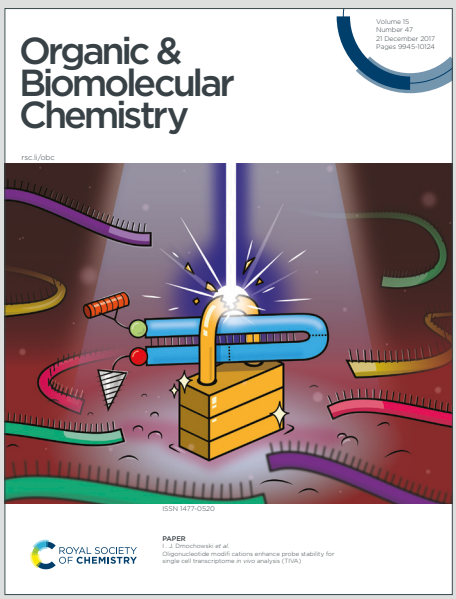
Organic & Biomolecular Chemistry

Accepted Manuscript

This article can be cited before page numbers have been issued, to do this please use: C. M. Thompson, A. D. McDonald, H. Yang, S. Cavagnero and A. R. Buller, *Org. Biomol. Chem.*, 2020, DOI: 10.1039/D0OB00868K.



ROYAL SOCIETY OF CHEMISTRY

rsc.li/obc

View Article Online
DOI: 10.1039/D0OB00868K

COMMUNICATION

Modular Control of L-Tryptophan Isotopic Substitution via an Efficient Biosynthetic Cascade

Received 00th January 20xx,
Accepted 00th January 20xx

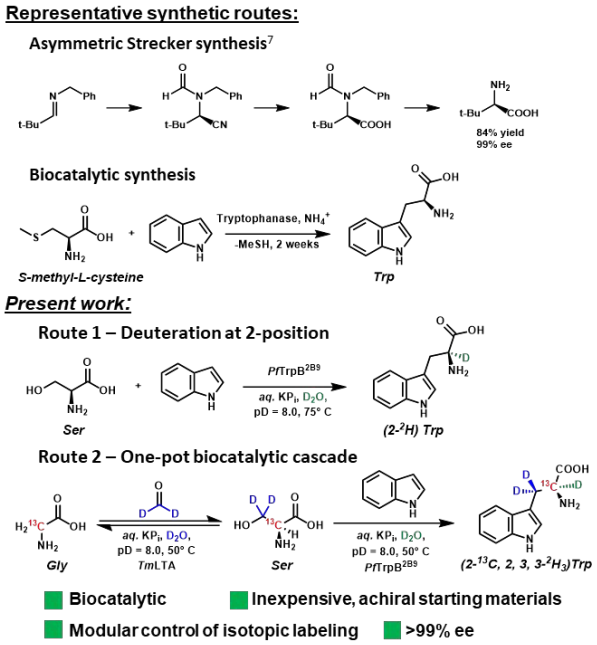
DOI: 10.1039/x0xx00000x

Clayton M. Thompson,[a] Allwin D. McDonald [a], Hanming Yang [a], Silvia Cavagnero [a]*, and Andrew R. Buller[a]*

Isotopologs are powerful tools for investigating biological systems. We report a biosynthetic-cascade synthesis of Trp isotopologs starting from indole, glycine, and formaldehyde using the enzymes L-threonine aldolase and an engineered β-subunit of tryptophan synthase. This modular route to Trp isotopologs is simple and inexpensive, enabling facile access to these compounds.

Isotopically substituted amino acids are valuable tools for the mechanistic and structural analysis of biological systems.[1] Among the standard amino acids, L-tryptophan (Trp) is the most structurally complex and serves as the precursor to diverse natural products and clinically used compounds.[2] Selective isotopic substitution of Trp is often essential for determining the metabolic fate of individual atoms and the kinetic properties of enzymes that manipulate Trp.[3,4] Further, the low prevalence of Trp in proteins makes this amino acid attractive for selective substitution methodologies for protein NMR.[5,6] In each of these cases, however, access to the requisite Trp isotopolog is a central hurdle.

