Discovery and Engineering of the L-Threonine Aldolase from *Neptunomonas Marine* for Efficient Synthesis of β-Hydroxy-α-Amino Acids via C–C Formation

Yuanzhi He¹, Siyuan Li², Jun Wang¹, Xinrui Yang², Jiawei Zhu¹, Qi Zhang¹, Li Cui¹, Zaigao Tan¹, Wupeng Yan¹, Yong Zhang¹, Luyao Tang¹, Lin-Tai Da²*, and Yan Feng¹*

[1] State Key Laboratory of Microbial Metabolism, Joint International Research Laboratory of Metabolic & Developmental Sciences, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai, China

[2] Key Laboratory of Systems Biomedicine (Ministry of Education), Shanghai Center for Systems Biomedicine, Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai, China

*Correspondence: yfeng2009@sjtu.edu.cn; darlt@sjtu.edu.cn

**ABSTRACT:** L-Threonine aldolases (LTAs) are attractive biocatalysts for synthesizing β-hydroxy-α-amino acids (HAAs) via C–C bond formation in pharmaceuticals, although their industrial applications suffer from low activity and diastereoselectivity. Herein, we describe the discovery of a new LTA from *Neptunomonas marine* (NmLTA) that displays both ideal enzymatic activity (64.8 U/mg) and diastereoselectivity (89.5% diastereomeric excess; de) for the desired product L-threo-4-methylsulfonylphenylserine (L-threo-MPTS). Using X-ray crystallography, site-directed mutagenesis, and computational modeling, we propose a “dual-conformation” mechanism for the diastereoselectivity control of NmLTA, whereby the incoming 4-
methylsulfonylbenzaldehyde (4-MTB) could potentially bind at the \textit{Nm}LTA active site in two distinct orientations, potentially forming two diastereoisomers (\textit{threo-} or \textit{erythro-}form products). Importantly, two key \textit{Nm}LTA residues H140 and Y319 play critical roles in fine-tuning the binding mode of 4-MTB, supported by our site-mutagenesis assays. Uncovering of the catalytic mechanism in \textit{Nm}LTA guides us to further improve the diastereoselectivity of this enzyme. A triple variant of \textit{Nm}LTA (N18S/Q39R/Y319L; SRL) exhibited both improved diastereoselectivity (de value > 99\%) and enzymatic activity (95.7 U/mg) for the synthesis of \textit{l-threo-}MPTS compared with that of wild type. The preparative gram-scale synthesis for \textit{l-threo-}MPTS with the SRL variant produced a space-time yield of up to 9.0 g L$^{-1}$h$^{-1}$, suggesting a potential role as a robust \textit{C–C} bond synthetic tool for industrial synthesis of HAAs at a preparative scale. Finally, the SRL variant accepted a wider range of aromatic aldehyde derivatives as substrates and exhibited improved diastereoselectivity toward \textit{para-}site substituents. This work provides deep structural insights into the molecular mechanism underlying the catalysis in \textit{Nm}LTA and pinpoints the key structural motifs responsible for regulating the diastereoselectivity control, thereby guiding future attempts for protein engineering of various LTAs from different sources.

**KEYWORDS:** \textit{l-threonine aldolase}, protein engineering, diastereoselectivity mechanism, \(\beta\)-hydroxyl-\(\alpha\)-amino acids, crystal structure

**INTRODUCTION**

Carbon–carbon (\(\textit{C–C}\)) bond forming reactions are highly effective strategies for constructing complex molecules by connecting small building blocks.$^{1,2}$ The aldol reaction is an attractive approach to stereoselectively synthesize functionalized chiral amino acids via \(\textit{C–C}\) formation.$^{3-6}$ Enzymatic catalysis has been widely applied in aldol reaction for environmental-friendly and
efficient synthesis of numerous valuable chemicals and pharmaceuticals by enabling ideal stereoselectivity. In particular, microbial L-threonine aldolases (LTAs), which are pyridoxal-5’phosphate (PLP)-dependent enzymes, have received considerable attention for functionalizing the C–C bond by catalyzing the aldol reaction of various aldehydes and α-amino acids in one-step to form diverse β-hydroxy-α-amino acids (HAAs) with two adjacent chiral centers (C\textsubscript{α} and C\textsubscript{β}) (Scheme 1). HAAs belong to an important subclass of chiral intermediates with antimicrobial, antitumor, and immunosuppressive activities, and are therefore widely used to produce valuable pharmaceuticals (Figure S1). For example, L-threo-4-methylsulfonylphenylserine (L-threo-MPTS) serves as an important intermediate in the synthesis of the glycopeptide antibiotics florfenicol and thiamphenicol, and L-threo-4-nitrophenylserine is a key precursor for the production of chloramphenicol. Although LTA has strict stereoselectivity at the C\textsubscript{α}-site, the stereoselectivity at the C\textsubscript{β}-site is much less conserved. However, initial efforts to improve the diastereoselectivity of \textit{Pseudomonas} sp. LTA (PsLTA) using error-prone PCR only had a limited effect on increasing the diastereomeric excess (de) value.

