Dermorphin is a long-acting and potent opioid peptide while the [L-Ala2]dermorphin is almost fully devoid of biological activity [48]. The presence of D-amino acid in contryphan-R initially isolated from *Conus radiatus* results in

greater stability compared to [L-Trp4]contryphan-R that was observed to be readily degraded to a colored derivative [18]. A recent study has linked the increased stability of contryphan-Sm (originally isolated from *Conus stercusmuscarum*) [49] to the structural features achieved due to the presence of D-Trp4, as compared with [L-Trp4]contryphan-Sm [50].

CONCLUSION

This study involved the *in silico* analysis of racemase and target propeptide interactions in order to define mechanisms for L-to-D AA isomerization. The observed racemase-propeptide interactions suggest the occurrence of at least two modes of action for racemase function. One mode involves conformational transitions of racemase between open and closed forms, while another retains a closed dimeric conformation. Both modes involve differential binding of the target propeptide sequence at either the pro-region or the mature peptide region. The shift in binding provides a means of orienting the target AA residues with the catalytic domains of racemase. The observed predominance of particular interaction types for the two racemase-propeptide systems suggests possible ways of designing D-AA incorporation for potential target sequences based on their compatibility with the racemases. While the predominant interaction types for different racemase binding conformations have been defined, further analysis of the significance of specific residue interactions for L- to D-AA isomerization targeting must be carried out to verify their predicted importance for targeted peptide modification.

ACKNOWLEDGEMENTS

The support of the NIMBB-UPD (to MHRC and NADB) and the UPB (to ECJ) is greatly appreciated. This work is partly based on the B. S. (MBB) thesis of MHRC.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

REFERENCES

1. Kreil, G. 1997, Annu. Rev. Biochem., 66, 337.

2. Bai, L., Sheeley, S. and Sweedler, J. V. 2010, *Bioanal. Rev.*, 1, 7.
3. Jimenez, E. C. 2020, *Curr. Prot. Pept. Lett.*, 21, 622.
4. Olivera, B. M., Walker, C., Cartier, G. E., Hooper, D., Santos, A. D., Schoenfeld, R., Shetty, R., Watkins, M., Bandyopadhyay, P. and Hillyard, D. R. 1999, *Ann. N. Y. Acad. Sci.*, 870, 223.
5. Richter, K., Egger, R. and Kreil, G. 1987, *Science*, 238, 200.
6. Richter, K., Egger, R., Negri, L., Corsi, R., Severini, C. and Kreil, G. 1990, *Proc. Natl. Acad. Sci. USA*, 87, 4836.
7. Yasuda-Kamatani, Y., Nakamura, M., Minakata, H., Nomoto, K. and Sakiyama, F. 1995, *J. Neurochem.*, 64, 2248.
8. Matsuo, R., Kobayashi, S., Morishita, F. and Ito, E. 2011, *Comp. Biochem. Physiol. B Biochem. Mol. Biol.*, 160, 89.
9. Heck, S. D., Siok, C. J., Krapcho, K. J., Kelbaugh, P. R., Thadeio, P. F., Welch, M. J., Williams, R. D., Ganong, A. H., Kelly, M. E., Lanzetti, A. J., Gray, W. R., Phillips, D., Parks, T. N., Jackson, H., Ahlijanian, M. K., Saccomano, N. A. and Volkmann, R. A. 1994, *Science*, 266, 1065.
10. Fujita, K., Minakata, H., Nomoto, K., Furukawa, Y. and Kobayashi, M. 1995, *Peptides*, 16, 565.
11. Yasuda-Kamatani, Y., Kobayashi, M., Yasuda, A., Fujita, T., Minakata, H., Nomoto, K., Nakamura, M. and Sakiyama, F. 1997, *Peptides*, 18, 347.
12. Morishita, F., Nakanishi, Y., Kaku, S., Furukawa, Y., Ohta, S., Hirata, T., Ohtani, M., Fujisawa, Y., Muneoka, Y. and Matsushima, O. 1997, *Biochem. Biophys. Res. Commun.*, 240, 354.
13. Dutertre, S., Lumsden, N. G., Alewood, P. F. and Lewis, R. J. 2006, *FEBS Lett.*, 580, 3860.
14. Checco, J. W., Zhang, G., Yuan, W. D., Le, Z. W., Jing, J. and Sweedler, J. V. 2018, *J. Biol. Chem.*, 293, 16862.
15. Heck, S. D., Faraci, W. S., Kelbaugh, P. R., Saccomano, N. A., Thadeio, P. F., Volkmann, R. A. 1996, *Proc. Natl. Acad. Sci. USA*, 93, 4036.
16. Soyez, D., Toullec, J. Y., Ollivaux, C. and Geraud, G. 2000, *J. Biol. Chem.*, 275, 37870.
17. Montecucchi, P. C., de Castiglione, R., Piani, S., Gozzini, L. and Erspamer, V. 1981, *Int. J. Pept. Protein Res.*, 17, 275.
18. Jimenez, E. C., Olivera, B. M., Gray, W. R. and Cruz, L. J. 1996, *J. Biol. Chem.*, 271, 28002.
19. Jimenez, E. C., Craig, A. G., Watkins, M., Hillyard, D. R., Gray, W. R., Gulyas, J., Rivier, J. E., Cruz, L. J. and Olivera, B. M. 1997, *Biochemistry*, 36, 989.
20. Stamper, G. F., Morollo, A. A. and Ringe, D. 1998, *Biochemistry*, 37, 10438.
21. Roy, A., Kucukural, A. and Zhang, Y. 2010, *Nat. Protocols*, 5, 725.
22. Yang, J. and Zhang, Y. 2015, *Nucl. Acids Res.*, 43, W174.
23. Yang, J., Yan, R., Xu, D., Poisson, J. and Zhang, Y. 2015, *Nat. Methods*, 12, 7.
24. Guex, N. and Peitsch, M. C. 1997, *Electrophoresis*, 18, 2714.
25. Kozakov, D., Beglov, D., Bohnuud, T., Mottarella, S. E., Xia, B., Hall, D. R. and Vajda, S. 2013, *Proteins Structure Function Bioinformatics*, 81, 2159.
26. Kozakov, D., Hall, D. R., Xia, B., Porter, K. A., Padhorny, D., Yueh, C., Beglov, D. and Vajda, S. 2017, *Nat. Protoc.*, 12, 255.
27. Vajda, S., Yueh, C., Beglov, D., Bohnuud, T., Mottarella, S. E., Xia, B., Hall, D. R. and Kozakov, D. 2017, *Proteins Structure Function Bioinformatics*, 85, 435.
28. Liang, J., Han, Q., Tan, Y., Ding, H. and Li, J. 2019, *Front. Mol. Biosci.*, doi:10.3389/fmolb.2019.00004.
29. GNU, P. 2007, Free Software Foundation, Bash (3.2.48), Unix Shell Program.
30. Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., Li, Q., Shoemaker, B. A., Thiessen, P. A., Yu, B., Zaslavsky, L., Zhang, J. and Bolton, E. E. 2019, *Nucl. Acids Res.*, 47, D1102.
31. Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S. and Olson, A. J. 2009, *J. Comput. Chem.*, 16, 2785.
32. Trott, O. and Olson, A. J. 2010, *J. Comput. Chem.*, 31, 455.