While there are diverse synthetic methods that can produce amino acids, making isotopically substitution compounds comes with unique challenges. For instance, synthetic routes should be planned starting from the cheapest possible isotopically substituted precursor. The typical concerns over catalyst loading and atom economy are secondary to yield with respect to the expensive, isotopically labelled or substituted reagents. Traditional synthetic routes that can access isotopically substituted Trp include the asymmetric Strecker synthesis (Scheme 1)[7] and reductive amination.[8] These approaches, however, often require isotopically enriched starting materials that are not commercially available or are prohibitively expensive. Additionally, most applications of amino acids like Trp benefit from high enantiopurity and chiral small molecule catalysts that are often difficult to obtain.



Scheme 1. Synthetic and proposed routes to L-tryptophan (Trp). $Kp_i = K_xD_{3-x}PO_4$, potassium phosphate buffer.

Enzymatic Trp synthesis is an attractive alternative with the potential for both high selectivity and low environmental impact. For example, previous studies on enzyme mechanism utilized a (2-^{2}H)Trp that was prepared biocatalytically from indole and *S*-methyl cysteine in D_2O using the enzyme tryptophanase (Enzyme Commission (EC) 4.1.99.1, Scheme 1). Although this procedure was successful, it afforded only 40% yield after a 2-day reaction.[9] The synthesis of a [3,3-$^2H_2$]Trp presented a more substantial challenge. To make this isotopolog, methionine γ-lyase (EC 4.4.1.11) was used to catalyze the exchange of the three protons on the 2- and 3-positions of *S*-ethyl-L-cysteine with D_2O solvent over three days in a 47% yield, followed by reaction with indole and

[a] *Department of Chemistry, University of Wisconsin-Madison, 1101 University Ave. Madison, WI. 53706.*

† Footnotes relating to the title and/or authors should appear here.

Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/x0xx00000x

tryptophanase.[10] Methionine γ-lyase accomplishes this labeling through desaturation at C3 during the course of its reversible reaction. Unfortunately, the resulting Trp was composed of a mixture of mono, di-, and tri-deutero isotopologs. Biocatalytic approaches have also been used to prepare various ^{13}C-substituted Trp by the action of tryptophan synthase (EC 4.2.1.20) on the correspondingly substituted ^{13}C-Ser.[11] This ^{13}C-Ser was prepared from isotopically substituted glycine (Gly) and formaldehyde using the enzyme Ser hydroxymethyltransferase (SHMT, EC 2.1.2.1). This method is particularly attractive, as Gly isotopologs are relatively cheap. However, long reactions times (>80 h) were required, and contamination with unreacted Gly necessitated *tert*-butyloxycarbonyl (Boc)-protection followed by column chromatography to isolate the ^{13}C-Ser, reducing yields of the combined process to < 30%. Here, we propose an affordable and efficient method to generate Trp starting from Gly, formaldehyde, and indole in a one-pot, two-step enzymatic cascade reaction that leads to the facile synthesis of enantiopure Trp isotopologs.

We began by considering the synthesis of Trp from Ser and indole. Previous efforts used wild-type tryptophan synthase from *Salmonella typhimurium*, which is a multi-enzyme complex that is tedious to prepare for synthetic applications.[12] The β-subunit of the complex (TrpB) is responsible for the pyridoxal phosphate (PLP)-dependent condensation of Ser and indole. Recently, the TrpB protein from the hyperthermophilic archeon *Pyrococcus furiosus* (*Pf*TrpB) was engineered for activity outside of its native complex.[13,14] Detailed studies have subsequently shown that a stand-alone variant containing eight mutations, *Pf*TrpB^{2B9}, has structural and kinetic properties that are nearly identical to those of the parent complex.[15] We overexpressed *Pf*TrpB^{2B9} in *E. coli* and purified the enzyme using Ni-affinity chromatography.[15] This procedure routinely yielded > 500 mg protein L^{-1} culture. This high expression level is particularly important for practical synthetic applications. To achieve the quantitative deuteration at the 2-position of Trp, we deployed *Pf*TrpB^{2B9} for the condensation of Ser and indole in deuterated buffer (Scheme 1, Route 1). This biocatalytic reaction produced (2-^{2}H)Trp in 89% isolated yield (18.2 mg, Figure 1, Table S1). Given the high yield and excellent properties (expression, speed, thermal stability, etc.) of the catalyst, this synthesis of (2-^{2}H)Trp represents a significant improvement over previous routes. However, synthesis of more sophisticated Trp isotopologs, including the desirable 2-^{13}C substitution, is hampered by the significant cost of the corresponding Ser isotopolog.