In comparison, structure-based engineering strategies have become more efficient owing to the availability of an increasing number of LTA structures from various sources. These structures reveal that LTA forms a tetrameric protein whose active-site residues are located within the interface between adjacent monomers. The LTA conformation was captured in various states, including the apo state, and in complexes with varied substrates in binary and ternary forms. Thus, these crystal structures provide atomic-level detail of the molecular recognitions between LTA and its substrates, thereby guiding further enzyme engineering. Several groups have designed a number of LTA variants with improved stereoselectivity at the C\textsubscript{β}-site based on these crystal structures, or homolog models. For example, Lin’s group increased the de value of...
Actinocorallia herbida LTA (AhLTA) from 58% to 81% for L-threo-MPTS by modulating the substrate binding pocket,\(^1\) and they further enhanced the AhLTA de value to 93.7% via rational design by combining molecular dynamics simulations and free-energy calculations.\(^2\) Cai et al. improved the diastereoselectivity of Pseudomonas putida LTA (PpLTA) from 32% to 85% for L-threo-4-nitrophenylserine via computer-aided directed evolution.\(^3\) In addition, Zheng and colleagues conducted a series of structure-based designs of LTA variants to produce an enhanced de value up to 99% for L-threo-MPTS.\(^4\) This group also proposed that the substrate 4-methylsulfonylbenzaldehyde (4-MTB) could potentially bind to LTA along two distinct channels, enabling in the formation of two diastereomeric products.\(^5\)–\(^8\)

Although the diastereoselectivities of certain LTAs have been improved via protein engineering, their enzymatic activities were considerably undermined,\(^9\)–\(^11\) which limited their applications in HAA-related pharmaceutical synthesis. Currently, it remains challenging to simultaneously improve the activity and C-\(\beta\)-site stereoselectivity of LTAs. Moreover, given that the structure of the aldehyde-bound LTA complex is still unknown, the key structural motifs involved in regulating the LTA diastereoselectivity remain obscure, hampering further protein engineering. Intriguingly, the enzymes from marine microorganisms can exhibit remarkable activities owing to their habitat-related characteristics, such as using unique carbon sources, including petroleum, polycyclic aromatic hydrocarbons, and fatty acid residues from marine animals.\(^12\)–\(^15\) Therefore, we sought to explore new LTAs from marine sources to mine highly active LTAs for C–C condensation with high diastereoselectivity as well as ideal activity.

Herein, we describe the discovery of a highly active Neptunomonas marine LTA (NmLTA) that synthesized L-threo-MPTS with the highest enzymatic activity and diastereoselectivity among the currently reported native LTAs. By combining X-ray crystallography and computational
modeling, we pinpointed the key regulatory motifs in \( Nm \)\( LTA \) responsible for controlling the diastereoselectivity of the desired product, which prompted us to propose a novel diastereoselectivity mechanism. Furthermore, site-directed and combinatorial mutagenesis assays identified a triple mutant of \( Nm \)\( LTA \), N18S/Q39R/Y319L (termed SRL), that could further improve the \( \text{de} \) value > 99% for \( L \)-threo-MPTS, more importantly, with a \(~1.5\)-fold increase in enzyme activity compared with that of wild type (WT). This work thus sets a solid foundation for future applications of LTA in biocatalytic synthesis of industrially important functional HAAs.

RESULTS AND DISCUSSION

Screening and characterization of LTAs from marine microbes. To search for LTAs with potential enzymatic activities, we used the synthetic reaction of \( L \)-threo-MPTS from 4-MTB and glycine as a model reaction (Scheme 1). A neighbor-joining phylogenetic tree was constructed for several LTAs from different sources according to their sequence similarity, which included 18 marine and 6 terrigenous microbial LTAs, among which three LTAs with available crystal structures were also included.\(^{26,28} \) Terrigenous-sourced LTAs were grouped in the same clade, whereas the ones from marine source could be divided into five clades (Figure 1A). For each marine clade, we selected one representative LTA namely \( MILTA \) (from \( Mameliella \) \( sp. \)), \( SpLTA \) (from \( Sulfitobacter pontiacus \), \( AILTA \) (from \( Actibacterium lipolyticum \)), \( Nm \)\( LTA \), and \( RpLTA \) (from \( Rhodopirellula \) \( sp. \)), and overexpressed these in heterologous \( Escherichia coli \) (\( E. coli \)) using pET22b with a his-tag. For comparison, three terrigenous LTAs with available crystal structures, i.e., \( AjLTA \) (from \( Aeromonas jandaei \) DK-39), \( PpLTA \) (from \( Pseudomonas putida \)), and \( EcLTA \) (from \( E. coli \)), were overexpressed in \( E. coli \) (Figure S2).\(^{26,28} \) We then examined the enzymatic activities of these LTAs against the model reaction using high performance liquid chromatography.
(HPLC) assays.\(^{38}\) N,N-dimethylformamide (DMF) was added as a co-solvent supporting the solubility of 4-MTB in the reaction (Table S3).