33. DeLano, W. 2002, The PyMOL Molecular Graphics System. DeLano Scientific, San Carlos, CA, USA. <http://www.pymol.org>
34. Martin, A. J. M., Vidotto, M., Boscaroli, F., Di Domenico, T., Walsh, I. and Tosatto, S. C. E. 2011, *Bioinformatics*, 27, 2003.
35. Piovesan, D., Minervini, G. and Tosatto, S. C. E. 2016, *Nucl. Acids Res.*, 44, W367-W374.
36. Corpet, F. 1988, *Nucl. Acids Res.*, 16, 10881.
37. The UniProt Consortium. 2019, *Nucl. Acids Res.*, 47, D506.
38. Jimenez, E. C., Watkins, M., Juszczak, L. J., Cruz, L. J. and Olivera, B. M. 2001, *Toxicon*, 39, 803.
39. Pi, C., Liu, Y., Peng, C., Jiang, X., Liu, J., Xu, B., Yu, X., Yu, Y., Jiang, X., Wang, L., Dong, M., Chen, S. and Xu, A. L. 2006, *Biochimie*, 88, 131.
40. Thakur, S. S. and Balaram, P. 2007, *Rapid Commun. Mass Spectrom.*, 21, 3420.
41. Galakatos, N. and Walsh, C. 1989, *Biochemistry*, 28, 8167.
42. Karshikoff, A. and Jelesarov, I. 2008, *Biotech. Biotechnol. Equipment*, 22, 606.
43. Pallaghy, P. K., He, W., Jimenez, E. C., Olivera, B. M. and Norton, R. S. 2000, *Biochemistry*, 39, 12845-12852.
44. Heck, S. D., Faraci, W. S., Kelbaugh, P. R., Saccomano, N. A., Thadeio, P. F. and Volkmann, R. A. 1996, *Proc. Natl. Acad. Sci. USA*, 93, 4036.
45. Bansal, P. S., Torres, A. M., Crossett, B., Wong, K. K., Koh, J. M., Geraghty, D. P., Vandenberg, J. I. and Kuchel, P. W. 2008, *J. Biol. Chem.*, 283, 8969.
46. Broccardo, M., Erspamer, V., Falconieri Erspamer, G., Improta, G., Linari, G., Melchiorri, P. and Montecucchi, P. C. 1981, *Br. J. Pharmacol.*, 73, 625.
47. Basso, N., Marcelli, M., Ginaldi, A. and De Marco, M. 1985, *Peptides*, 6(Suppl 3), 177.
48. Giagnoni, G., Mennuni, L., Pecora, N., Basilico, L., Parolaro, D. and Gori, E. 1987, *Pharmacol. Res. Commu.*, 19, 173.
49. Jacobsen, R., Jimenez, E. C., Grilley, M., Watkins, M., Hillyard, D., Cruz, L. J. and Olivera, B. M. 1998, *J. Pept. Res.*, 51, 173.
50. Bascos, N. A. D. and Jimenez, E. C. 2018, *Curr. Top. Pept. Prot. Res.*, 19, 27.