Previous efforts to synthesize isotopologs of Ser from Gly utilized the PLP-dependent SHMT enzyme, which uses the expensive tetrahydrofolate cofactor to deliver formaldehyde. This route was further stymied by the need for [Boc]-derivatization and isolation of the Ser isotopolog.[11] We hypothesized that Ser could be produced from formaldehyde and Gly using an L-threonine aldolase (LTA, EC 4.1.2.5). We reasoned that even though the synthetic direction of the LTA reaction is unfavorable (equilibrium lies towards Gly and formaldehyde), the coupled transformation with *Pf*TrpB^{2B9} will provide a strong thermodynamic driving force and obviate the need for intermediate isolation (Scheme 1, Route 2). We chose the *C*-terminally His-tagged LTA from *Thermatoga maritima* (*Tm*LTA) due to its high thermal stability and successful prior applications in biocatalysis.[16] In our hands, heterologous overexpression of *Tm*LTA in *E. coli* yielded ~1 g of protein L^{-1} culture (Figure S1). We began developing the two-enzyme cascade on analytical scale by combining *Tm*LTA and *Pf*TrpB^{2B9}

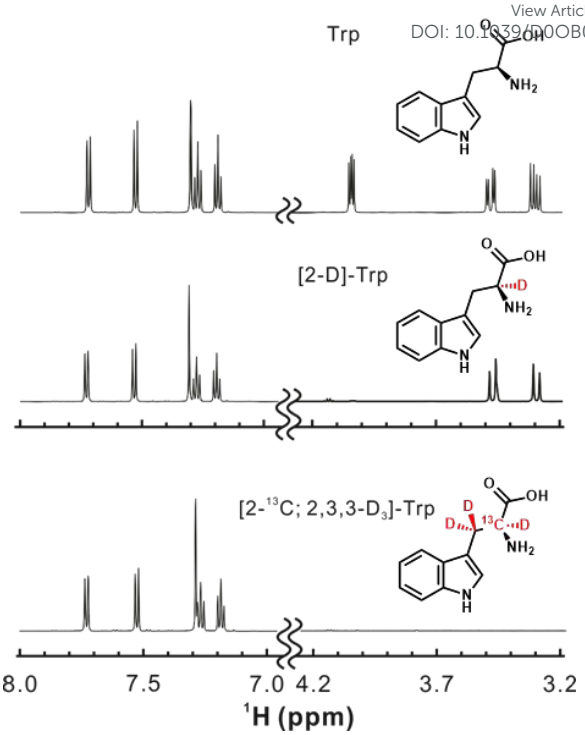


Figure 1. ^{1}H-NMR spectra of synthesized Trp isotopologs. ^{1}H spectra of unlabeled commercial Trp (top), (2-^{2}H)Trp (middle), and (2-^{13}C, 2, 3, 3-$^{2}H_3$)Trp (bottom). See supporting information for more detail.

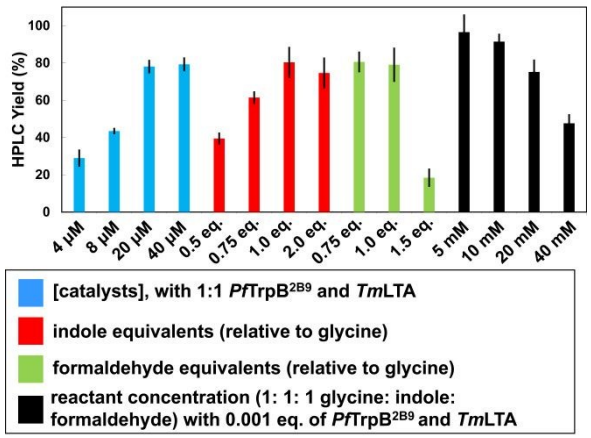


Figure 2. Optimization of *Tm*LTA-*Pf*TrpB^{2B9} cascade reaction. Reference reaction conditions are 5 mM Gly, 5 mM indole, 5 mM formaldehyde, 0.005 mM LTA and *Pf*TrpB^{2B9}, and 0.02 mM PLP and modified as indicated. Reactions were run in quadruplicates and analyzed by UPLC-MS. Uncertainties in percent yield relative to glycine are reported as ± standard error.