Most of the tested LTAs favored the \textit{threo}-configuration product, whereas only \textit{PpLTA} and \textit{EcLTA} preferentially produced the \textit{erythro}-configuration (Figure 1B and Table S1). Notably, all marine LTAs had a de value of > 50\% for \textit{l-threo}-MPTS. Among these, \textit{NmLTA} exhibited the highest de value of 89.5\% and an efficient enzymatic activity (64.8 U/mg) that was \(~\)1.4-fold higher than the highest native LTA reported so far (Figure 1B and Table 1).\(^{33}\) Further analyses showed that \textit{NmLTA} had a maximum enzymatic activity at 30 °C and pH 8.0, and the efficiency could be further improved by increasing the PLP concentration (Figure S4). For thermal stability, \textit{NmLTA} remained highly active when incubated at 25 °C for over 140 min and remained 60\% of its maximal at 50 °C over 60 min (Figure S5). Kinetic analyses for \textit{NmLTA} determined the respective \textit{k}_{\text{cat}} and \textit{K}_{\text{m}} values of 190.4 s\(^{-1}\) and 4.5 mM for \textit{l-threo}-MTPS and 805.7 s\(^{-1}\) and 12.3 mM for the natural substrate \textit{l-threonine} (Tables S2 and S7). Thus, \textit{NmLTA} is a promising enzymatic tool for establishing C–C bond owning to the ideal activity, which established a good foundation for continued attempts to improve the diastereoselectivity and activity.

\textbf{Structural features of PLP-bound \textit{NmLTA} complex.} To investigate the structural features of \textit{NmLTA}, we used X-crystallography to obtain the crystal structure of \textit{NmLTA} complexed with the PLP cofactor. After screening 1,600 crystallization conditions, the complex structure was resolved at a resolution of 2.8 Å (PDB ID: 7YVR) (Figures 2A, S6, and Table S4), which prompted us to dig deeper into the catalytic mechanism of the \textit{C\_beta} stereoselectivity. \textit{NmLTA} belongs to the canonical PLP-dependent fold-type I enzyme family and has a homo-tetramer (Figure 2A).\(^{11}\) The functional group of PLP forms a Schiff base (internal aldimine) with the highly conserved K214 residue and directly interacts with \textit{NmLTA} residues at the monomer interfaces (inserted in Figure
2A). PLP can form π–π stacking interaction with the imidazole group of H95 and establish several hydrogen bonds (HBs) with D181 and R184 via the hydroxyl and imine groups. Moreover, G70, S71, and K239 stabilized the PLP phosphate group. Comparisons of the active-site residues of NmLTA with that of other LTAs (PpLTA, EcLTA, and AjLTA)\textsuperscript{26, 28} indicate that these residues are structurally highly conserved, highlighting their critical roles in substrate binding and catalysis (Figures 2B and S3).

We then compared the potential substrate binding pockets for different LTAs, as two large potential binding cavities (A and B) can be observed in all LTAs (Figure 2C). Analyses of the surface electrostatic potential for both cavities revealed that NmLTA exhibits similar electrostatic properties to PpLTA, with the former displaying more positive charges in cavity A than the latter because of the presence of R322. In sharp contrast, EcLTA and AjLTA demonstrated quite different electrostatic surfaces compared with those of NmLTA. Additional structural analyses of cavity A show that NmLTA contains a unique binding channel along which Y319 and R322 are located (Figure 2D), whereas either nonpolar (e.g., F314 for PpLTA) or neutral residues (e.g., S304 for EcLTA) are present in other LTAs (Figure 2D). Considering that the above four LTAs demonstrate distinct activities and diastereoselectivities for the synthesis of L-threo-MPTS (Figure 1B), we speculated that the varied residues and charge distributions in the LTA active site may dictate their different lead substrate specificities.

**Identification of potential NmLTA residues responsible for diastereoselectivity control.**

After the binding of PLP to LTA, the catalysis proceeds first by binding of the α-amino acid (e.g., glycine) into the LTA active site, forming the PLP-Gly conjugate and displacing K214. According to the formerly proposed catalytic mechanism,\textsuperscript{39} the protonation of 4-MTB is likely promoted by a base-mediated water bridge. The activated PLP-Gly nucleophile will then attack the incoming
aldehyde, leading to the C–C bond formation (Figure 3A). To scrutinize the structural features
during the enzymatic aldol reaction, we first performed molecular docking for the intermediate
PLP-Gly with NmLTA. The energy-minimized NmLTA-PLP-Gly complex revealed a similar
binding mode for PLP to that observed in the NmLTA-PLP complex. In addition, the carboxyl
group of PLP-Gly could establish stable salt bridges with R184 and R329 and one HB with S16
(Figure 3B). This computational model was further supported by site-directed mutagenesis, which
indicated that mutations of R184A and R329A could completely abolish the activity of NmLTA,
whereas the S16A mutant reduces the activity by approximately 80% (Figure 3C).

Using NmLTA-PLP-Gly model, we docked the aldehyde substrate 4-MTB in the NmLTA
active sites. 4-MTB could potentially bind to NmLTA in two different binding sites (namely,
cavities A and B). Notably, in cavity B, 4-MTB formed few direct contacts with the NmLTA
residues, and moreover, the distance between the nucleophile PLP-Gly and the aldehyde group of
4-MTB was too distant (6.3 Å) to perform the reaction (Figure S7). Therefore, cavity B is an
inappropriate binding site for 4-MTB. In comparison, in cavity A, 4-MTB fitted into a region
flanked by R322, Y319, H140, and H95 where 4-MTB was sandwiched between H140 and Y319
and formed HBs with R322 and Y319 via the sulfonic-acid group (Figure 3D). Intriguingly, we
observed two distinct orientations of the aldehyde group in 4-MTB, which probably gives rise to
the two different aldol products (namely, threo- or erythro-form at the Cβ-site). The threo-
configuration, which is the major product of NmLTA, is located closer to H140 and forms one
additional HB with H95 that is absent in the erythro-form, giving rise to a more favorable docking
free energy (−4.9 and −3.8 kcal/mol for threo- and erythro-forms, respectively). Collectively, our
findings suggest a “dual-conformation” mechanism for regulating the diastereoselectivity control
of NmLTA (Figure 3D).
These structural insights guided further attempts at enzyme engineering to improve the diastereoselectivity of \textit{Nm}LTA. Previous studies reported that the activities of various LTA mutants were profoundly decreased, although with enhanced diastereoselectivity, such as for engineered \textit{Bn}LTA, \textit{Lm}LTA, and \textit{Cs}LTA (Table 1).\textsuperscript{30, 32, 33} Here, to simultaneously improve the activity and diastereoselectivity, we first selected all the LTA residues located within 6 Å around the “dual-conformation” 4-MTB, which identified 11 residues, namely S16, N18, Q39, P40, Y41, H95, D99, D138, H140, Y319, and R322 (Figure 3E). Among these residues, the counterpart residues of residue H95 and D99 in other LTAs are known to play vital roles in substrate specificity and recognition, respectively.\textsuperscript{23, 40} Consistently, site-directed mutagenesis assays also proved that the H95A mutation could completely abolish the activity of \textit{Nm}LTA (Figure S8). We also validated that P40, Y41, D138, and R322 are critical for specific activity or diastereoselectivity of \textit{Nm}LTA and that the individual residue substitution would considerably impair either only the activity (i.e., D138) or both activity and diastereoselectivity (i.e., P40, Y41, D99, and R322) (Figure S8). Therefore, the remaining five residues (i.e., S16, N18, Q39, H140, and Y319) were selected for further systematic investigations.

**Site-saturation and combinatorial mutagenesis of selected \textit{Nm}LTA residues.** For each of above five chosen residues, we performed saturation mutagenesis using the NNK codon, and diastereoselectivity and activity screening were implemented using 4-MTB and glycine as substrates with a molecular ratio of 1:10, followed by HPLC analysis (Figures 4A and 4B). For each single mutation, both the de value and enzymatic activity were measured for \textit{l}-\textit{threo}-MPTS. For H140-site, all single substitutions produced a considerable decrease in the de value compared with that of the WT, and the H140L mutation had the lowest de value (\(-46.2\%\)) (Figure 4A), indicating its essential role in regulating the diastereoselectivity. By contrast, we observed
profoundly improved diastereoselectivity for several substitutions at the other four sites (Figure 4C). Specifically, we obtained a total of twelve NmLTA variants with elevated de values ranging from 91.1% to 94.3% in the following order: S16A > S16G > Y319D > Y319L > Y319S > Q39R > N18S > Q39A > Q39L > Q39K > Q39H > N18T (Figure 4C). For enzymatic activity, except for Y319D and mutations at S16 and N18, all other positive substitutions displayed essentially equal (e.g., Q39H, Q39L, Q39A, and Y319L) or even higher (e.g., Q39R, Q39K, and Y319S) activities compared with that of the WT (Figures 4B, 4C, and Table S6).