Journal Name COMMUNICATION

DOI: 10.1039/D0OB00868K

and explored the effect of variable substrate and catalyst loading. Given that our primary goal was not to explore mechanistic differences among substrate isotopologs, we found it operationally simplest to compare cascade efficiencies under different conditions by evaluating endpoint-product formation. We observed that excess formaldehyde did not result in increases in product formation (Figure 2). Excess formaldehyde increased formation of a species whose mass (m/z = 217) is consistent with formation of an undesired β-carboline through a spontaneous Pictet-Spengler reaction. Indeed, comparing reaction with a full equivalent or only 0.75 equivalents of formaldehyde gave similar amounts of product (Figure 2). Increasing the concentration of indole relative to the other substrates did not increase Trp formation. Reactions proceeded to higher conversions as the substrates were diluted, achieving almost quantitative conversion at 5 mM of each substrate with the ideal ratio of 1:1:1 Gly, formaldehyde, and indole. The catalyst loadings to achieve this conversion were fairly high for enzymatic reactions, 20 μM of each enzyme, corresponding to 0.4 mol% catalyst. In practical terms, however, the high catalyst loadings are compensated for by the high expression of each enzyme.

This new biocatalytic cascade is highly modular with respect to isotope incorporation (Scheme 1, Route 2). The (2-^{13}C)Trp substitution can be incorporated from commercially available (2-^{13}C)Gly, and (2-^{2}H) substitution (deuteration at the α-position) is achieved via exchange with a deuterated solvent, as shown above. Substituting at the 3-position can be controlled via the source of formaldehyde, which is commercially available as a variety of isotopologs. As an exemplar of this the new cascade methodology, we employed the two-enzyme cascade to synthesize the complex (2-^{13}C, 2, 3, 3-$^{2}H_3$)Trp isotopolog starting from the appropriate isotopically enriched precursors. This particular Trp isotopolog was chosen because of its potential as a resonance-assignment aid for protein NMR spectroscopy. In addition, this compound showcases the modular nature of the *Pf*TrpB^{2B9}-LTA cascade. In order to characterize the extent of ^{13}C incorporation at position 2 of Trp and D incorporation at positions 2 and 3, we carried out UPLC-MS as well as ^{1}H and ^{13}C NMR analysis (Figure 1, S5-S8). These data show a complete absence of the 2-H in (2-^{2}H)Trp. In addition, the 2-^{13}C signal, 2- and 3-H signals are absent from the (2-^{13}C, 2, 3, 3-$^{2}H_3$)Trp. The enantiomeric excess (ee) for both Trp isotopologs was found to be > 99% via derivatization with a Marfey's reagent analog and subsequent UPLC-MS analysis (SI Fig. S2).[17] Hence, the optimized reaction conditions led to formation of isotopically pure (2-^{13}C, 2, 3, 3-$^{2}H_3$)Trp in 60% isolated yield (4.8 mg).

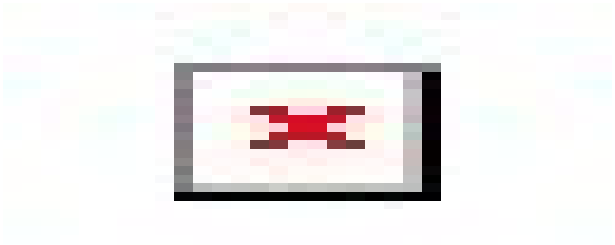


Figure 3. T_1 and T_2 measurements for (2-^{13}C, 2, 3, 3-$^{2}H_3$)Trp and unlabeled commercial Trp.

To further characterize the above Trp isotopolog and probe its potential utility, we carried out 2-^{13}C NMR longitudinal (T_1) and transverse (T_2) relaxation measurements on unlabeled Trp and (2-^{13}C, 2, 3, 3-$^{2}H_3$)Trp (Figure 3). Theory predicts that deuterium substitution results in increased T_1 and T_2, consistent with our experimental observations. This result likely arises from the decrease in the magnetic dipole-dipole interaction due to introduction of the selective deuterium substitution.[18] The decreased ^{13}C relaxation rates can be extremely useful for protein NMR analysis, which typically suffers from short T_2s due to slow protein tumbling rates. In addition, the 2-C signal of uniformly ^{13}C-substituted Trp is split by ^{13}C-^{13}C J coupling from the neighboring nuclei. In contrast, this coupling is not present (2-^{13}C, 2, 3, 3-$^{2}H_3$)Trp. The above combined factors result in a more intense, sharper 2-^{13}C resonance.