To further enhance the diastereoselectivity and activity of NmLTA, we conducted combinatorial mutagenesis based on the above twelve single mutants. Screening of the double-mutant libraries produced six candidates that exhibited enhanced diastereoselectivity, including S16G/Y319L, Q39R/Y319L, N18S/Y319L, S16A/Q39R, Q39R/Y319D, and S16A/Y319L (Figure 4D). Intriguingly, Q39R/Y319L showed the highest diastereoselectivity among all double-mutant variants (de value of 97.6%), and critically, it also had a 2-fold higher specific activity (133.0 U/mg) than that of the WT (64.8 U/mg) (Figures 4D, 4E, and Table S6). In addition, the $k_{cat}$ and substrate performance values for the Q39R/Y319L mutant increased to 400.3 s$^{-1}$ and 160.1, respectively, which were ~2 and 6-fold higher than that of the WT (Table S7). We then performed a third round of screening based on these double mutants and identified two triple mutants, namely N18S/Q39R/Y319L (SRL) and N18S/Q39R/Y319D (SRD), with further improved the de values of 99.3% and 98.5%, respectively (Figure 4D and Table S6). Notably, the SRL variant had a specific activity of 95.7 U/mg and a $k_{cat}/K_m$ of 78.9 s$^{-1}$mM$^{-1}$, which were ~1.5-fold and 1.9-fold higher than those of the WT, respectively. Moreover, the SRL variant displayed a ~10.7-fold higher substrate performance than that of the WT (297.5 vs. 27.8, respectively) (Figure 4E and
Table S7). Altogether, the SRL variant of NmLTA not only showed ideal diastereoselectivity for L-threo-MPTS but also displayed improved enzymatic activities compared with that of the WT.

**Dual-conformation regulation for the diastereoselectivity control of NmLTA.** As described above, our computational studies indicated that 4-MTB could potentially bind at the NmLTA cavity A in two distinct orientations, clamped by the sidechains of H140 and Y319 sidechains. Thus, the activated PLP-Gly could potentially attack the carbonyl group of 4-MTB from two possible directions, resulting in the threo- or erythro-product (Figure 5D and Figure S9). We therefore proposed a “dual-conformation” regulation for the diastereoselectivity control of NmLTA whereby H140 and Y319 play key roles in fine-tuning the substrate conformation. The proposed mechanism was supported by site-directed mutagenesis, which indicated that substitutions of H140 with any other residues could profoundly decrease the de value for the threo-configuration product and most of these mutants underwent diastereoselective inversion, whereas mutations of Y319 to Leu, Ser, or Asp had a reverse effect. In comparison, we performed a saturation mutation of V155 that is located in cavity B of NmLTA. Notably, the mutation of the counterpart residue V155 in BnLTA (I143R) and CsLTA (V143R) enhanced de values (Table 1).31, 33 By contrast, none of the variants at the V155 site exerted any considerable influence on the diastereoselectivity for the synthesis of L-threo-MPTS (Figure S10), supporting our conclusion that cavity B does not participate in regulating the substrate diastereoselectivity. This work therefore rules out the formerly proposes “two pathways” mechanism as observed in other LTA,30, 31, 33 and proposes a novel diastereoselectivity mechanism for NmLTA. Interestingly, a similar mechanism was also observed in a pyruvate-dependent aldolase, where the pyruvate enolate could attack the si or re face of the substrate succinic semialdehyde to yield racemic products.41
We next compared the structural difference between the WT NmLTA and the SRL variant by constructing an SRL model based on the WT crystal structure. Structural inspections revealed that the SRL variant can profoundly alter the binding mode of the incoming aldehyde in the active site. Specifically, the Y319L substitution weakened the packing interactions between the aromatic ring of the aldehyde and Y319 as observed in the WT, which subsequently pushed the substrate toward H140, leading to the preferred formation of the threo-product (Figures 5A, 5B, and 5D). This finding was consistent with the above observations that all the single mutants at H140 yielded reduced de value for l-threo-MPTS (Figures 4A and 5D). In addition, the Q39R mutation provided an additional anchor point to stabilize the sulfonic acid tail of 4-MTB together with R322, which likely increased enzymatic activity (Figures 5A and 5B). Moreover, the N18S mutation likely imposed unsubstantial structural perturbations on PLP-Gly by altering the HB interaction with Y41, slightly improving diastereoselectivity (Figure S11). Consistently, our binding free-energy calculations showed that the SRL variant of NmLTA demonstrated stronger binding affinity between the aldehyde and protein compared with that of the WT (Figure 5C), and that this was mainly caused by the additional salt bridge interaction between R39 and the sulfonic-acid group of 4-MTB. Notably, previous studies of other LTAs had identified counterpart residues of NmLTA H140, Y319, and N18 (e.g., H136 and Y314 in AhLTA, H142 and Y319 in LmLTA, Y8H in BnLTA and N16A in AhLTA), suggesting that these residues may serve as universal sites for diastereoselectivity control (Figure S12). The Q39 position, however, has not yet been examined for other LTAs, therefore serving as a potential engineering site for other LTAs.

**Catalytic performance of the SRL variant in preparative gram-scale l-threo-MPTS.** To evaluate the performance of NmLTA for the potential industrial synthesis of l-threo-MPTS, we first examined the thermal stability of the SRL variant. The results showed that the SRL variant
maintained > 98% activity after heating at 25 °C for 140 min and still retained 70% of the activity at 50 °C for 60 min, demonstrating improved thermostability over that of the WT (Figure S5). We then evaluated the gram-scale synthesis of L-threo-MPTS in a 1 L scale reactor by including 100 mM 4-MTB and 1 M glycine with 1,400 U WT NmLTA or the SRL variant. Measurement of the de value and yield of L-threo-MPTS indicated that the SRL variant always maintained strict diastereoselectivity with increased conversion rate, whereas for the WT, the de value slightly decreased as the conversion rate increased. For the SRL variant, the yield of L-threo-MPTS was 21.0 g/L, with a de value of 99.3% and space-time yield of 9.0 g L⁻¹ h⁻¹. Thus, the SRL variant greatly improved the catalytic performance of NmLTA, displaying much stricter diastereoselectivity compared with that of the WT (Figure 6). Therefore, the SRL variant could potentially serve as a robust C–C bond synthetic tool for industrial synthesis of HAAs on a preparative scale.

Substrate scope of the evolved NmLTA variant. To further examine the substrate scope of the WT NmLTA and the SRL variant, we investigated the enzymatic activities and diastereoselectivities of LTAs toward a representative library of benzaldehyde derivatives. The WT NmLTA and the SRL variant could recognize and catalyze all tested substrates except for the 4-OH substituted benzaldehyde, with the conversion rates varying between 58% and 80% (Figure 7). As listed in Figure 7, all para-substituted substrates had increased de values except for 4-F (de value > 80%), in particular, for 4a, the de value could reach up to 85.7%. For the meta-site, only 3-Br-substituted substrate improved the de value, whereas substitution with 2-Cl or 2-NO₂ at ortho-site increased the de values. Moreover, two tested disubstituted substrates, 3,4-CH₃ and 3,5-CH₃, undermined the enzymatic performance, which is likely due to the large steric hindrances caused by the substituents. For substrate benzaldehyde, the SRL variant shows a considerable
increase in the de value (12a). Altogether, the SRL variant shows a broad substrate scope for the
synthesis of diverse HAAs, which are important synthetic precursors for bioactive compounds and
therapeutic drugs (e.g., 5a for chloramphenicol). Notably, different substrates with varied
substituents could apparently impose considerable influences on the enzymatic activities and
diastereoselectivity control of the SRL variant, likely caused by the perturbed binding mode of the
substrates. Despite that, the SRL variant appeared more tolerant toward the para-substituted
substrates by electronegative or neutral groups (except for 4-F).

CONCLUSION

Herein, we report a robust LTA from marine N. marine (NmLTA) that established the C–C
bond with high enzymatic activity and strict diastereoselectivity control of the Cβ-site. X-ray
crystallography, computational modeling, and site-directed mutagenesis allowed us to design a
triple mutant of NmLTA, namely the SRL variant, that exhibits > 99% de value for L-threo-MPTS,
and importantly, demonstrates ~1.5-fold increased enzymatic activity (95.7 U/mg) compared with
that of WT to achieve simultaneous enhancement of activity and diastereoselectivity for the
synthesis of L-threo-MPTS. Meanwhile, the SRL variant of NmLTA exhibited the highest specific
activity for synthesizing L-threo-MPTS among the currently reported LTAs (Table 1).33 Moreover,
the preparative gram-scale aldol reaction of L-threo-MPTS catalyzed by the SRL variant produced
a space-time yield of up to 9.0 g L⁻¹h⁻¹. Furthermore, the substrate spectrum assays indicated that
the SRL variant could accept a wider range of aromatic aldehyde derivatives as substrates and
show improved diastereoselectivity by tuning the para-site substituents. Our work identified the
key structural motifs in NmLTA responsible for regulating its activity and diastereoselectivity and
proposed a novel molecular mechanism underlying the diastereoselectivity control of C–C bond
formation, thereby accelerating the application of the aldol reaction using engineered LTAs for the biocatalytic preparation of industrially important functional HAAs.

AUTHOR INFORMATION

Supporting Information

Chemicals and Materials used in this study; Methods of gene mining, synthesis, expression, protein purification, site-directed mutagenesis, and screening of variant; Activity, kinetic, and substrate specificity assay; Crystallization and protein structure determination; Molecular docking, molecular dynamic simulations, and calculations of molecular mechanics Poisson-Boltzmann Surface Area; HPLC and MS Analysis; Primers and relative DNA sequences used in this study.