In conclusion, we have developed a simple and modular one-pot method for the synthesis of enantiopure Trp isotopologs. The simpler (2-^{2}H)Trp can be accessed through a single enzymatic reaction. Independent control of substitution at the 2- and 3- positions can be achieved through a dual enzyme cascade, as demonstrated by the synthesis of (2-^{13}C, 2, 3, 3-$^{2}H_3$)Trp. We note that each of the enzymes, *Pf*TrpB^{2B9} and *Tm*LTA, can be overexpressed to exceptionally high protein density in *E. coli* (0.5 − 1.0 g protein L^{-1} culture) facilitating preparative-scale reactions. We envision that this modular approach could be implemented with any source of selectively or uniformly substituted glycine, formaldehyde, or indole, which are all commercially available, to yield diverse isotopically-substituted Trps. The resulting products promise to serve as valuable tools in mechanistic biochemical studies as well as structural-biology investigations.

Conflicts of interest

A.R.B is an inventor on a patent on the use of *Pf*TrpB^{2B9} for the synthesis of tryptophans.

Acknowledgments

We are grateful to Jon Ellis for helpful scientific discussions and to Aadhishre Kasat and Meghan Campbell for the synthesis of FDVA. S.C. thanks the National Institute of Health (grant R01GM125995) and the UW-Madison Graduate School for funding. A.R.B. gratefully acknowledges support from the National Institute of Health New Innovator Award (DP2GM137417). The Bruker AVANCE 400 NMR spectrometer was supported by NSF grant CHE-1048642, and the Bruker AVANCE 600 NMR spectrometer was supported by NIH grant S10 OD012245.

References

View Article Online
DOI: 10.1039/D0OB00868K

1 M. D. Tsai, E. Schleicher, R. Potts, G. E. Skye and H. G. Floss, *J. Biol. Chem.*, 1978, **253**, 5344–5349.
2 L. M. Alkhalaf and K. S. Ryan, *Chem. Biol.*, 2015, **22**, 317–328.
3 R. S. Phillips, B. Sundararaju and N. G. Faleev, *J. Am. Chem. Soc.*, 2000, **122**, 1008–1014.
4 A. S. Y, I. Semak, A. Pisarchik, T. Sweatman, A. Szczesniewski and J. Wortsman, *FEBS*, 2002, **511**, 102–106.
5 M. Kainosho and P. Gu, *Q. Rev. Biophys.*, 2009, **42**, 247–300.
6 M. Kainosho, Y. Miyanoiri, T. Terauchi and M. Takeda, *J. Biomol. NMR*, 2018, **71**, 119–127.
7 S. J. Zuend, M. P. Coughlin, M. P. Lalonde and E. N. Jacobsen, *Nature*, 2009, **461**, 968–971.
8 B. Amir-Heidari and J. Micklefield, *J. Org. Chem.*, 2007, **72**, 8950–8953.
9 D. M. Kiick and R. S. Phillips, *Biochemistry*, 1988, **27**, 7339–7344.
10 R. S. Phillips, B. Sundararaju and N. G. Faleev, *J. Am. Chem. Soc.*, 2000, **122**, 1008–1014.
11 J. P. G. Malthouse, T. B. Fitzpatrick, J. J. Milne, L. Grehn and U. L. F. Ragnarsson, *J. Pept. Sci.*, 1997, **3**, 361–366.
12 R. S. Phillips, *Tetrahedron Asymmetry*, 2004, 15, 2787–2792.
13 A. R. Buller, S. Brinkmann-Chen, D. K. Romney, M. Herger, J. Murciano-Calles and F. H. Arnold, *Proc. Natl. Acad. Sci.*, 2015, **112**, 14599–14604.
14 M. Herger, P. Van Roye, D. K. Romney, S. Brinkmann-Chen, A. R. Buller and F. H. Arnold, *J. Am. Chem. Soc.*, 2016, **138**, 8388–8391.
15 A. R. Buller, P. Van Roye, J. K. B. Cahn, R. A. Scheele, M. Herger and F. H. Arnold, *J. Am. Chem. Soc.*, 2018, **140**, 7256–7266.
16 K. Fesko, *Appl Microbiol Biotechnol*, 2016, **100**, 2579–2590.
17 P. Marfey, *Carlsb. Res. Commun.*, 1984, **49**, 591–596.
18 K. Pervushin, R. Riek, G. Wider and K. Wüthrich, *Proc. Natl. Acad. Sci.*, 1997, **94**, 12366–12371.
19 D. K. Romney, J. Murciano-Calles, J. E. Wehrmüller and F. H. Arnold, *J. Am. Chem. Soc.*, 2017, **139**, 10769–10776.
20 D. G. Gibson, L. Young, R.-Y. Chuang, J. C. Venter, C. A. Hutchison and H. O. Smith, *Nat. Methods*, 2009, **6**, 343–345.
21 A. Krezel and W. Bal, *J. Inorg. Biochem.*, 2004, **98**, 161–166.