Corresponding Author

* Yan Feng – State Key Laboratory of Microbial Metabolism, Joint International Research Laboratory of Metabolic & Developmental Sciences, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai, China. E-mail: yfeng2009@sjtu.edu.cn

* Lin-Tai Da – Key Laboratory of Systems Biomedicine (Ministry of Education), Shanghai Center for Systems Biomedicine, Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai, China. E-mail: darlt@sjtu.edu.cn

Funding Sources

This work was financially supported by the Ministry of Science and Technology of the People's Republic of China (grant no. 2020YFA0907700) and National Natural Science Foundation of China (grant no. 32271306 and 22177072).
The authors declare no competing financial interest.

ACKNOWLEDGEMENTS

We thank staffs of beamlines BL18U1 and BL19U1 at Shanghai Synchrotron Radiation Facility for their assistance in diffraction data collection. We thank Prof. Juan Lin (College of Chemical Engineering, Fuzhou University, China) for providing the product standard.

REFERENCES


(39) Rocha, J. F.; Sousa, S. F.; Cerqueira, N. M. F. S. A. Computational studies devoted to the catalytic mechanism of threonine aldolase, a critical enzyme in the pharmaceutical industry to synthesize \(\beta\)-hydroxy-\(\alpha\)-amino acids. *ACS Catal.* **2022**, *12* (9), 4990-4999.
(40) Chen, Q.; Chen, X.; Feng, J.; Wu, Q.; Zhu, D.; Ma, Y. Improving and Inverting Cβ-
stereselectivity of threonine aldolase via substrate-binding-guided mutagenesis and a stepwise

(41) Coincon, M.; Wang, W.; Sygusch, J.; Seah, S. Y. Crystal structure of reaction intermediates
in pyruvate class II aldolase: substrate cleavage, enolate stabilization, and substrate specificity. *J.
Scheme 1. The aldol reaction catalyzed by LTA for the synthesis of \( \alpha \)-threo-MPTS.
Figure 1. Mining and identification of LTAs with ideal activity and diastereoselectivity. (A) A simplified neighbor-joining phylogenetic tree showing the evolutionary relationship between characterized terrigenous LTAs and the newly discovered marine microbe LTAs. Marine and terrigenous microbe-sourced clades are colored in orange and gray background, respectively. Five candidate marine LTAs (e.g., MILTA, SpLTA, AlLTA, NmLTA, and RpLTA) are labeled at the tree branches with red triangles. Three formerly characterized LTAs from Bacillus nealsonii (BnLTA), Leishmania major (LmLTA), and Actinocorallia herbida (AhLTA) are labeled on the tree branches with blue triangles. Three LTAs (e.g., AlLTA, PpLTA, and EcLTA) with available crystal structures are labeled at the tree branches with green triangles. All eight LTAs used in this study are highlighted in bold. (B) Specific activity and diastereoselectivity of LTAs for L-threo-MPTS. Reaction conditions: 100 mM aldehyde, 1 M glycine, 50 μM PLP, 10% DMF, and 20 μg/mL purified LTAs in 1 mL of 50 mM HEPES-NaOH buffer (pH 8.0) at 25 °C and 250 rpm within 30 min.
Figure 2. Crystal structure of NmLTA and comparisons with those of other LTAs. (A) Overall structure of PLP-bound WT NmLTA complex. The asymmetric unit of protein contains four monomers (A–D) in the P121 space group (left panel). The PLP cofactor (magenta sticks) is attached to the highly conserved K214 via forming a Schiff base and interacts with the LTA residues in two adjacent monomers (cyan and yellow). Key hydrogen bonds formed between PLP and the surrounding residues are highlighted in gray dashed lines. (B) Structural comparisons of active site residues from different LTAs, including NmLTA (yellow), EcLTA (green), AjLTA (cyan), and PpLTA (deep salmon). The PLP cofactors from all sources are shown as magenta sticks. Crystal structures of NmLTA, PpLTA, EcLTA, and AjLTA are derived from the Protein Data Bank identities: 7YVR, 5VYE, 4LNJ, and 3WGB, respectively. (C) Electrostatic potential surface of the substrate channel in four LTAs, colored from red (acidic) to blue (basic). PLP, PLR (4′-deoxypyridoxine phosphate), PLP-Gly, and PLP-K214 (LLP) are shown as magenta sticks. (D) Comparisons of the residues along the substrate channel in four LTAs.
Figure 3. Computational modeling of substrate-bound NmLTA complexes. (A) Schiff base exchange mechanism of NmLTA. (B) Docking model of PLP-Gly-bound NmLTA complex. Key HBs are highlighted in dashed lines. (C) Alanine scanning of residues surrounding the carboxyl group of PLP-Gly. (D) Docking models of ternary NmLTA complexes where PLP-Gly and 4-MTB are shown as magenta and green sticks, respectively. Two docking conformations of 4-MTB are provided, corresponding to two configuration products (R/S configuration at Cα-site). (E) Highlight of the residues within 6 Å surrounding the 4-MTB.
Figure 4. Saturation site-directed mutagenesis for evolving diastereoselectivity of NmLTA. (A) Diastereoselectivities of NmLTA mutants for the product l-threo-MPTS. (B) Relative activities of NmLTA mutants for the product l-threo-MPTS and the catalytic activity of WT NmLTA was defined as 100%. (C) The de value and specific activities of several positive mutants. (D) Development of combinatorial mutation strategy of LTAs with enhanced de value starting from the WT NmLTA. (E) Specific activities of WT and different NmLTA variants. Reaction conditions: 100 mM 4-MTB, 1 M glycine, 50 μM PLP, 10% DMF, and 20 μg/mL purified LTAs in 1 mL of 50 mM HEPES buffer (pH 8.0) at 25 °C and 250 rpm within 30 min.
**Figure 5.** The “dual-conformation” regulation in *NmLTA*. (A) Conformation of 4-MTB in WT *NmLTA*. Residues Q39, H140, Y319, and R322 line the active site channel (labeled by purple, cyan, and yellow, respectively). Gray dashed lines represent hydrogen bond interactions. (B) The conformation of 4-MTB in the SRL variant. The grey dashed lines represent hydrogen bond interactions. (C) Binding free-energy calculation of the WT *NmLTA* and the SRL variant using the MM/PBSA algorithm during the initial 1 ns. (D) Dual-conformation regulation for the diastereoselectivity control of *NmLTA*. 
Figure 6. Catalytic performances of WT *Nm*LTa and the SRL variant for the synthesis of L-threo-MPTS. Reaction conditions: 100 mM 4-MTB, 1 M glycine, 50 μM PLP, 10% DMF, and 1400 U LTAs in 1 L of 50 mM HEPES buffer (pH 8.0) at 25 °C and 250 rpm within 160 min.
Figure 7. Diastereoselectivities and conversions of WT NmLTA and the SRL variant for the synthesis of HAA 1a–18a. Reaction conditions: 100 mM aldehyde, 1 M glycine, 50 μM PLP, 10% DMF, and 1.5 U LTAs in 1 mL of 50 mM HEPES buffer (pH 8.0) at 25 °C and 250 rpm within 120 min. See Figures S13–S29 for all liquid chromatograms and mass spectrograms.
Table 1. Comparison of the de values and specific activities for the synthesis of L-threo-MPTS by *Nm* LTA and other identified LTAs.

<table>
<thead>
<tr>
<th>Organism</th>
<th>Enzyme</th>
<th>WT or variant</th>
<th>de value</th>
<th>Specific Activity (U/mg)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Pseudomonas</em> sp.(^{12,14})</td>
<td><em>Ps</em> LTA</td>
<td>WT</td>
<td>53.0%</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>V200I</td>
<td>71.0%</td>
<td>-</td>
</tr>
<tr>
<td><em>Bacillus nealsonii</em> (^{30,31})</td>
<td><em>Bn</em> LTA</td>
<td>WT</td>
<td>30.3%</td>
<td>2.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Y31H/N305R</td>
<td>93.1%</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Y8H/Y31H/I143R/N305R</td>
<td>&gt; 99%</td>
<td>-</td>
</tr>
<tr>
<td><em>Actinocrallia herbida</em> (^{19,23})</td>
<td><em>Ah</em> LTA</td>
<td>WT</td>
<td>61.0%</td>
<td>2.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Y314R</td>
<td>81.0%</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N16A/E98S/Y314R</td>
<td>93.7%</td>
<td>-</td>
</tr>
<tr>
<td><em>Leishmania major</em> (^{32})</td>
<td><em>Lm</em> LTA</td>
<td>WT</td>
<td>26.8%</td>
<td>17.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Y45H/Y319F/V321W</td>
<td>96.3%</td>
<td>12.1</td>
</tr>
<tr>
<td><em>Cellulosilyticum</em> sp.(^{33})</td>
<td><em>Cs</em> LTA</td>
<td>WT</td>
<td>37.0%</td>
<td>46.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Y8H/V143R/H305L</td>
<td>&gt; 99%</td>
<td>9.3</td>
</tr>
<tr>
<td><em>Neptunomonas marina</em></td>
<td><em>Nm</em> LTA</td>
<td>WT</td>
<td>89.5%</td>
<td>64.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N18S/Q39R/Y319L</td>
<td>&gt; 99%</td>
<td>95.7</td>
</tr>
</tbody>
</table>