View Article Online
DOI: 10.1039/D0OB00868K

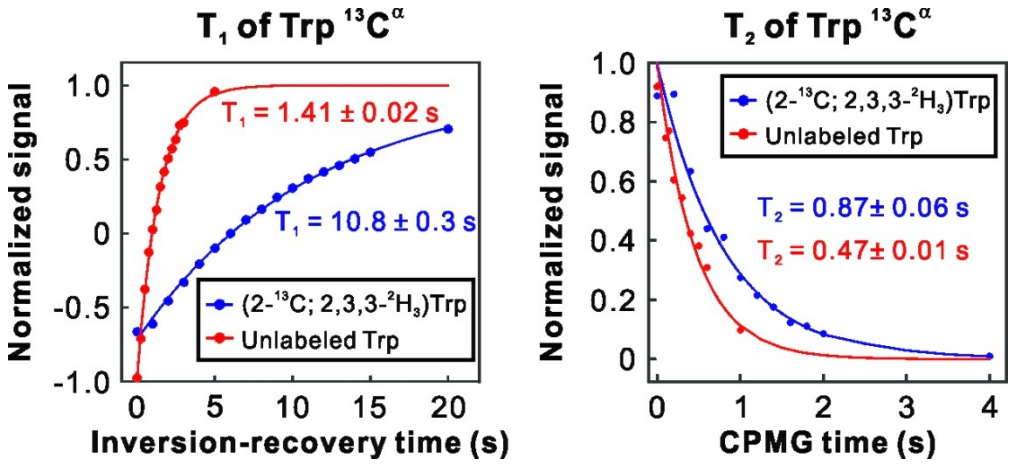
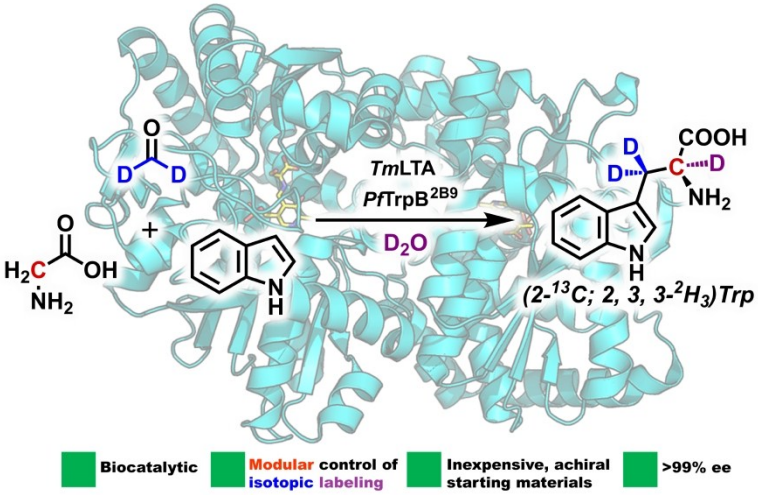


Figure 3. T1 and T2 measurements for (2-13C, 2, 3, 3-2H3)Trp and unlabeled commercial Trp.

View Article Online
DOI: 10.1039/D0OB00868K



338x190mm (300 x 300 DPI)